

Musik für interaktive Medien

Arrangement- und Interpretationstechniken



Dissertation

zur Erlangung des akademischen Grades

**Doktoringenieur
(Dr.-Ing.)**

angenommen durch die Fakultät für Informatik
der Otto-von-Guericke-Universität Magdeburg.

vorgelegt von:	Dipl.-Inform. Axel Berndt
Gutachter:	Prof. Dr. Holger Theisel Prof. Dr. Andreas Nürnberger Prof. Dr. Niels Knolle
Eingereicht am:	18. April 2011
Verteidigt am:	14. Juli 2011

Zusammenfassung

Musik kommt bei der Vertonung interaktiver Medien in vielerlei Weise zum Einsatz. Dabei stellt die interaktive Beeinflussung der Musik die größte Herausforderung dar. Dieser sogenannten *nichtlinearen Musik* wird eine flexible Wandelbarkeit abverlangt, sogar noch während sie spielt. Die vorliegende Arbeit geht auf einige medientheoretische Fragestellungen ein und stellt technische Lösungen vor. Es werden die unterschiedlichen Erscheinungsformen und Aufgaben der Musik in interaktiven Medien behandelt. Die technischen Herangehensweisen an nichtlineare Musik werden systematisiert und zwei Lösungsansätze vorgestellt, der *Echtzeit-Arrangeur* und das *reaktive Orchester*. Diese arbeiten auf der Ebene des Arrangements und der ausdrucksvollen Interpretation. Es werden Modelle zur Beschreibung interpretatorischer Merkmale entwickelt, die auch musikwissenschaftliche und musikpsychologische Relevanz und Anwendungen haben.

Abstract

Music is applied in manifold ways in scoring interactive media. The biggest challenge in this respect is posed by the interactive manipulation of music. This so-called *nonlinear music* must provide for flexible mutability even while it is playing. The following research study addresses some media-theoretical issues and presents technical approaches. The different manifestations and functions of music in interactive media are treated. The technical approaches to nonlinear music are systematized, and two solutions to the problem of flexible mutability are elaborated: the *Realtime Arranger* and the *Reactive Orchestra*. These are shown to operate at the level of arrangement and expressive performance. Models are developed for the description of the performance features. These models are of relevance and have applications in the fields of musicology and music psychology.

Danksagung

Diese Arbeit entstand am Institut für Simulation und Graphik (ISG) der Fakultät für Informatik der Otto-von-Guericke-Universität Magdeburg. Hier bot sich mir eine Arbeitsumgebung, in der ich viel Inspiration und Unterstützung erfahren habe. Dafür danke ich vor allem Prof. Dr. Knut Hartmann und Prof. Dr. Holger Theisel. Doch meine Ambitionen, in diesem interessanten interdisziplinären Fachgebiet Musikinformatik zu arbeiten, wären vielleicht schon im Keim erstickt, gäbe es da nicht die Lehrer, Betreuer und Projektpartner, die sich begeistern ließen und mich seit meinen ersten Projekten im Studium unterstützen: Prof. Dr. Niels Knolle, Prof. Dr. Andreas Nürnberger, Prof. Dr. Maic Masuch, Dr. Ian Pitt und Dr. Carsten Lange.

Ich danke allen, mit denen ich im Laufe der Zeit »gemeinsame Sache« machen durfte: Sebastian Stober, Lars Stockmann, Daniel Büchner, Sven Heinicke, Judith Trappe, Jochen Miksch und all die Studenten von den Computermusik-Seminaren. Die gemeinsamen Projekte waren oft ein Quell neuer Ideen. Ganz besonderer Dank dafür gebührt Dr. Niklas Röber, dessen erbaulicher Zuspruch mir auch durch die schwierigen Phasen geholfen hat. Erbaulich und immer wieder inspirierend war auch die Atmosphäre in der Arbeitsgruppe *Visual Computing* und mit den Kollegen, mit denen ich im Laufe der Zeit das Büro teilte. Unter diesen möchte ich insbesondere meinen Korrekturlesern Dr. Christian Rössl, Marc Mörig, Janick Martinez Esturo, Alexander Kuhn und Dirk Joachim Lehmann danken. In Bezug auf alle bislang benannten Punkte gilt mein ganz besonderer Dank Tilo Hähnel, meinem Mitstreiter bei der Erforschung der ausdrucksvollen Interpretation von Musik. Die gemeinsame Arbeit markiert die produktivsten und inspirierendsten zweieinhalb Jahre, die man sich nur wünschen kann.

Große Unterstützung habe ich außerdem vom Sekretariat und den Technikern des ISG erhalten. Auch hierfür bin ich herzlich dankbar. Weiterhin danke ich Dr. Eamon Kiernan, dessen hervorragendes englisches Lektorat mir bei vielen meiner Veröffentlichungen eine unschätzbare Hilfe war. Schließlich gebührt mein herzlicher Dank auch meiner Familie, die mir stets den nötigen Rückhalt gegeben und mich voll unterstützt hat.

Veröffentlichungen

Wissenschaftliche Schriften und Vorträge

- Berndt, A. (2006), Repräsentation musikalischer Metainformationen, Diploma thesis, Otto-von-Guericke University, Department of Simulation and Graphics, Magdeburg, Germany.
- Berndt, A., Hartmann, K., Röber, N. & Masuch, M. (2006), Composition and Arrangement Techniques for Music in Interactive Immersive Environments, *in* 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute/Sonic Studio, Piteå, Sweden, pp. 53–59.
- Berndt, A. & Hartmann, K. (2007a), Audio-Interactive Counterpoint, *in* 'Sound, Music and the Moving Image', Institute of Musical Research, University of London, England. invited talk.
- Berndt, A. & Hartmann, K. (2007b), Strategies for Narrative and Adaptive Game Scoring, *in* 'Audio Mostly 2007: 2nd Conf. on Interaction with Sound', Fraunhofer IDMT, Interactive Institute/Sonic Studio Piteå, Ilmenau, Germany, pp. 141–147.
- Hartmann, K., Büchner, D., Berndt, A., Nürnberger, A. & Lange, C. (2007), Interactive Data Mining & Machine Learning Techniques for Musicology, *in* 'Conf. on Interdisciplinary Musicology 2007 (CIM07)', Tallinn, Estonia.
- Berndt, A. & Hartmann, K. (2008), The Functions of Music in Interactive Media, *in* U. Spierling & N. Szilas, eds, 'Interactive Digital Storytelling (ICIDS) 2008', Springer, Erfurt, Germany, pp. 126–131. LNCS 5334.
- Berndt, A. & Theisel, H. (2008), Adaptive Musical Expression from Automatic Realtime Orchestration and Performance, *in* U. Spierling & N. Szilas, eds, 'Interactive Digital Storytelling (ICIDS) 2008', Springer, Erfurt, Germany, pp. 132–143. LNCS 5334.
- Stockmann, L., Berndt, A. & Röber, N. (2008), A Musical Instrument based on Interactive Sonification Techniques, *in* 'Audio Mostly 2008: 3rd Conf. on Interaction with Sound', Interactive Institute/Sonic Studio Piteå, Piteå, Sweden, pp. 72–79.
- Berndt, A. (2009), Musical Nonlinearity in Interactive Narrative Environments, *in* G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 355–358.

- Berndt, A. & Hähnel, T. (2009a), *Der Computer als Interpret*, in 'Gespräche um Telemann', Zentrum für Telemannpflege und -Forschung, Magdeburg, Germany. invited talk.
- Berndt, A. & Hähnel, T. (2009b), *Expressive Musical Timing*, in 'Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 9–16.
- Berndt, A. (2010), *Decentralizing Music, Its Performance, and Processing*, in M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 381–388.
- Berndt, A. & Hähnel, T. (2010), *Modelling Musical Dynamics*, in 'Audio Mostly 2010: 5th Conf. on Interaction with Sound—Sound and Design', Interactive Institute/Sonic Studio Piteå, ACM, Piteå, Sweden, pp. 134–141.
- Hähnel, T. & Berndt, A. (2010a), *Expressive Articulation for Synthetic Music Performances*, in 'Proc. of New Interfaces for Musical Expression (NIME) 2010', University of Technology Sydney, Sydney, Australia, pp. 277–282.
- Hähnel, T. & Berndt, A. (2010b), *Synthetic and Pseudo-Synthetic Music Performances: An Evaluation*, in 'Proc. of the 3rd Int. Conf. of Students of Systematic Musicology (SysMus10)', University of Cambridge, Cambridge, UK.
- Berndt, A. (2011a), *Diegetic Music: New Interactive Experiences*, in M. Grimshaw, ed., 'Game Sound Technology and Player Interaction: Concepts and Developments', IGI Global, Hershey, PA, pp. 60–76.
- Berndt, A. (2011b), *Musical Tempo Curves*, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, University of Huddersfield, Huddersfield, UK.
- Hähnel, T. & Berndt, A. (2011a), *Eighth-Notes Performances: Kinds of Inégalité*, in 'Audio Mostly 2011: 6th Conf. on Interaction with Sound—Sound and Context', University of Coimbra, Interactive Institute/Sonic Studio Piteå, ACM, Coimbra, Portugal.
- Hähnel, T. & Berndt, A. (2011b), *Studying Interdependencies in Music Performance: An Interactive Tool*, in 'Proc. of New Interfaces for Musical Expression (NIME) 2011', University of Oslo, Norwegian Academy of Music, Oslo, Norway, pp. 48–51.

Notenveröffentlichungen

Berndt, A. (2008), *Liturgie für Bläser*, 2nd edn, Musikverlag Bruno Uetz, Halberstadt, Germany. transl.: Liturgy for Brass.

Berndt, A. (2012), Nun sich der Tag geendet hat, *in* F. Plewka, M. Schmeiß, T. Sterzig & J. Häußler, eds, 'Mitteldeutsches Bläserheft'. accepted and in press.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Problem: Musik ist linear	3
1.2	Problem: Musik ist endlich	5
1.3	Zielstellung und Aufbau der Arbeit	6
2	Geschichte und Funktion von Musik in interaktiven Medien	7
2.1	Vorgeschichte bis zur Filmmusik	8
2.2	Entwicklung der Musik in interaktiven Medien	11
2.3	Narrative Funktionen der Musik	15
2.3.1	Das filmische Erbe	15
2.3.2	Einbeziehung der Interaktivität	19
3	Verwandte Arbeiten	27
3.1	Sequentielles Arrangement	27
3.2	Paralleles Arrangement	30
3.3	Ausdrucksvolle Interpretation	32
3.4	Reharmonisierung	35
3.5	Komponierte Überleitung	37
3.6	Generative Musik	38
3.7	Zusammenfassung	39
3.8	Einordnung der weiteren Beiträge dieser Arbeit	40

4 Die Musik-Engine als Echtzeit-Arrangeur	43
4.1 Verteilte Musik	44
4.2 Komposition eines interaktiven Musikübergangs	47
4.3 Musik, nicht Physik	48
4.4 Bewegliche Soundquellen	53
4.5 MIDI-Portierung	54
4.6 Experimente und Diskussion	55
4.6.1 Kompositionsstudien	55
4.6.2 Demonstrationsszenarien	58
5 Modellierung ausdrucksvoller Interpretationen	61
5.1 Allgemeine Bemerkungen zur Interpretationsanalyse	64
5.2 Klang	69
5.3 Timing	70
5.3.1 Makro-Timing	72
5.3.2 Mikro-Timing	83
5.4 Dynamik	90
5.4.1 Makro-Dynamik	92
5.4.2 Mikro-Dynamik	100
5.5 Artikulation	106
5.6 Diskussion	115
5.7 Zusammenfassung	119
6 Das reaktive Orchester	121
6.1 Engine-Architektur	125
6.2 Überleitung	132
6.2.1 Strukturanalyse	133
6.2.2 Generiere Überleitung	134
6.3 Ergänzungen	142
6.4 Diskussion	143

7 Musik als Medium zur direkten Interaktion	151
7.1 Szenenmusik	153
7.1.1 Funktionen	154
7.1.2 Design-Prinzipien	155
7.1.3 Nichtlinearität und Interaktivität	156
7.2 Visualisierte Musik	159
7.2.1 Musik als Diegese	159
7.2.2 Interaktivität in der musikalischen Domäne	161
8 Schlussbemerkungen	167
A Narrative Funktionen der Musik: Ergänzung	171
A.1 Nach Zofia Lissa	171
A.2 Nach Johnny Wingstedt	178
B <i>iMuse</i> im Überblick	183
C Stilistischer Pluralismus	187
D Performance-Übungen	189
E Tempostudie	193
Literaturverzeichnis	195

Einleitung

MUSIK ist eine vergängliche Erscheinung. Sie existiert nur so lange, wie eine oder mehrere Schallquellen das sie umgebende Übertragungsmedium in Schwingung versetzen. Unser auditiver Sinn nimmt diese Schwingungen wahr und interpretiert sie als Geräusche, Klänge, Töne, Rhythmen, Melodien, Akkorde, ästhetische Strukturen, künstlerische Ideen, Kompositionen, Musik. Die Aufführung eines Musikstücks war lange Zeit ein einmaliges Erlebnis, das sich nicht so einfach wiederholen ließ, schon gar nicht ohne Interpreten. Selbst dann konnte jede Aufführung anders sein – andere Interpreten oder Interpretationen, andere Instrumente mit anderen Klängen und andere Akustik. Aus dieser Vergänglichkeit erwächst auch ihre Lebendigkeit. Musik erklingt nicht von allein, sie muss zum Klingen gebracht werden und kann sich bei jeder Aufführung erneuern. Uns Hörern bleibt dann nur die kurze Zeit ihrer Aufführung, um aufzupassen, hinzuhören und zu erleben.

Als im 18. und 19. Jahrhundert mit den Notenrollen, Wachs- und Stiftpapieren und Schallplatten die ersten Tonträger auftauchten, konnte das Vergängliche endlich festgehalten und das Hörerlebnis beliebig oft wiederholt werden. Der Klang, die Interpretation, die Musik war nun konserviert und blieb unveränderlich. Es war der Beginn einer Erfolgsgeschichte, welche eine riesige weltumspannende Industrie, die Musikindustrie, begründete. In unserer heutigen Medienkultur nimmt diese *konservierte Musik* einen nicht zu unterschätzenden Stellenwert ein. Unsere Hörerfahrung und musikalische Erlebniswelt wird von ihr dominiert und geprägt.

Doch hält unsere heutige Medienkultur auch einen medialen Modus bereit, dem die konservierte Musik in ihrer Unveränderlichkeit nicht gewachsen ist: die Interaktivität, die bidirektionale Kommunikation und das aufeinander bezogene Handeln zweier oder mehrerer Parteien. Ist die Rede von *interaktiven Medien*, so ist wenigstens eine dieser Parteien eben jenes Medium, das nach Goertz (2004)

besser als »Medienanwendung«¹ bezeichnet werden sollte. Ihre Inhalte, wie etwa Text, Bild, Animation, Geräusch, Sprache und Musik, bezeichnen Kühhirt & Rittermann (2007) als »Medienobjekte«². »Eine herausragende Eigenschaft objektbasierter [audiovisueller] Anwendungen ist die Möglichkeit, das Erscheinen und das Verhalten ihrer Inhalte durch Interaktionen in vielfältiger Weise beeinflussen zu können.«³ Trifft das auch auf Musik zu?

Auf der visuellen Ebene ist diese Interaktivität längst Normalität. Mit Bildelementen kann selbstverständlich interagiert werden, auch außerhalb von Bildbearbeitungsprogrammen. Natürlich können die Bildelemente auch ein eigenes komplexes Verhalten aufweisen, etwa im Rahmen einer Simulation, das durch Interaktion beeinflusst werden kann. In musikalischer Hinsicht kann davon heute nicht die Rede sein. Hier kommt fast ausschließlich die konservierte Musik zum Einsatz, vorgefertigte Aufnahmen, die feststehen und durch den Anwender kaum weiter beeinflusst werden können, außer in speziellen Musikbearbeitungsprogrammen. Wie groß die Diskrepanz zur visuellen Interaktivität ist, wird deutlich, wenn Musik als Teil eines audiovisuellen Werkes in Verbindung mit der visuellen Ebene tritt. Hier ist als eines der anschaulichsten, vielseitigsten und dadurch auch eines der herausforderndsten Beispiele das Video- und Computerspiel zu nennen, das sich in der Ästhetik seiner Präsentation häufig am Film orientiert.

Vom Film jedoch wissen wir, dass Musik eine sehr organische Verbindung mit den anderen Inhalten eingehen und eine *dialektische Einheit* bilden kann. Aber der Film ist linear, d.h. nicht-interaktiv. Ihm kann auch die lineare konservierte Musik beigeordnet werden. Das Computerspiel hingegen ist interaktiv. Es wartet nicht auf die Musik und richtet sich auch sonst nicht nach ihr; nein, die Musik muss folgen, muss reagieren, erst recht, wenn sie vielleicht sogar selbst das Objekt ist, auf welches die Interaktion abzielt. Hier muss die Musik zwei widerstreitenden Kriterien gerecht werden. Zum einen muss sie verharren können. Solange keine Interaktion geschieht, die eine musikalische Veränderung auslöst, muss auch die Musik bei dem gerade geforderten Ausdruckscharakter bleiben. Exakte Wiederholungen ein und des selben Musikstücks sind jedoch wegen ihrer statischen die Glaubwürdigkeit der Inszenierung untergrabenden Wirkung unerwünscht (Sanger 2003, Marks 2008, Collins 2008). Deshalb wird die eigenständige Variabilität von sich wiederholendem Material (Variieren von Endloschleifenmusik) gefordert. Zum anderen wird der Musik eine zeitnahe und musikalisch schlüssige Wandelbarkeit als Reaktion auf äußere Einflüsse, wie etwa Benutzerinteraktionen, abverlangt. Zur Filmmusik schrieb die Musiktheoretikerin Zofia Lissa (1965): »Die raschen Veränderungen des Inhaltes der Szene

¹Goertz (2004), Seite 106.

²Kühhirt & Rittermann (2007), Seite 16.

³Kühhirt & Rittermann (2007), Seite 17.

zwingen zu parallelen Veränderungen der Musik, und zwar sowohl in Klangcharakter, ihren Funktionen, der Instrumentierung als auch zuweilen sogar des Stils.«⁴

Die Beiordnung musikalischer zu inhaltlichen Elementen setzt eben diese Wandlungsfähigkeiten der Musik voraus und ist heute kaum organisch zu nennen. Die Musik tritt »in kleineren und größeren Fragmenten zutage, [die] nicht selten mitten in einer Phrase [abreißen].«⁵ Dies schrieb Lissa zur frühen Filmmusik, die durch geeignete Kompositions- und Arrangementstechniken dieses Defizit heute überwunden hat. Die Parallele zur Musik in interaktiven Medien heute ist aber augenfällig. Auch hier steht sie oft fragmentarisch und zusammenhanglos in einem indifferenten Verhältnis zu den übrigen Inhalten, was die folgenden Abschnitte noch detaillierter besprechen werden. Die durch ihre Konservierung verlorene Wandelbarkeit und Lebendigkeit wird ihr im Kontext sich ständig wandelnder interaktiver Medien zum Verhängnis. Diese Arbeit soll dazu beitragen, dieses Defizit zu überwinden. Es werden Lösungen vorgestellt, durch die nichtlineare Aspekte in musikalische Daten eingearbeitet werden, so dass anstelle feststehender unveränderlicher Musikstücke ein ganzer Raum an musikalischen Möglichkeiten aufgespannt wird, in welchem interaktive Einflussnahme in vielfältiger Weise möglich wird.

1.1 Problem: Musik ist linear

Die Vertonung interaktiver Medien wird heute von der konservierten Musik, den vorgefertigten und weitgehend feststehenden Musikaufnahmen, dominiert. Jede Note ist komponiert, jeder Ton von Interpreten eingespielt, Abmischung und akustische Effekte vom Tontechniker eingearbeitet. Die Musik ist bereit zur Wiedergabe, das Bild quasi fertig gezeichnet und bereit, angezeigt zu werden.

Nachträgliche Veränderungen, gar während der Wiedergabe, sind nur noch in einem sehr engen Rahmen möglich. Die Wiedergabe kann unterbrochen werden. Die Lautstärke kann erhöht oder verringert werden. Bis zu einem gewissen Grad lassen sich auch Tonhöhe und Abspielgeschwindigkeit variieren, bevor unvermeidlich Verfremdungseffekte deutlich wahrnehmbar werden. Ein Flüstern wird eben nicht zum Schreien nur durch Erhöhung der Amplitude. Ein tiefer Ton klingt nicht wie ein hoher, eine rasche Tonfolge nicht wie eine lang ausgespielte.

⁴Lissa (1965), Seite 45.

⁵Lissa (1965), Seite 21.

Obertonspektrum und Hüllkurve⁶ eines jeden Tones spiegeln die kleinsten Variationen in der Spielweise wieder und prägen das klangliche Resultat in eigentümlicher Weise. Selbst der Kontext, in dem ein Ton gespielt wird, beeinflusst seine Klangeigenschaften. Ein eigentlich leicht zu singender Ton kann angestrengt und gepresst klingen, wenn er ausgehend vom vorhergehenden Ton unangenehm liegt. Selbst das Spiel der Mitmusiker beeinflusst jeden einzelnen von ihnen und sein Spiel.

Soll dies alles berücksichtigt werden, wird der signalverarbeiterische Aufwand nicht nur enorm, er gerät an die Grenzen des derzeit Möglichen. Man bedenke, wie schwer und unzufriedenstellend es bis heute ist, auch nur einzelne Ton-Anfänge (Onsets) in einer Aufnahme, d.h. einem komplexen Wellensignal, zu detektieren. Ton-Enden automatisch zu finden, ist ungleich schwerer und selbst unter günstigen Studiobedingungen nur unzureichend möglich. Einen Überblick geben Bello, Daudet, Abdallah, Duxbury, Davies & Sandler (2005).

In Anbetracht all dieser Schwierigkeiten scheinen der *harte* und *weiche Schnitt* für den Großteil der Entwickler interaktiver Medien, insbesondere von Computerspielen, immer noch am unkompliziertesten. Beim *harten Schnitt* wird die Wiedergabe des gerade laufenden Musikstücks einfach abgebrochen und das nächste Musikstück gestartet. Das entspricht dem Skippen am CD-Player. Es entsteht ein abrupter, zur Musik asynchroner und aus musikalischer Sicht äußerst bedenklicher Bruch, den bereits Zofia Lissa (1965) anmahnte und der den Spielemusikkomponisten ein steter Dorn im Auge ist (Sanger 2003, Marks 2008, Collins 2008). Um diese unmusikalische Abruptheit abzuschwächen, wird beim *weichen Schnitt* eine kurze Kreuzblende (Cross-Fade) durchgeführt. Sie geschieht immer noch asynchron, wobei im Allgemeinen musikalisch unpassende klangliche und rhythmische Überlagerungen dort entstehen, wo das alte Musikstück ausblendet und das neue einblendet, beide also gleichzeitig erklingen.

Alle diese Nebeneffekte sind nicht generell unerwünscht. Wenn Abruptheit und Überraschung gewollt sind, kommt der harte Schnitt selbst in modernen Filmen zum Einsatz, dies aber eben mit der entsprechenden ausdrucksmäßigen Absicht. In jedem Kunstwerk dient alles Formale und Konstruierte zugleich auch einer inhaltlichen Aussage oder Absicht. Nichts wird ohne Grund gesagt, gezeigt oder getan. Als reguläre Mittel zum Musikwechsel sind harter und weicher Schnitt aber eben nicht genügend im musikalischen Zusammenhang verankert, daher sehr besonders und auffällig. Wenn ein solcher Musikwechsel dadurch allzu offensichtlich erzwungen ist, anstatt als schlüssige Konsequenz organisch hervorzugehen, überzeugt er nicht mehr. Das gilt insbesondere dann, wenn der

⁶Hüllkurven beschreiben die lautstärkemäßige Formung der Töne vom Tastenanschlag bis zum Loslassen der Taste und, wenn der Ton danach noch verklingt, auch darüber hinaus. Der Amplitudenverlauf wird üblicherweise in die vier Phasen *Attack*, *Decay*, *Sustain* und *Release* (ADSR) unterteilt (Bernstein & Cooper 1976, Roads 1996, Boulanger 2000).

damit ausgedrückte inhaltliche Wandel eigentlich fließend verläuft (z.B. das Betreten einer anderen Region in einer kontinuierlichen Spielwelt). Harter und weicher Schnitt sind Ausrufezeichen, die auf das Konstruierte erst recht hinweisen, Glaubwürdigkeit und Nachvollziehbarkeit durch Inkonsistenz und Inkohärenz zerstört und nur zum Versagen der ursprünglichen Absicht beitragen.

1.2 Problem: Musik ist endlich

Ein weiteres Problem der vorgefertigten statischen Musik ist ihre feste zeitliche Länge, oder in diesem Fall Kürze. Ein Beispiel: In einem virtuellen Haus mag jedes Stockwerk oder sogar jeder Raum eine eigene Hintergrundmusik besitzen. Besucht der Spieler einen Raum, erklingt die zugehörige Musik. Bleibt er sehr lange in diesem Raum, ist bald das Ende des Musikstücks erreicht. Üblicherweise beginnt es nun unverändert von vorn, was vom Hörer auf Dauer als langweilig und sogar störend empfunden wird. Collins (2009) beschreibt, dass die Musik dann sogar manuell vom Spieler im Einstellungs Menü ausgeschaltet wird.

In interaktiven Medien ist es allerdings nahezu unmöglich, die Länge einer Szene und ihren Fortgang vorauszusagen. Dementsprechend unmöglich ist es, Länge und Struktur der dazugehörigen Musik entsprechend vorzubereiten. Das gleiche Problem ist auch auf der Sprachausgabe- und Klangeffekt-Ebene bekannt. Ein virtueller Charakter, der dem Spieler immer die gleiche Antwort gibt (z.B. die Spielfigur, die nicht ausführbare Aktionen immer mit »Das geht nicht.« quittiert), verliert seine Persönlichkeit-suggestierende Kraft und verkommt zum mechanischen Gerät. Auch immer gleich klingende Klangeffekte und Geräuschkulissen verlieren mit der Zeit nicht nur an Faszination sondern an Glaubwürdigkeit. Das ist einer der Hauptgründe für Farnell (2007), einen wesentlich flexibleren Geräuschsinteseansatz auszuarbeiten.

Wenn dem Hörer nach einiger Zeit des immer gleichen dies bewusst wird, büßt die Inszenierung einer konsistenten virtuellen Welt ihre Überzeugungskraft ein. In gewisser Hinsicht ist damit das Ende der Welt erreicht. So gelungen auch ihre graphische Inszenierung sein mag, auf auditiver Ebene ist der Blick hinter die hohlen Kulissen freigegeben. Die Immersion, das sich hineinversetzt Fühlen in eine Welt oder einen Charakter, geht in diesem Moment verloren. Der Spieler wird zum Bediener.

1.3 Zielstellung und Aufbau der Arbeit

Beide Problemfelder (Linearität und Endlichkeit der Musik) werden in dieser Arbeit behandelt, wobei der Schwerpunkt auf dem Linearitätsproblem liegt, d.h. auf der kurzfristigen schlüssigen Wandelbarkeit des musikalischen Ausdruckscharakters. Die dafür entwickelten Lösungen eröffnen auch neue Perspektiven für das Variieren von Endlosschleifenmusik.

Dem soll jedoch zunächst eine medientheoretische Aufarbeitung vorausgehen, in welcher die Rolle der Musik in interaktiven Medien, speziell in Video- und Computerspielen, untersucht wird (Kapitel 2). Die Korrelation mit dem entwicklungsgeschichtlichen Vorgänger, der Filmmusik, birgt wertvolles Erkenntnispotential. Aus den Betrachtungen folgen unter anderem Rückschlüsse auf die ausdrucksmäßige Bandbreite und das notwendige Maß an musikalischer Flexibilität.

Obwohl Aspekte der musikalischen Nichtlinearität und Interaktivität in der Musikinformatik kein unbekanntes Thema sind, gibt es noch immer keine theoretische Aufarbeitung oder Systematisierung existierender Lösungsansätze und noch offener Lösungsmöglichkeiten. Diese Grundlagenarbeit leistet Kapitel 3 und setzt damit den Rahmen, in welchen sich die im weiteren Verlauf dieser Arbeit entwickelten Lösungsansätze eingliedern.

Diese loten insbesondere das Potential aus, das mit den Mitteln des Arrangements (Kapitel 4) und der ausdrucksvollen Interpretation (Kapitel 6) möglich ist. Hier ist allein auf Grundlage traditioneller Musikformate nicht viel möglich, so dass neben den eigentlichen algorithmischen Lösungen auch Modellierungsaspekte eine wichtige Rolle spielen. Wie beschreibt man bspw. eine mehrdimensionale interaktive Musikszene? Auch die Modellierung ausdrucksvoller Interpretationen (Kapitel 5) ist ein ausführlich behandelter Themenkomplex, der nicht nur als Grundlage für die späteren sehr differenzierten Musikbearbeitungsmechanismen bedeutsam ist. Interessant ist das auch für die Gebiete Music Information Retrieval (Feature Description) und Musikwissenschaft (Interpretationsforschung).

Kapitel 7 geht schließlich den Weg hin zur interaktiven Musik konsequent weiter und beschreibt, wie Musik zum Medium für direkte Interaktionen werden kann. Dafür wird das Phänomen der *diegetischen Musik*, also der »Musik im Bild«, näher untersucht und für interaktive Medien systematisiert. Es werden Design-Prinzipien erarbeitet und die Vielfalt an Möglichkeiten zur Nichtlinearität und Interaktivität aufgezeigt. Das Kapitel zeigt einige der vielfältigen Anwendungsmöglichkeiten für die hier entwickelten aber auch bereits bestehende Techniken auf und gibt Motivationen für künftige Entwicklungen. Kapitel 8 schließt diese Arbeit.

Geschichte und Funktion von Musik in interaktiven Medien

MUSIK spielt in audiovisuellen Medien eine außerordentlich wichtige Rolle. Kaum ein Film kommt ohne sie aus. Selbst in den kürzesten Werbespots ist Musik als verstärkendes, die Werbebotschaft unterstreichendes Mittel unerlässlich. Auch in interaktiven Medien ist sie omnipräsent. Schon im schlichsten DVD-Menü darf die auf den Film einstimmende Hintergrundmusik nicht fehlen, ganz zu schweigen von professionell und teuer produzierten Video- und Computerspielen, deren Budgets denen von Hollywood-Filmen in nichts nachstehen.

Aber die Musik steht selten im Vordergrund. Das Rampenlicht der bewussten Wahrnehmung gilt zumeist Bild und Sprache, während Geräuschkulisse und Musik im Hintergrund bleiben, unterbewusst wahrgenommen und verarbeitet werden. Jedoch besagt »die unter- und unbewusste Wahrnehmung [...] nichts über einen untergeordneten Stellenwert innerhalb der Dramaturgie. Je unbewusster Musik wirkt, desto mehr kann sie den Bildbetrachter in einem vom Filmemacher gewünschten Sinne konditionieren und seine Rezeption des Bildes stimulierend lenken. Die suggestive Kraft, die auf diese Weise von der Musik ausgeht, kann den Zuschauer für emotionale Gehalte des Filmbildes sensibilisieren und ihn für die Wahrnehmung der analogen Bildgehalte [...] öffnen«¹, beschreibt der Linguist und Musikwissenschaftler Norbert J. Schneider (1990). Die Musik soll unterbewusst wahrgenommen werden, um vorbei am rationalen Denken direkt im limbischen System die ihr innewohnende emotionale Ladung zu entfalten (Kungel 2004, Brown, Martinez & Parsons 2004).

Zudem wird durch die Kombination mit den anderen Medien ein Assoziationsraum aufgespannt, durch den Musik zum Symbolträger werden kann. Schneider (1990) spricht in diesem Zusammenhang von einem »Semantisierungspro-

¹Schneider (1990), Seite 72.

zess«². Ein musikalisches Element kann durch die Montage mit anderen medialen Schichten (wie etwa der Bildebene) zum Symbol für außermusikalische Sachverhalte werden. Ein berühmtes Beispiel ist der *Imperial March* von John Williams (1977) und seine Assoziation mit der Filmfigur *Darth Vader*. »Musik wird stets mit ›Bedeutung‹, d.h. mit einem kommunikativen Inhalt, eingesetzt«³. Dieses Kapitel soll der Rolle, welche Musik speziell in audiovisuellen Medien wie dem Film einnimmt, nachgehen und dabei eine Übertragung auf interaktive Medien vornehmen. Dem liegen folgende Veröffentlichungen zu Grunde:

- (Berndt & Hartmann 2007a),
- (Berndt & Hartmann 2008).

Ein geschichtlicher Abriss der Entwicklung von Filmmusik (Abschnitt 2.1) lässt in der Gegenüberstellung mit der vergleichsweise noch kurzen Entwicklungsgeschichte von Musik in interaktiven Medien, speziell Computerspielen, einige Parallelen erkennen (Abschnitt 2.2). Das erlaubt einen Ausblick auf ausstehende Entwicklungen und künftige Tendenzen für die Musik in interaktiven Medien. Hier fügen sich die weiteren Beiträge dieser Arbeit in ein entwicklungsgeschichtliches Gesamtbild ein. Abschnitt 2.3 wird dies konsequent fortsetzen, das medientheoretische Potenzial von Musik in interaktiven Medien über den aktuellen Stand in Forschung und Praxis hinaus ausarbeiten und ist damit zugleich Hintergrund und Motivation für diese Arbeit.

2.1 Vorgeschichte bis zur Filmmusik

Die Verbindung der Musik mit anderen Medien hat eine Tradition, die so alt ist wie die Musik selbst. Zusammen mit Text und Sprache ergibt sich Gesang. Im griechischen Drama kommt die Musik mit dem Bühnenbild und einer dramatischen Handlung zusammen. In Tanz und Ballett wird sie zum Regulator für Bewegung und Geste. Der Schritt von den szenischen Bühnenwerken der Renaissance zur Oper verwebt die Musik noch enger mit Handlung und Text. »Die Verbindung [...] mit der dramatischen Handlung vollzog sich also bereits auf dem Boden des Theaters.«⁴ Selbst die barocke Kirchenmusik nahm mit Formen wie dem Oratorium und der Historie an diesem Entwicklungsprozess teil. Die *Weihnachtshistorie* von Heinrich Schütz (1664) verstünde sich heute vielleicht als »Soundtrack zum Krippenspiel«. Die Musik ist hier gegenüber den anderen

²Schneider (1990), Seite 84.

³Schneider (1990), Seite 80.

⁴Lissa (1965), Seite 19.

die dominierende Kunstart. In zunehmendem Maße beschränkt sie sich nicht mehr nur darauf, textuelle Inhalte zu illustrieren, sondern reflektiert auch über sie. Der Komponist wird zum Kommentator, zum Prediger, der die Worte mit musikalischen Mitteln auslegt.

»In den Dramen, den Theaterstücken spielt die Musik eine weniger organische Rolle, aber auch hier hat sie ihre künstlerischen Aufgaben.«⁵ Sie bereitet den Zuschauer auf die emotionale Stimmung einer Szene vor oder fasst diese zusammen. Sie begleitet Monologe oder Dialoge und erfüllt interpunktive Aufgaben.

Mit dem Theater ist der Film verwandt. In beiden dominiert die visuelle Schicht, das Bild. Auch Geräusch und Dialog gehören zu den bewusst rezipierten Elementen, denen sich die Musik in Struktur und Anlage im Allgemeinen unterzuordnen hat. Die ersten Beispiele für Musik im Film (genauer: im Stummfilm) sind jedoch Zeugnisse dafür, dass im Theater bereits etablierte Traditionen und Konventionen für den Film noch einmal neu entdeckt werden mussten. Die Musiker – meist Pianisten, in luxuriösen Kinos auch Orchester – hatten in erster Linie die Aufgabe, störende Nebengeräusche von Projektor und Umwelt zu maskieren (Adorno & Eisler 1947a, Lissa 1965).

Zunächst stand den Interpreten sogar die Wahl der Begleitmusiken völlig frei. Eine zum konkreten Film gehörende spezielle Komposition lag selten vor. Die Musikstücke sollten zwar möglichst zu den jeweiligen Situationen passen, waren aber in hohem Maße von Geschmack und Laune des Pianisten und dem ihm verfügbaren Repertoire an Notenliteratur und Darstellungsmöglichkeiten abhängig. »Klassische Werke standen damals gleich neben banalen Schlagern oder modernen Tänzen, sie traten in kleineren und größeren Fragmenten zutage, rissen nicht selten mitten in einer Phrase ab.«⁶

Hier zeigte sich, dass die Musikstücke, obgleich sie eine mehr oder minder zur Szene passende Stimmung vermitteln mochten, doch nicht für die Kombination mit dem Medium Film geeignet waren. »Bald jedoch wurde man sich dessen bewusst, dass zwischen der musikalischen Begleitung des Films und dessen Ablauf weder ein Antagonismus von Charakter und Ausdruck eintreten noch auch nur ein indifferentes Verhältnis herrschen durfte.«⁷ Ein erster Versuch, dieser Forderung nachzukommen, war die Anlage von speziellen Musikarchiven, welche die Musikauswahl mit einem umfangreichen Bestand an ausdrucksmäßig oder programmatisch klassifizierten Charakterstücken unterstützten. Als erstes dieser Musikarchive entstand 1913 das *Sam Fox Moving Picture Music*. Solche musikalischen Kinotheken werden auch heute noch von budgetschwachen TV-Produktionen, Serien und Dokumentarfilmen ausgiebig genutzt (Kungel 2004).

⁵Lissa (1965), Seite 19.

⁶Lissa (1965), Seite 21.

⁷Lissa (1965), Seite 21.

Wurde Musik speziell zum Film komponiert, lag sie dem Filmmaterial in gedruckter Form (Cue Sheets) oder als Grammophonplatten bei. Die Synchronisation war auch hier noch sehr problematisch. Erst der Tonfilm, bei dem Bild und Ton auf das gleiche Band aufgezeichnet werden, legte die technischen Grundlagen für ihre volle Synchronisierung (Lissa 1965).

Der Tonfilm brachte auch die Ebene der Sprache und Geräusche wieder ins Spiel. Die Musik, die bislang mitunter auch die Aufgabe hatte, der Geisterhaftigkeit und Distanziertheit der stummen Bilder (wie es Adorno & Eisler (1947a) beschreiben) Abhilfe zu verschaffen, konnte nun selektiver und dramaturgisch motivierter eingesetzt werden. Sie musste nun nicht mehr über die komplette Laufzeit des Films hinweg spielen. Ihre dramaturgische Bedeutung für das Gesamtkunstwerk stieg, je mehr sie aus dem Fokus der bewussten in die unterbewusste Wahrnehmung verdrängt wurde.

Verschiedene Möglichkeiten der Kombination von Film und Musik haben überdies zu verschiedenen Ausprägungen geführt. Im Dokumentar- und Spielfilm sind Bild und Sprache die klar dominierenden Elemente. Die Musik tritt nur punktuell ins Rampenlicht, etwa an dramaturgisch wichtigen Stellen oder als erzählerisches Element. In Opernverfilmungen, Musikfilmen und Musikvideos ist hingegen die Musik die vorherrschende Schicht. Die zu hörenden Stücke sind im Allgemeinen autonome eigenständige Werke, was Konsequenzen für die Struktur der Erzähl- und der Bildebene nach sich zieht. Ihre sonst stärker ausgeprägte Sprunghaftigkeit muss sich hier der musikalischen Kontinuität anpassen. Ein ungefähres Gleichgewicht zwischen Bild und Musik lässt sich hauptsächlich bei Experimental- und Trickfilmen beobachten.

Stets kann die Musik dabei auf zweierlei Weise in Erscheinung treten, als diegetisches oder nicht-diegetisches Element. Diegetische Musik wird auch als »Musik im Bild« oder Szenenmusik bezeichnet. Sie bzw. ihre Klangquelle ist Teil der dargestellten Welt (z.B. ein Autoradio, ein Straßenmusikant, eine Musicbox) und innerhalb dieser wahrnehmbar. Nicht-diegetische Musik, auch bezeichnet als »Musik aus dem Off«, ist hingegen nur für den Zuschauer hörbar, vergleichbar mit der Erzählerstimme, die einen Film kommentiert aber selbst nicht vor Ort anwesend ist. Beide Erscheinungsformen der Musik kommen in Filmen und auch in Computerspielen vor. Während Kapitel 7 sich ausschließlich mit der diegetischen Musik befassen wird, legt dieses Kapitel den Fokus auf die nicht-diegetische Musik.

»Die Musik im Film hat demnach ihre Vorformen, ihre Vorgeschichte, ihre Geschichte und ihre Entwicklungsperspektiven. Sie hat auch ihre allgemein angenommenen Konventionen, [...] die sich auch verändern.«⁸ Für die Musik in in-

⁸Lissa (1965), Seite 23.

teraktiven Medien und besonders in Video- und Computerspielen zeichnet sich nun ein sehr ähnlicher Entwicklungsprozess ab.

2.2 Entwicklung der Musik in interaktiven Medien

Die unrealistischen, ja surrealen, Soundeffekte der frühen Videospiele sind zunächst ein Zugeständnis an die begrenzten technischen Möglichkeiten. Hierin spiegeln sich die Anfänge der Klangsynthese wider. So ist etwa das Piepsen im Spiel *Pong* (Bushnell & Alcorn 1972) kein echtes Geräusch, sondern ein einzelner Ton, denn ein klar definiertes Wellenmuster ist technisch einfacher zu erzeugen als die komplexen Frequenzstrukturen mancher Geräusche. Auch komplexere Tonstrukturen ersetzen das Geräusch. Macht die Spielfigur einen Sprung, erklingt dazu oft ein aufwärts glissierender Ton, der die Sprungbewegung tonmalerisch nachahmt. Sammelt die Spielfigur ein Power-Up⁹ ein, erklingt zur Belohnung ein kleiner motivhafter Einwurf.

Dieser Ersatz für echte – manchmal auch gar nicht existente – Soundeffekte hat zu einer ganz eigenen Geräuschästhetik geführt, die eher der Musik denn dem Sounddesign verwandt ist (Collins 2006a,b, 2007). Das führte in Einzelfällen sogar bis zur Verschmelzung mit der eigentlichen Hintergrundmusik. Im Spiel *Super Mario Bros.* (Nintendo EAD 1985) sind die Soundeffekte so konzipiert, dass sie sich in die Harmonik der Hintergrundmusik einfügen und damit zu einem Teil dieser werden.

Bereits 1979 wurde im Spiel *Asteroids* (Rains & Logg 1979) eine ähnliche Synthese versucht. Zu dem Grundmetrum, das von zwei alternierenden Tönen propagiert wird, fügen sich die Laser- und Explosionseffekte zu einer erstaunlich rhythmischen Collage. Die Umsetzung dieses Konzeptes wie auch generell aller frühen Spielemusiken wird noch sehr stark von technischen Gegebenheiten limitiert. Für *Asteroids* stand nur ein Synthesizer mit einem einzigen Audiokanal zur Verfügung. Das hatte zur Folge, dass immer dann, wenn ein Soundeffekt erklang, das Grundmetrum pausierte. Beide konnten nicht gleichzeitig erklingen.

Die Titelmusik des Spieleklassikers *Pac-Man* (Toru 1980) war noch einstimmig. Für die Vertonung von *Super Mario Bros.* auf dem *Nintendo Entertainment System* (kurz: NES) standen immerhin schon vier Kanäle zur Verfügung. In die CPU¹⁰ der NES sind insgesamt fünf Soundkanäle integriert, auf denen verschiedene Wellenmuster erzeugt werden können. Zwei Kanäle sind für die Erzeugung

⁹Power-Ups sind Objekte in der Spielwelt, bei deren Einsammeln der Spielfigur besondere Fähigkeiten, ein Extraleben o.ä. verliehen wird.

¹⁰Abkürzung für Central Processing Unit

von Rechteckwellenformen mit variabler Pulsbreite reserviert, ein Kanal für eine Dreieckwellenform, ein Rauschgenerator und ein weiterer Soundkanal, der für Soundeffekte reserviert ist. Sopran und Alt der Musik von *Super Mario Bros.* wurden auf den beiden Rechteckkanälen ausgegeben, der weichere Klang der Dreieckwellenform wurde für den Bass verwendet und der Rauschgenerator für die Perkussion.

In diesen Möglichkeiten (Rauschen und Grundwellenmuster wie Sinus, Rechteck und Dreieck) erschöpft sich auch schon die Klangvielfalt der meisten Systeme jener Zeit. Auch dynamische Differenzierungen (Lautstärkeunterschiede) und Hüllkurvencharakteristiken (der Amplitudenverlauf eines Tones) waren nicht möglich, was die Arbeit mit interpretatorischen Ausdrucksmitteln extrem limitierte.

Hinzu kommt, dass die frühen Spielemusiken selten aus der Feder professioneller Komponisten stammen, denn Nichtinformatikern erschloss sich dieses stark technisch geprägten Milieu nur selten. Es waren die Programmierer, die sich aus dem Schatz ihrer eigenen Hörerfahrungen und Musikvorlieben inspirieren ließen. Die Musiken waren häufig popularmusikalisch im Stile der jeweiligen Zeit geprägt und von eingängiger Melodik, Harmonik, Struktur und gefälligem Duktus bestimmt. Ende der 1980er Jahre wurde eine Reihe von Video- und Computerspielen veröffentlicht, die auf zeitgleich aktuellen Filmen beruhten. Hier ist ein erster Versuch zu erkennen, Prinzipien der Filmästhetik auf das interaktive Medium zu übertragen, darunter auch musikalische Prinzipien. Das geht aber kaum darüber hinaus, dass verschiedene Spielabschnitte (Levels) mit eigenen Themen aus der Filmvorlage unterlegt wurden. Hier dominiert nach wie vor der Unterhaltungszweck, nicht der dramaturgische Effekt.

Die Anforderungen an die musikalische Qualität stiegen mit den zunehmend differenzierteren Möglichkeiten der Soundkarten und -chips. Ende der 1980er und zu Beginn der 1990er Jahre war die AdLib-Soundkarte (AdLib Inc. 1987) zu großer Verbreitung gelangt. Ihr OPL2-Soundchip (Yamaha YM3812) ermöglichte u.a. eine größere klangliche Vielfalt auf 9 Kanälen, die Definition eigener Hüllkurven und 64 Lautstärkeabstufungen. Auch der sich immer weiter durchsetzende MIDI-Standard¹¹ ist der musikalischeren Qualität förderlich, findet er doch auf Heimcomputersystemen (u.a. Comodore C64, Comodore Amiga und PC) ebenso Anwendung wie in professionellen Tonstudios. Dort ist er auch heute noch nicht wegzudenken. Lediglich die auffallend synthetischen Instrumentenklänge, damals noch als qualitativer Nachteil empfunden, standen der immer professioneller werdenden musikalischen Vertonung nach. Heute sind die

¹¹MIDI ist die Abkürzung für *Musical Instrument Digital Interface* und ist ein Protokoll für die Übermittlung, Aufzeichnung und Wiedergabe von Steuerinformationen digitaler Musikinstrumente, das 1982 von der MIDI Manufacturers Association veröffentlicht wurde.

»alten Klänge« Assoziationsträger, werden bewusst eingesetzt und als synthetische Klänge erlebt, wie etwa im Stück *Gameboy Tune* von Tomáš Dvořák aus dem Soundtrack zum Spiel *Machinarium* (Amanita Design 2009).

Bereits die ersten beiden Spiele der Weltraum-Action-Serie *Wing Commander* (Origin Systems 1990, 1991) lassen nun eine deutlich an der Filmmusik orientierte und professioneller umgesetzte Vertonung erkennen, deren Einsatz dramaturgisch motiviert ist. Im Gegensatz zu solchen Action-orientierten Spielen dominiert bei Adventure-Spielen die Erzählung einer Geschichte und das Lösen von Rätseln; der Musik kommt hier mehr als in anderen Spielegenres jener Zeit eine narrative und immersive Funktion zu.

In den zahlreichen Adventures der Firma LucasArts¹² kommt zu diesem Zweck eine der ersten bedeutenden Musik-Engines zum Einsatz, die *Interactive Music Streaming Engine* (kurz: *iMuse*). Ihre Entwickler, Land & McConnell (1994), haben mit dieser Pionierarbeit versucht, jener Erkenntnis Rechnung zu tragen, die über ein halbes Jahrhundert zuvor bereits die Filmmusik lernte. »Eine Summe verschiedener, mehr oder weniger parallel nebeneinander verlaufender künstlerischer Erscheinungen ergibt noch kein einheitliches, organisches Kunstwerk. Notwendig ist ihre Synthese [...]. Die Beiordnung musikalischer Fragmente zum Bild war locker und rein äußerlich [...]. Die Sprünge von einem Fragment zum anderen kamen ohne Kadenz, Modulationen usw. zustande; sehr häufig entstand eine störende Diskrepanz zwischen Bild und Musik.«¹³ *iMuse* war ein erster Versuch, die musikalische Begleitung durch Wahrung einiger Grundsätze des Arrangements organischer dem Spielgeschehen beizuordnen, so wie es auch die Filmmusik einst lernte. Ein Überblick zu *iMuse* wird in Anhang B gegeben.

Mit dem Aufkommen des CD-ROM-Mediums machte die Musik in interaktiven Medien jedoch wieder einen Schritt zurück. Dadurch, dass nun mehr Speicherplatz verfügbar wurde, konnten speicherintensivere Audioformate wie MP3, Wave und natürlich auch CD-Audio-Tracks verwendet werden, die sehr hohe Klangqualitäten ermöglichen. Die Aufnahmen können sogar von menschlichen Musikern im Studio eingespielt werden, was lediglich aus Budgetgründen oft nicht gemacht wird. Alle diese Formate haben aber einen entscheidenden Nachteil. Die Aufnahmen sind unveränderlich statisch und können nur abgespielt oder unterbrochen werden. Ein Anpassen der Musiken an interaktive Geschehnisse etwa durch geschickte Arrangiertechniken ist bei dieser Unveränderlichkeit nicht möglich. So bleibt *iMuse* für lange Zeit der fortschrittlichste Versuch, dem indifferenten Verhältnis zwischen Spielgeschehen und Musikbegleitung Abhilfe zu verschaffen.

¹²LucasArts wurde 1982 von George Lucas gegründet und war bis 1991 als Lucasfilm Games firmiert.

¹³Lissa (1965), Seite 101.

Die 1990er Jahre sind auch die Zeit für eine besondere Ausprägung von Video- und Computerspielen, in denen die Musik eine ähnliche Bedeutung erhält wie in Musikfilm und Musikvideo. In *Loom* (Lucasfilm Games 1990) und *The Legend of Zelda – Ocarina of Time* (Nintendo EAD 1998) sind Töne, gespielt auf virtuellen Instrumenten, das Mittel der Interaktion mit der Spielwelt. Die gespielten Melodien können dabei auch motivisch in der eigentlichen Begleitmusik verarbeitet sein und so einen Assoziationsraum aufspannen, der ihnen narrative Bedeutung verleiht.

Die Spieleentwickler von Bemani (1999a,b,c) gingen noch einen anderen Weg und nutzten in Spielen wie *Guitar Freaks*, *DrumMania* und *HipHopMania* Nachbildungen echter Instrumente als Controller (Gitarre, Drum Kit, Scratch Turntable). In *Donkey Konga* (Namco 2003) ist das Eingabegerät in Form und Bedienungsweise Bongos nachempfunden. Konami's (1998) *Dance Dance Revolution* wird über eine druckempfindliche Matte gesteuert, auf welcher der Spieler zur Musik passende Tanzschritte ausführt. In dieser Art Spielen tritt die Musik in autonomer Form als sie selbst auf. Sie steht nicht in Bezug zu Spielinhalten; sie ist selbst Spielinhalt. Mit dieser speziellen Erscheinungsform von Musik als Element der Diegese (der virtuellen Welt) wird sich Kapitel 7 detailliert beschäftigen. Daher wird in diesem Kapitel nicht weiter darauf eingegangen.

Heute hat die Musik in interaktiven Medien sich in vielerlei Hinsicht gegenüber ihren frühen ausschließlich unterhaltenden Erscheinungsformen um einiges weiterentwickelt. Video- und Computerspiele werden von professionellen Komponisten vertont, von professionellen Musikern eingespielt und professionell produziert. Ihr Einsatz in Spielen dient nicht mehr der bloßen Unterhaltung, auch wenn das zuweilen immer noch vorkommt, sondern ist durch Dramaturgie, Narration und Immersion motiviert. Das Anwendungsfeld für Musik ist darüber hinaus deutlich breiter geworden. Programme und Betriebssysteme benutzen kleine musikalische Einwürfe als funktionales auditives Feedback für den Benutzer, der bspw. auf eingehende E-Mails und wichtige Systemmeldungen hingewiesen werden soll, und wenn es nur signalisiert, dass das Betriebssystem fertig hochgefahren ist. In diesen Beispielen ist die Musik nicht ausschließlich schmückendes Beiwerk, sondern gewinnt mehr und mehr an funktionaler Bedeutung. Schließlich finden aus der Musik stammende Konzepte sogar außerhalb der Musik Anwendung. Sanderson, Wee, Seah & Lacherez (2006) und Stockmann (2008) beschreiben, wie bestimmte Tonfolgen und rhythmische Muster für Sonifikationsaufgaben genutzt werden, u.a. für Anzeige- und Alarmfunktionen im medizinischen Bereich.

Die Musik in interaktiven Medien hat also auch ihre »Vorformen, ihre Vorgeschichte, ihre Geschichte und ihre Entwicklungsperspektiven«¹⁴. Ihre ersten For-

¹⁴Lissa (1965), Seite 23.

men litten an technischen und musikalischen Unwegsamkeiten. Die Neubesinnung auf tradierte Konventionen besonders aus dem direkten Vorfahren (Film-musik) führte zu einer Professionalisierung in den Bereichen Komposition und Produktion. Noch hat sich diese neue Musik jedoch nicht aus den Schatten ihrer Vorfahren erhoben und zu eigenen Formen, Ausprägungen und Perspektiven gefunden. Der Grund dafür ist, dass es bis heute immer noch weitgehend an einer organischen Verflechtung von musikalischem Wandel und interaktivem Geschehen fehlt.

2.3 Narrative Funktionen der Musik

In so komplexen Gefügen wie dem Film oder dem Computerspiel hat die Musik mehr zu repräsentieren als nur sich selbst, um ihrem Dasein Berechtigung zu geben. Sie erfüllt Funktionen, die über ihren bloßen Unterhaltungszweck hinausgehen. Diese sog. *funktionale Musik* unterscheidet sich von der *autonomen Musik*. Sie steht in Bezug zu Texten, Bildern oder anderen außermusikalischen programmatischen Konzepten und Gestaltungsmedien. Im Gegensatz zur autonomen Musik muss sie diese auch nicht in sich vereinnahmen (etwa Text in Form von Gesang), um für sich allein stehen zu können. Sie braucht nicht für sich allein stehen, denn sie hat ihre Funktionen nur im Kontext des sie einbindenden Mediums (Kungel 2004).

Einige der bedeutsamsten Versuche, diese Funktionen zu systematisieren sollen nun vorgestellt werden. Sie sind zu einem Großteil für die Filmmusik formuliert worden, lassen sich aber auf das interaktive Medium übertragen. Eine vergleichbare Aufarbeitung, die das interaktive Moment mit einbezieht, fehlt bislang und soll durch die folgenden Ausführungen gegeben werden.

2.3.1 Das filmische Erbe

Parallelismus und Kontrapunkt

Sergej Eisenstein (1940) entwickelte die Theorie des Synchronismus und unterschied drei Arten. Hat der Ton eine direkte Referenz im Bild (z.B. Vogelgezwitscher und der dazu gehörende Vogel), spricht Eisenstein vom *natürlichen Synchronismus*. Der *künstlerische Synchronismus* stellt hingegen eine Aufweichung dar, indem lediglich die Struktur des Klangs in Bewegung, Rhythmus, Farbe usw. eine Entsprechung im Bild hat. Schließlich führt er noch die *Chromophonie* als Entsprechung von Klangfarbe und Licht im Bild ein.

Bereits 12 Jahre zuvor verfassten Eisenstein, Pudowkin & Alexandrow (1928) ihr berühmtes *Manifest zum Tonfilm*, welches als einer der frühesten Versuche einer Systematisierung gilt und noch heute Gültigkeit hat. Sie unterscheiden zwischen *Parallelismus* im Sinne inhaltlicher Übereinstimmung zwischen Musik und Bild (der zuvor beschriebene Synchronismus ist lediglich eine spezielle bildzentrierte Ausprägung des Parallelismus) und *Kontrapunkt*, der sich dadurch auszeichnet, dass Musik und Bild einander inhaltlich ergänzen und sogar widersprechen. Hier kann die Musik dem Bild eine weitere Dimension analoger Inhalte hinzufügen und ermöglicht dadurch neue Ausdrucksformen, die gerade aufgrund ihrer Subtilität äußerst wirkungsvoll sein können. Schneider (1990) beschreibt, wie mächtig diese Wirkung sein kann: »Die Musik setzt die Zurechnungsfähigkeit des Kinobesuchers herab, macht ihn verführbar.«¹⁵ Auch Adorno & Eisler (1947b) ermutigen zu einer ausgeprägteren Arbeit mit dem dramaturgischen Kontrapunkt. Sie fordern, dass die Musik, »anstatt sich in der Konvention der Nachahmung des Bildvorgangs oder seiner Stimmung zu erschöpfen, den Sinn der Szene hervortreten lässt, indem sie sich in Gegensatz zum Oberflächengeschehen stellt.«¹⁶

Jedoch lassen sich Parallelismus und Kontrapunkt nicht immer klar voneinander trennen. Um der Grauzone, in der keine klare Zuordnung möglich ist, besser gerecht zu werden, unterscheidet Wolfgang Thiel (1981) drei Ausprägungen: Die *Bildillustration* stellt die strenge Form des Parallelismus dar, die aus dem amerikanischen Fachjargon als *Underscoring* und *Micky Mousing* bekannt ist. Die *affirmative Bildinterpretation und -Einstimmung*, auch als *Mood Technique* bezeichnet, fügt dem Bild eigene Inhalte, Informationen und Blickwinkel hinzu, um seine Aussage zu unterstützen, intensivieren, konkretisieren, nicht aber zu kommentieren oder gar zu widersprechen. Dies ist Aufgabe der *kontrapunktierenden Bildinterpretation und -kommentierung*, welche Adorno und Eisler als »dramaturgischen Kontrapunkt« bezeichnen. Auch das sog. Pauli-Modell entspricht dieser Systematisierung weitgehend: Hansjörg Pauli (1976) bezeichnet die drei Ausprägungen als *paraphrasierend*, *polarisierend* und *kontrapunktierend*.

Konkretisierung

Die bisher vorgestellten Systematisierungen sind knappe und stark verallgemeinernde Kategorisierungen. Mit ihrer eigenen Klassifizierung der Funktionen der auditiven Schicht im Film, besonders der Musik, schafft es Zofia Lissa (1965), außerordentlich konkret zu werden. Ihr Systematisierungsversuch gilt bis heute als einer der bedeutsamsten und einflussreichsten. Sie unterscheidet die folgenden achtzehn Kategorien, welche den obigen Kategorien (paraphrasierend,

¹⁵Schneider (1990), Seite 72.

¹⁶Adorno & Eisler (1947b), Seite 30.

polarisierend, kontrapunktierend) nicht immer eindeutig unterzuordnen sind. Eine ausführliche Erläuterung dieser Funktionen wird in Anhang A.1 gegeben.

1. Musikalische Illustration
2. Musikalische Unterstreichung von Bewegungen
3. Musikalische Stilisierung realer Geräusche
4. Musik als Repräsentation des dargestellten Raums
5. Musik als Repräsentation der dargestellten Zeit
6. Deformation des Klangmaterials
7. Musik als Kommentar
8. Musik in ihrer natürlichen Rolle
9. Musik als Ausdrucksmittel psychischer Erlebnisse
10. Musik als Grundlage der Einfühlung
11. Musik als Symbol
12. Musik als Mittel zur Antizipierung des Handlungsinhalts
13. Musik als formal einender Faktor
14. Mehrfunktionalität und Mehrschichtigkeit der Filmmusik
15. Funktionen der Geräuscheffekte
16. Funktion der Rede im Film
17. Funktion der Stille im Film
18. Nichtfunktionelle Zuordnung der audio-visuellen Schichten im Tonfilm

1965, das Jahr der Veröffentlichung von Lissas *Ästhetik der Filmmusik*, war die Filmmusik noch deutlich in den Traditionen der autonomen Musik verhaftet. Kompositionstechniken, die speziell den Erfordernissen im Film gerecht werden, waren gerade dabei, sich zu entwickeln. Dazu gehört insbesondere die Fähigkeit, musikalisch schnell auf inhaltliche Entwicklungen zu reagieren. Schon Adorno & Eisler (1947b) formulierten, »dass der von der traditionellen Musik und in der akademischen Formenlehre niedergelegte Formenschatz für den Film weithin unbrauchbar ist. [...] Höhepunkte müssen unter Umständen von

der Musik unmittelbar, ohne alles Crescendo, oder nur mit der knappsten Vorbereitung erreicht werden.«¹⁷ Heute, ein halbes Jahrhundert nach Lissas Untersuchungen, hat die Musik diese formal offene und unverbindliche Feingliedrigkeit erreicht und die nichtfunktionellen Überbleibsel musikhistorischer Vorformen (Lissas Kategorie 18) weitgehend unnötig gemacht.

Mit ihren achtzehn Kategorien gelang Lissa eine derart detaillierte Systematisierung, die nahezu alle filmmusikalischen Erscheinungen bis 1965 und darüber hinaus abdeckt, dass selbst ein halbes Jahrhundert später der Filmmacher Reinhard Kungel (2004) nur Detailergänzungen vornimmt.

Erst der akribische Systematisierungsversuch des Theater- und Filmmusikkomponisten sowie Musikwissenschaftlers Johnny Wingstedt (2004, 2008) klassifiziert die narrativen Funktionen der Musik im multimedialen Umfeld (in erster Linie für den Film) und deckt dabei auch neue Funktionen ab, die sich erst nach Lissa herausgebildet haben oder durch ihre Systematik nicht erfasst bzw. differenziert werden konnten. Er unterscheidet sechs Klassen von Funktionen, die sich z.T. in weitere Unterkategorien unterteilen. Eine ausführlichere Darstellung wird in Anhang A.2 gegeben; hier folgt eine geraffte Darstellung.

Emotionale Klasse: Die emotionalen bzw. emotionalisierenden Funktionen der Musik sind am häufigsten anzutreffen, gewöhnlich in Kombination mit Funktionen anderer Klassen. Musiken, die Funktionen dieser Klasse bedienen, helfen dem Betrachter, sich in Personen, Umgebungen und Situationen einzufühlen.

Informative Klasse: Die Funktionen, die den drei Unterkategorien dieser Klasse angehören, eint die Absicht, Informationen zu vermitteln, nicht Emotionen. Wie Schneider (1990) bereits feststellte und auch Marshall & Cohen (1988) und Cohen (1998) in ihren eigenen Untersuchungen belegen, wird die Bedeutung der Musiken mit informativer Funktion sehr konkret und eindeutig wahrgenommen. Entsprechend mächtig sind diese Musiken, die

1. analoge Inhalte vermitteln,
2. assoziative Inhalte evozieren und
3. Wiedererkennung aufbauen.

Beschreibende Klasse: Anders als die informative Klasse, die eher passiv Assoziationen und Analogien vermittelt, beschreiben die Funktionen dieser Klasse hauptsächlich physikalische Phänomene in der Szene, nehmen aktiv am Geschehen teil. Dazu zählen die

¹⁷Adorno & Eisler (1947b), Seite 86.

1. Beschreibung des Settings und die
2. Beschreibung physikalischer Aktivität.

Führende, leitende Klasse: Funktionen dieser Klasse lenken das Auge, die Gedanken, Meinung und Erwartung des Zuschauers. Sie sind ein Werkzeug zum Hervorheben und Verstärken visuell schwacher Details. Solche Musiken können auch den Zugang zu einem tieferen Verständnis einzelner Szenen oder des Gesamtwerkes erleichtern. Dabei sind sie

1. indikativ oder
2. maskierend.

Zeitliche Klasse: Musik wird oft auch bezeichnet als Zeitkunst, als gestaltete Zeit. Sie kann ohne die zeitliche Dimension nicht existieren. Naheliegender Weise ist sie gerade für zeitbezogene Funktionen prädestiniert.

1. Musik wirkt dem Bildschnitt und der daraus resultierenden Episodenhaftigkeit entgegen und sorgt für Kontinuität.
2. Musik definiert Struktur und Form.

Rhetorische Klasse: Die Musik kann der Szene eine weitere eigenständige und häufig verallgemeinernde Inhaltsschicht hinzufügen. Diese wirkt im Sinne eines Kommentars, einer Auseinandersetzung mit dem Gezeigten und als Mittel zur Beeinflussung der Interpretation des Zuschauers. Die große suggestive Kraft des Films liegt in dieser Musik begründet, die das subtile Vermitteln solcher Inhalte überhaupt erst ermöglicht.

Zusammenfassung

Die vorgestellten Systematisierungsversuche haben ihre Richtigkeit in Jahren und Jahrzehnten musikwissenschaftlicher Untersuchungen, sowie praktischer Anwendungen und Umsetzungen im Film bewiesen. Unterschiede rühren in erster Linie vom jeweils unterschiedlichen Abstraktionsgrad und dem jeweiligen musikhistorischen Entwicklungsstand her. Es ist nicht die Absicht dieser Arbeit, diesen Fundus um einen weiteren Versuch zu erweitern. Er soll vielmehr die Grundlage und Ausgangspunkt für die weiteren Betrachtungen sein, die auch das interaktive Moment mit einbeziehen.

2.3.2 Einbeziehung der Interaktivität

Als eine der vielseitigsten Erscheinungsformen interaktiver Medien, in der die auditive Schicht exzessiv genutzt wird, erweist sich das Video- und Computerspiel. Hier haben sich Geräusch und Musik als feste Bestandteile etabliert. Die

Vielfalt ihrer Funktionalitäten deckt nicht nur all jene Funktionen ab, die im vorigen Abschnitt beschrieben wurden und z.T. auch im nicht-spielerischen Kontext auftreten, sondern geht darüber hinaus. Einige Beispiele lassen sich in den aus der Praxis heraus entstandenen Schriften von Sanger (2003), Langer (2007), Marks (2008) finden. Es ist also angebracht, dem Auditiven in Computerspielen besondere Aufmerksamkeit zu widmen. Natürlich stehen im Fokus dieser Arbeit vorrangig die musikalischen Aspekte.

Dabei fällt auf, dass der Parallelismus bis heute das vorherrschende Prinzip bei der Vertonung von Computerspielen ist. Die musikalische Illustration von Bewegungen und Geräuschen (Micky Mousing) hat zu einer ganz eigenen Geräuschästhetik geführt, welche sogar in die Zeichentrickverfilmungen der *Super Mario Bros.*-Spiele übernommen wurde. Weiterhin lassen sich die Musiken zunehmend auch der *affirmativen Bildinterpretation und -Einstimmung* zuordnen. Sie vermitteln Stimmungen, welche die Grafiken aufgrund technischer Beschränkungen nicht immer überzeugend vermitteln können und verhelfen der virtuellen Welt und virtuellen Charakteren zu mehr Glaubwürdigkeit. Von allen Autoren wird die Funktion der Einfühlung in das Szenarium als die prominenteste genannt.

Es lassen sich nahezu alle musikalischen Elemente der filmischen Inszenierung auch auf die interaktive Inszenierung übertragen. Letztlich bedeutet es nur, dass bestimmte Musiken im Zusammenhang mit entsprechenden Ereignissen erklingen. Betritt der Spieler einen Ort, erklingt die zugehörige Immersions-Musik. Erscheint ein bestimmter Nicht-Spieler-Charakter in der Szene, wird dessen Musik zugeschaltet. Treten bestimmte Ereignisse im Spielverlauf ein, wird die dafür vorgesehene Musik gestartet.

Die Musik und generell die auditive Schicht im Spiel haben eine nicht zu unterschätzende Wirkung auf den Spieler und bergen damit ein oft verschenktes Potential (Garcia 2006, Moffat & Kiegler 2006, Röber & Masuch 2005a). Kristine Jørgensen (2006) stellt eine Besonderheit fest, welche nur in interaktiven Medien auftritt: Im Gegensatz zur nicht-diegetischen Musik im Film, die sich nur an den Zuschauer richtet und damit kein die Handlung beeinflussender Faktor ist, hat die nicht-diegetische Musik im Spiel sehr wohl Einfluss auf den Spielverlauf, denn sie beeinflusst das Spielerverhalten.

Im Spiel ist der Zuschauer ein aktiver Teilnehmer, eine Variable, auf die Musik, Sprache und Geräusch anders als bisher Bezug nehmen können. Zusätzlich zur diegetischen und nicht-diegetischen Schicht, die bereits aus dem Film bekannt sind, führt Jørgensen (2006, 2010) die *trans-diegetische* Schicht ein. Hier kommunizieren Elemente der Spielwelt und eben auch die Musik mit dem Spieler. Ein typisches Beispiel sind die Gedankenreden des Protagonisten, die dem Spieler mitteilen: »Das geht nicht!«, wenn er bspw. eine verschlossene Tür zu öffnen ver-

sucht. Auch das »Ja, Sir!« eines angewählten Soldaten in Strategiespielen ist der trans-diegetischen Schicht zuzuschreiben. Aber lassen sich die Elemente dieser Schicht wirklich von der diegetischen und nicht-diegetischen abgrenzen? Die nicht-diegetische Erzählerstimme ist auch schon aus dem Film bekannt. Die Gedankenrede ist jedoch im eigentlichen Sinne nicht an den Spieler gerichtet, sondern an den Protagonisten selbst, in dessen Haut der Spieler steckt. Als Teil der Spielwelt nimmt er hier also etwas wahr, dessen Quelle ebenfalls der Spielwelt angehört und in ihr auch wahrgenommen wird, also diegetisch ist. Auch die Antwort des Soldaten auf das, wenn auch nicht sichtbare, virtuelle Pendant des Spielers, nämlich seinen Befehlshaber, kann als diegetisch klassifiziert werden.

Die trans-diegetische Schicht stellt also keine Differenzierung von der diegetischen und nicht-diegetischen Schicht dar. Mit diesem Konstrukt versucht Jørgensen vielmehr, all jene diegetischen Elemente zu erfassen, die intentional an den Spieler gerichtet sind, und diejenigen nicht-diegetischen Elemente, die sich auf die Diegese auswirken können. Zum einen können sie eigenständig als inhärenter Teil der Spielmechanik erklingen; in diesem Fall erfüllen sie Anzeigefunktionen (z.B. gibt der Klang der Fußstapfen Aufschluss über die Beschaffenheit des Fußbodens) oder Hinweisfunktionen (ein Soldat meldet, dass er angegriffen wird). Zum anderen erklingen sie als Antwort auf Spielerinteraktionen (»Ja, Sir!«) und stellen somit ein bestätigendes Feedback für den Spieler dar. Zwischen Sprache, Geräuschen und Musik differenziert Jørgensen nicht extra und stellt die folgenden fünf Klassen von Funktionen der auditiven Schicht vor.

Aktions-orientierte Funktionen: Hier kommen Icons¹⁸ und Earcons¹⁹ zum Einsatz, die der Spielsteuerung sowohl durch ihre Anzeige- als auch Feedback-Funktion dienen.

Atmosphärische Funktionen: Musik und Geräuschkulisse haben die Aufgabe, die Einfühlung in die Spielwelt und in Situationen zu vermitteln. In einer Benutzerstudie belegt Jørgensen deren hohe immersive Kraft. Das Ausschalten von Musik und Geräuschkulisse wurde von den Probanden mit einem spürbaren Rückgang der Präsenz der virtuellen Umgebung und mit einer Reduktion der Spielwelt auf Regeln und Spielmechaniken beschrieben.

Funktionen zur Orientierung: Hier werden Informationen über Objekte und Orte in der Umgebung einschließlich deren Richtung ausgehend von der

¹⁸*Auditory Icons* sind diegetische Geräusche, die in der Welt naturgemäß vorkommen, wie etwa Fußstapfen, Schussgeräusche usw.

¹⁹*Earcons* sind nicht-diegetische Hinweissignale, hinweisgebende Musikeinsätze und Interface-Sounds.

Spielerorientierung vermittelt. Lissas Kategorie 4 (Repräsentation des dargestellten Raums), insbesondere die auditive Darstellung dessen, was außerhalb des Sichtbereiches liegt, findet hier ihre Entsprechung. Positional-Audio- und Surround-Technik sind Mittel, durch die der Informationsgehalt noch deutlich präzisiert und die Effektivität dieser Funktionen aufgewertet werden kann (Friberg & Gärdenfors 2004, Röber & Masuch 2005a,b).

Kontrolle-unterstützende Funktionen: Spiele, in denen komplexe Strukturen gesteuert werden müssen (etwa Strategiespiele und Wirtschaftssimulationen), können die komplette Bandbreite an Informationen meist nicht ausschließlich graphisch vermitteln, ohne die visuelle Aufnahmefähigkeit des Spielers zu überfordern. Die meisten Informationen können auch auf auditivem Wege vermittelt werden (z.B. durch Sprachmeldungen wie »Sie haben Post.«) und den visuellen Kanal ergänzen oder entlasten. Diese Komplementarität von visueller und auditiver Schicht wird auch von Kramer, Walker, Bonebright, Cook, Flowers, Miner & Neuhoff (1997) und Eldridge (2002) besprochen. Das Abschalten des auditiven Informationskanals hat bei Jørgensens Probanden zu Desorientierung geführt, obwohl alle spielrelevanten Informationen auch visuell verfügbar waren. Die Spiele wurden als erheblich schwerer beherrschbar und sogar unspielbar beschrieben (Jørgensen 2008). Speziell mit Musik und Sprache für auditive Interfaces beschäftigen sich Alty, Rigas & Vickers (2005). Auch im medizinischen Bereich werden für Anzeige- und Alarmzwecke auditive Ausgaben eingesetzt (Sanderson et al. 2006, Hermann, Baier, Stephani & Ritter 2006).

Funktionen zur Identifizierung: Musik kann Verwandtschaft von Situationen vermitteln. Etwa bei drohender Gefahr können typische Gestaltmerkmale erklingen, die der Spieler nur in diesem Zusammenhang hört und dadurch gewarnt wird. Gleiches gilt für Geräuscheffekte. Designat für solche Assoziationen können auch Charaktere und Objekte in der virtuellen Welt sein, die aus technischen oder Designgründen visuell vielleicht nur schwer identifizierbar sind. Das Auditive kann hier Abhilfe verschaffen, mehr noch; auch der Zustand des Designates lässt sich vermitteln. So klingen bspw. eine gut geölte und eine rostige klapperige Maschine verschieden.

Jørgensens Betrachtungen galten in erster Linie dem Einfluss und Nutzen der auditiven Schicht für die Bedienbarkeit und Beherrschbarkeit von Computerspielen. Außer den atmosphärischen Funktionen sind alle ihre Kategorien vom Standpunkt der Bedienungsunterstützung aus aufgestellt worden. Dadurch erklärt sich vielleicht auch die fehlende Differenzierung zwischen Musik, Sprache und Geräusch. Jede einzelne dieser auditiven Schichten kann aber noch eine

Vielzahl weiterer Funktionen wahrnehmen, die in dem Kontext von Jørgensens Untersuchung freilich nicht erfasst werden konnten und wohl auch nicht sollten.

Als eine weitere Erkenntnis aus ihrer Studie berichtet Jørgensen (2006), dass die Akzeptanz der Musik im Spiel – und auch ihr Verständnis – durch die Konventionen der Musik im Film bereits etabliert ist. Das macht auch Lenaerts (2006) deutlich, der sich weitgehend an Wingstedts Systematisierung hält. Wingstedt (2004) gibt anhand zweier Beispiele einen Eindruck davon, dass die gleichen narrativen Funktionen der Musik im Film auch im interaktiven Kontext anwendbar sind. Seine führenden, leitenden Funktionen können die Aufmerksamkeit des Benutzers lenken, auf Details fokussieren, maskieren. Jørgensens Funktionen (außer den atmosphärischen) gehören in eben diese Klasse. Wingstedts zweites Beispiel ist die für Kontinuität sorgende Funktion der Musik, die Szenen miteinander verschmilzt und dem Bildschnitt seine alles zerhackende Abruptheit nimmt. Er schreibt: »In interactive non-linear media, such as computer games, music's ability to provide continuity is an important quality with strong potential.«²⁰

Noch immer bezieht sich die Musik nur auf die den Spieler umgebende virtuelle Szene. Dies ist Filmmusik! Sie ist für einen außenstehenden Betrachter konzipiert. Sie hilft ihm, die dazugehörigen Szenen nachzuvollziehen, sich in sie hineinzusetzen, sich ein Urteil zu bilden – und doch bleibt er passiver Betrachter. Der Spieler aber ist aktiver Teilnehmer am virtuellen Geschehen und das Vertonungskonzept sollte ihn mit einbeziehen anstatt auszuschließen. Wie das aussehen kann, wird etwa von den Spielen *Black & White* und *Fable* (Lionhead Studios 2001, 2005) demonstriert. Hier reagiert die Spielwelt und das Verhalten der Nicht-Spieler-Charaktere auf die Handlungsweise des Spielers und fällt ein moralisches Urteil über ihn. Nicht-Spieler-Charaktere mögen dem Spieler freudig entgegenlaufen, wenn er die Szene betritt, oder aber angstvoll vor ihm fliehen. In gleicher Weise kann auch Musik auf die Handlungen des Spielers Bezug nehmen und erreicht ihn durch ihre Subtilität auf einer sehr persönlichen Ebene. Ein moralischer Kommentar bezieht sich nicht mehr auf die Handlungen einer anderen Person, sondern auf die eigenen. Die emotionalisierende Wirkung der Musik kann dem Spieler Stolz und Reue für sein Handeln vermitteln. Es werden Ausdrucksformen möglich, die so persönlich nicht einmal der Film darstellen kann.

Natürlich bieten sich nicht alle Funktionen für eine solche auf den Interakteur zentrierte Übertragung an. Die Musik, welche die Einfühlung in das Szenarium oder in andere Charaktere unterstützen soll, bezieht sich naturgemäß auf eben

²⁰Wingstedt (2004), Seite 8.

diese, nicht auf den Spieler. Es folgt eine Darstellung der Funktionen, die Musik im Bezug auf den Spieler wahrnehmen kann.

Unterstützung: Wie bereits bei Jørgensen (2006) beschrieben wurde, kann die auditive Schicht und als Teil von ihr auch die Musik den Spieler bei der Handhabung des Spieles unterstützen, kann den visuellen Kanal entlasten und relevante Informationen gegenüber weniger relevanten hervorheben. Dringende Nachrichten können bspw. lauter signalisiert werden. Die Orientierung kann unterstützt werden; Kapitel 4 wird speziell hierzu ein Beispiel zeigen. In dieser Funktion kommuniziert die Musik eigentlich nur von innen nach außen. Sie tritt nicht in einen »Dialog« mit dem Spieler. Sie wird aber in ihrer Anlage stark von spielerischen Erfordernissen geprägt sein und kann auf verschiedenen Schwierigkeitsgraden auch unterschiedlich viel Unterstützung leisten.

Selbstverständnis: Durch musikalische Illustration der Bewegungen, welche die Spielfigur, also das virtuelle Pendant des Spielers, ausführt, und durch Deformation des Klangmaterials kann die Selbstreflexion des Spielers beeinflusst und z.B. eine selbstironische Haltung geweckt werden. Auch das Gegenteil ist möglich. Die Handlungen erhalten durch Unterstreichung oder Verfremdung eine ungeahnt ernste Bedeutung.

Konsequenzen vermitteln: Die Handlungen des Spielers können Auslöser sein für komplexe Ereignisketten, deren Konsequenzen erst nach einiger Zeit, vielleicht sogar erst nach Stunden, sichtbar werden. Mit dem auslösenden Ereignis einhergehend kann ein neuer musikalischer Gedanke (z.B. ein neues Motiv) eingeführt werden, der weitergeführt wird, quasi das Räderwerk hinter den Kulissen repräsentiert, in den Hintergrund tritt und immer dann wieder hervortritt, wenn die Konsequenzen der einstigen Handlung zum Vorschein kommen. Der Assoziationsraum der Musik wird hier durch die Interaktionen des Spielers aufgebaut. Andere Interaktionen können zu anderen Querverweisen führen, so dass unterschiedliche Spieldurchläufe auch musikalisch sehr verschieden sein können.

Steuerung von Spielerverhalten und Spieldynamik: Durch Musik kann die Handlungsweise des Spielers beeinflusst werden (Kim & André 2004a,b). Schnelle rhythmische Musik kann das Tempogefühl in einem Autorennen langsamer erscheinen lassen, als es tatsächlich ist. Das Musiktempo ist so schnell, dass es dem Geschehen vorausseilt, davonläuft, das szenische Geschehen hinter sich zurücklässt. Der Effekt ist schon im Film bekannt. Musik, die vom Szenentempo weniger stark abweicht, kann ein deutlich intensiveres Tempogefühl vermitteln (Kungel 2004). Durch die formbildende Eigenschaft der Musik kann sie den Handlungen des Spielers außer-

dem eine bestimmte »Ästhetik« geben und ihn unbewusst dazu anspornen, im Takt zu bleiben. Auch in der realen Welt ist ein solcher Effekt bekannt: Laute rhythmische Musik vermag, dem Hörer ein bestimmtes Schrittempo aufzuzwingen und beeinflusst seine Herzfrequenz (Rubisch, Husinsky, Doppler, Raffaseder, Horsak, Ambichl & Figl 2010). Von dieser Wirkung der Musik wird heute schon im Ausdauertraining und in der Bewegungstherapie Gebrauch gemacht (Bräuninger & Blumer 2004). Für die Vermittlung eines bestimmten Spielgefühls offenbart sich hier großes Potential. Entspricht das Spielerverhalten nicht der beabsichtigten Szenendynamik, kann eine entsprechend gewählte Musik nachhelfen und den Spieler etwa antreiben oder abbremsen.

Beeinflussung und Wertung von Spielerentscheidungen: Durch Musik, die psychische Inhalte vermittelt (Erinnerungen, Gefühle), und durch symbolbehafte Musik kann der Spieler bei seiner Entscheidungsfindung beeinflusst werden. Bestimmte Assoziationen können besonders betont und wieder andere ausmaskiert werden. Das lenkt die Gedanken des Spielers in eine entsprechende Richtung. Ein auf die gefällte Entscheidung folgender musikalischer Kommentar kann dann wieder Stellung dazu nehmen. Wie im Film offenbaren sich auch hier große Chancen und Gefahren in einem erzieherischen Kontext.

Entscheidungs-dramaturgie: In einem Spiel sind für gewöhnlich viele Entscheidungen zu fällen. Welche Antworten wählt man im Gespräch mit einem Nicht-Spieler-Charakter? Welchen Lösungsweg schlägt man ein? Welcher Fraktion²¹ schließt man sich an? Einige dieser Entscheidungen mögen sekundär und andere spielentscheidend sein. Mit musikalischen Mitteln lassen sich die wichtigen Entscheidungen gegenüber den anderen hervorheben und als wichtig kenntlich machen. Der Spieler wird dazu gebracht, diese Entscheidungen besser zu überdenken, als er es sonst vielleicht würde.

Die Musik in Computer- und Videospielen kann also zunächst all jene Funktionen wahrnehmen, die sie auch im Film bereits erfüllt. Dabei ist zu beachten, dass Länge und Reihenfolge von Szenen und Ereignissen in interaktiven Medien nicht feststehen. Außerdem ist der Spieler aktiver Teilnehmer am Geschehen in der Spielwelt, wird dabei aber durch nicht-diegetische Informationen (auch Musik) beeinflusst. Bei der Entwicklung eines Vertonungskonzeptes ist es

²¹In Rollenspielen gibt es oft mehrere »Interessensgemeinschaften« (Gilden, Parteien, Banden etc.), denen sich der Spieler anschließen kann. Mit dem Eintritt in eine dieser Fraktionen verändert sich das Verhalten der anderen ihm gegenüber und hat oft auch Auswirkungen auf den weiteren Spielverlauf, wie etwa in der *Gothic*-Serie (Piranha Bytes 2001, 2002, 2006).

deshalb wichtig, darauf zu achten, dass die nicht-diegetische Musik dem Spieler bestimmte Informationen nicht schon zu früh verrät und dadurch Überraschungseffekte und unerwartete Wendungen in der Handlung vorwegnimmt²². Zusätzlich ist es nun möglich, den Spieler in das Vertonungskonzept mit einzu beziehen. Bis heute kann dieses Potential nicht voll ausgeschöpft werden, fehlt es doch an etablierten musikalischen und technischen Lösungen, um Musik und interaktives Geschehen miteinander zu koordinieren. Das ist die Kernproblematik, der sich diese Arbeit im weiteren widmet.

²²Es kommt tatsächlich vor, dass dramatische Momente verpuffen, weil die Musik sie zu früh entlarvt.

Verwandte Arbeiten

DIE musikalisch kohärente Vertonung interaktiver Medien ist ein Problem, welches besonders die Video- und Computerspieleindustrie bereits seit den 1980er Jahren beschäftigt. Folglich sind einige der bislang bedeutsamsten Lösungsansätze in diesem Umfeld entstanden und haben ihre praktische Relevanz in verschiedensten Spieleapplikationen unter Beweis stellen können. Wesentlich jünger ist das wissenschaftliche Interesse an dieser Problematik, das erst um die Jahrtausendwende zu keimen begann.

Egal, ob in wissenschaftlichem oder nichtwissenschaftlichem Umfeld entstanden, ist allen Lösungsansätzen eines gemein. Um Musik mit dem interaktiven Geschehen zu verbinden, sind Eingriffe in den musikalischen Ablauf und in die Musik selbst notwendig. Darüber hinaus gibt es jedoch keine systematische Aufarbeitung des Themengebietes. Dieses Kapitel soll die Lücke schließen. Dem liegt die folgende Veröffentlichung zu Grunde:

- (Berndt 2009).

Nachfolgend werden nach der Art der Eingriffe insgesamt sechs grundsätzliche Lösungsstrategien unterschieden. Ausgehend vom Arrangement vorgefertigter statischer Musikfragmente werden zunehmend flexiblere Konzepte eingeführt, die differenzierter in das musikalische Material eingreifen, es schließlich sogar generieren. Einen Überblick gibt Abbildung 3.1 Für jeden Ansatz werden repräsentative Vertreter benannt.

3.1 Sequentielles Arrangement

Die frühesten Ansätze zur nichtlinearen Musik finden sich bereits in der Klassik (ca. 1730 bis 1830) mit den *musikalischen Würfelspielen* von Kirnberger (1767)

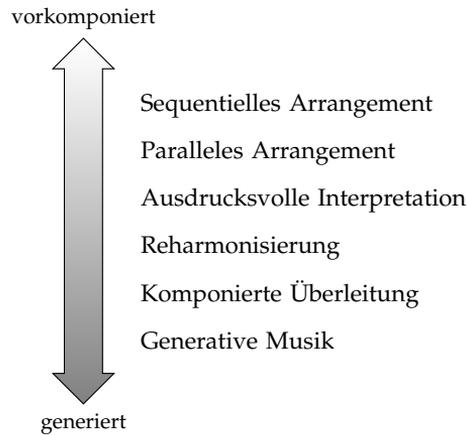


Abbildung 3.1: Eine Übersicht der Lösungsstrategien zur nichtlinearen Musik, eingeordnet im Spannungsfeld zwischen der Arbeit mit vorgefertigtem in sich unveränderlichem Material und generativen Techniken, die musikalisches Material adaptieren und sogar erzeugen.

und Mozart (1787)¹. Die Musik wird als sequentielle Abfolge fest vorkomponierter Versatzstücke betrachtet. Aus einem größeren Korpus solcher Segmente können immer neue Sequenzen zusammengestellt werden. In den klassischen Würfelspielen sind die Segmente von der Länge genau eines Taktes. In einer Jazz-Variante von Jörg Ratai (2005) sind sie zwei und drei Takte lang und umfassen ganze Jazz Changes². Bei Stockhausen (1956) sind es in sich abgeschlossene Figuren und Phrasen unterschiedlicher Länge.

Die Idee des sequentiellen Arrangements lässt sich für die Vertonung interaktiver Anwendungen adaptieren. Die Musik wird nach wie vor als Abfolge von Segmenten betrachtet. Diese kann nun aber während der Wiedergabe neu arrangiert werden, d.h., die Entscheidung, welches Segment auf das aktuell spielende folgt, kann bis zu dessen Ablauf neu gefällt werden. Im Falle eines Musikwechsels wird dann statt dem nächsten regulär vorgesehenen Segment nahtlos das erste Segment der neuen Musik angehängt. Diesem kann noch ein überleitendes Segment vorangestellt sein. In dem Spiel *Gothic II* (Piranha Bytes 2002) wird der Übergang zusätzlich von einem perkussiven Effekt kaschiert, etwa ein Pauken- oder Beckenwirbel, der über die laufende Musik blendet, sie klanglich überdeckt und dadurch die Bühne freiräumt für den Einsatz der neuen Musik.

¹Eine Implementierung von Mozarts musikalischem Würfelspiel stammt von Berndt & Knolle (2001).

²Jazz Changes sind floskelhafte Harmoniefolgen.

Alle Segmente sind vorkomponiert und liegen als Audiodaten vor. Es ist Aufgabe von Komponist und Produzent, für eine gute metrische, melodische, harmonische und klangliche Konnektivität der Segmente zu sorgen. Brüche, Sprünge, ästhetisch unpassende Wendungen im melodischen und harmonischen Ablauf müssen vermieden werden, ebenso rhythmische Stolperstellen. Die Segmente müssen musikalisch sauber aneinander anschließen. Der kompositorische Aufwand ist deshalb verhältnismäßig hoch. Da ein musikalischer Wandel immer erst am Ende des aktuell spielenden Segments geschehen kann, ist außerdem eine feine Granularität (kurze Segmente) zu Gunsten einer zeitnahen Reaktivität wünschenswert. Das schränkt jedoch die musikalischen Entfaltungsmöglichkeiten, die nun nur noch innerhalb eines Segments möglich sind, extrem ein, schließlich muss an dessen Ende wieder ein konnektiver Zustand erreicht werden.

In Computerspielen ist das sequentielle Arrangement die vorherrschende Lösungsstrategie, vorzufinden bspw. in *Unreal II* (Legend Entertainment 2003), der *No One Lives Forever*-Serie (Monolith Productions 2000, 2002) und in einigen Teilen der *Gothic*-Reihe (Piranha Bytes 2002, 2006). Weitere Anwendungsbeispiele sind die schnell pulsierenden Radiobetten in Verkehrsfunk und Kurzmeldungen. Auch im akademischen Umfeld sind Musik-Engines entstanden, welche das Prinzip des sequentiellen Arrangements umsetzen (Tobler 2004, Aav 2005, Chiricota & Gilbert 2007).

Besondere Erwähnung verdient die *Interactive Music Streaming Engine* (kurz: *iMuse*) von Land & McConnell (1994). Interessant ist sie vor allem wegen ihrer eleganten Lösung. Die Musikstücke werden hier durch zusätzlich eingefügte (in der Wiedergabe aber nicht hörbare) Marken segmentiert. Diese Marken sind Sprungpunkte zu anderen Marken im selben oder einem anderen Musikstück. Wenn ein Musikwechsel verlangt ist, wird das Sprungziel der von der aktuellen Wiedergabeposition aus nächsten Marke gesetzt und die Marke aktiviert. Die Wiedergabe wird von dort aus nahtlos am Sprungziel fortgesetzt. Diese Marken sind nicht nur geeignet, um zwischen verschiedenen Musikstücken zu verzweigen. Sie können auch innerhalb ein und des selben Musikstücks verwendet werden. Das entspräche dem Umarrangieren des Musikstücks. Im Falle eines Musikwechsels muss außerdem nicht direkt in das Zielmusikstück gewechselt werden. Zuvor kann in eine Überleitung gesprungen werden, an deren Ende eine weitere Sprungmarke in die eigentliche Zielmusik führt. Es kann mehrere verschiedene Überleitungen aus einem Musikstück heraus geben, so dass am Ende eines jeden Segments nicht immer der gleiche konnektive Zustand erreicht werden muss. Das schränkt die kompositorischen Entfaltungsmöglichkeiten deutlich weniger ein, ist natürlich auch mit mehr Aufwand verbunden. Weitere Informationen zu *iMuse* sind in Anhang B zusammengestellt.

3.2 Paralleles Arrangement

Nicht nur die zeitliche Sequentialität sondern auch die zeitgleiche Parallelität, das Zusammenspiel mehrerer Stimmen, bietet Ansatzpunkte zum Arrangement vorkomponierter Musik. Nicht immer müssen alle komponierten Stimmen und Instrumente spielen. Manche können nach Belieben (*ad Libitum*) gespielt werden oder eben entfallen. Die Musik kann durch unterschiedliche Besetzungsvarianten mehr oder weniger stark im Klang- und Ausdruckscharakter variieren. In der Chormusik J. S. Bachs werden *ad-Libitum*-Oberstimmen, meist durch Trompeten aufgeführt, als Klangkrone für besonders feierliche Choralverse eingesetzt. Diese Praxis ist heute noch üblich (Gaar 1997), auch um »das Choralspiel abwechslungsreich zu gestalten«³.

In der Barockzeit (ca. 1600 bis 1750) waren Musiken mit reduzierbarer Besetzung dem Genre der *Ruralkompositionen* zugeordnet. Sie wurden für ländliche Verhältnisse konzipiert, d.h. für kleine Gemeinden mit kleinen Kantoreien (Chören), die selten die volle vierstimmige Standardbesetzung aufbringen konnten und kaum über Instrumentalisten verfügten. Optionale Stimmen, die nicht besetzt werden konnten, durften entfallen. Zu den *Ruralkompositionen* zählen u.a. verschiedene Messen von Valentin Rathgeber (1733). In der Fortführung dieser Tradition entstanden die *Baukastenkompositionen* von Manz & Winter (1976), Berndt (2008).

Wenn Stimmen entfallen und damit der mehrstimmige Satz ausgedünnt wird, besteht natürlich die Gefahr, dass das klangliche Resultat unvollständig wirkt. Fehlende Töne in Akkorden und Passagen mit geringer thematischer Dichte werden oft als Löcher oder Leerlauf-Phasen wahrgenommen. Der Komponist muss also darauf achten, dass die Architektonik der Musik auch in reduzierter Besetzung nicht zusammenbricht. D.h., in jeder möglichen Stimmenkombination muss auf die Ausgewogenheit musikalischer Strukturmerkmale geachtet werden, so dass jede Variante in sich homogen ist.

Dabei müssen die Varianten nicht alle die gleichen Strukturmerkmale und den gleichen Ausdruckscharakter haben. Sie können sehr unterschiedliches musikalisches Material vorstellen, verschiedene Melodien, Gegenmelodien, Begleitstimmen, Rhythmik. Die Harmonik kann in einer Variante schlicht gehalten, in einer anderen sehr komplex mit zusätzlichen Akkordtönen angereichert sein. Durch unterschiedliche metrische Prägnanz⁴ der Varianten kann sogar der Eindruck unterschiedlicher Tempi erweckt werden. Sogar die Verschmelzung unterschiedlicher Stile innerhalb eines einzigen schlüssigen Werkes, der sog. *stilistische Pluralismus*, ist möglich: Während einige Stimmen zueinander bezogen

³aus dem Vorwort zum Notenheft »89 Oberstimmen zu Sätzen des Posaunen-Choralbuchs zum Evangelischen Gesangbuch« (Baur, Böhler, Dieckmann, Frieß, Günther, Meier & Nonnenmann 1999)

⁴Die metrische Prägnanz gibt die Dichte der Rhythmus-gebenden Impulse an.

in einem festen stilistischen Rahmen komponiert sind, reichern andere dies mit Mitteln eines anderen Stiles an. Trotzdem werden auch diese Stimmen immer noch einen, wenn auch fernen, Bezug zur primären Musik aufweisen. Mit ihnen können nun aber neue wieder andere Stimmen kombiniert werden, die sich noch weiter vom Ursprung entfernen, mit diesem vielleicht gar nicht mehr kombinierbar sind. Eine kurze Einführung in den stilistischen Pluralismus gibt Anhang C.

Diese Variabilität kann in interaktiven Medien ausgenutzt werden. Jede Stimme entspricht einem Track, der synchron mit allen anderen, auch den stummen, spielt. Durch Überblend-Techniken können die verschiedenen Stimmen- bzw. Trackkombinationen automatisch ineinander überführt werden, ohne dass der musikalische Fluss wie beim harten Schnitt unterbrochen wird. Tatsächlich findet eine derart differenzierte Musikmontage, die jeden Track separat und vollautomatisch behandelt, heute noch nicht statt. Stattdessen werden die Tracks direkt ohne jeden Übergang ein- und ausgeschaltet, was an musikalischen Strukturgrenzen gut funktioniert, ansonsten aber recht abrupt und wenig schlüssig wirkt. Diese Arbeit stellt in den Kapiteln 4 und 6 Lösungen für komplexere automatische Montage-Techniken vor.

Musik-Engines, welche die Mittel des parallelen Arrangements nutzen, wenn auch nur in vereinfachter Weise, werden u.a. von Lenaerts (2006), Wingstedt, Liedahl, Lindberg & Berg (2005), Wingstedt (2006) beschrieben. Auch *iMuse* bietet die Möglichkeit, einzelne Audio-Tracks und MIDI-Kanäle stumm zu schalten und bei Bedarf wieder hinzu zu nehmen. Sequentielles und paralleles Arrangement schließen sich also nicht gegenseitig aus. Tobler (2004) und Aav (2005) nutzen das sequentielle Arrangement auf der Ebene einzelner Tracks, um der Endlosschleifenmusik zu mehr Varianz zu verhelfen. Ganz im Sinne der klassischen Würfelspiele können bestimmte Segmente durch gleichwertige Substitute (ebenfalls vorgefertigte Versatzstücke) ersetzt werden. Anstelle einzelner Melodiepassagen können Varianten gespielt werden, während die Begleitung gleich bleibt oder ebenfalls variiert. Ob das »Original« oder eines von eventuell mehreren Substituten gespielt wird, entscheidet bei jedem Durchlauf der Zufall.

Aufwand und Komplexität beim Komponieren solcher Musiken, ja *musikalischen Landschaften*, erhöhen sich jedoch nocheinmal deutlich gegenüber der rein sequentiellen Lösung (Collins 2008). Nun müssen nicht mehr nur zeitlich aufeinander folgende Segmente aneinander passen. Auch zeitgleich erklingende Tracks müssen übereinander passen, miteinander »harmonieren«, inklusive all ihrer Substitutionsmöglichkeiten. Gerade das Miteinander-Harmonieren verleitet schnell zu musikalischer und ausdrucksmäßiger Ähnlichkeit, der man als Komponist aktiv entgegenwirken muss.

3.3 Ausdrucksvolle Interpretation

Oft bedarf es gar keines Wechsels zu einem grundsätzlich anderen Musikstück, um dem inhaltlichen Wandel einer Szene zu folgen. Das ist insbesondere dann nicht nötig, wenn die Szene an sich noch immer die gleiche ist⁵ und eine grundsätzlich andere Musik einen inhaltlichen Wandel ausdrücken würde, der in dieser Ausprägung so nicht stattfindet. Vielmehr verlangt es nach einer feineren Nuancierung, die nichtsdestoweniger ebenso ausdrücklich dargestellt werden muss.

Zeugnis des wohl frühesten Versuches, das musikalische Geschehen mit dem Spielgeschehen zu verknüpfen, ist der Spieleklassiker *Space Invaders* (Nishikado 1978). Eine sich ständig wiederholende absteigende Viertonfolge illustriert das Herabsinken angreifender Flugobjekte. Erreicht eines von ihnen die Erde, ist das Spiel verloren. Mit dem immer tiefer Herabsinken der Angreifer wird die Spielgeschwindigkeit und gleichermaßen das Tempo der Viertonfolge erhöht. Das sorgt für eine im Spielverlauf zunehmend spannungsgeladene Hektik. Solche Wirkungen wurden u.a. von Bartlett (1996), Sloboda & Juslin (2001), Moffat & Kiegler (2006) untersucht.

Das Tempo ist eines der vielfältigen interpretatorischen Gestaltungsmittel. Die Musik ist nach wie vor vorkomponiert und bleibt in ihrer Substanz unverändert. Sie kann aber in unterschiedlicher Weise dargeboten werden und dadurch in teils sehr unterschiedlichem Lichte erklingen. Die ausdrucksmäßige Bandbreite reicht von sanft und gesänglich bis aggressiv akzentuiert, von ruhig schwebend bis hektisch treibend. Hier kann mit Timing, Dynamik, Artikulation, metrischer Betonung und Akzentuierung, instrumentalen Spieltechniken, akustischen und technischen Effekten gearbeitet werden.

Auch die *Orchestration*, also der gezielte Einsatz der instrumentalen Klangfarben, ist ein interpretatorisches Mittel, mit dem musikalische Merkmale herausgearbeitet oder aber auch maskiert und Kontraste erzeugt, vergrößert bzw. verringert werden können. Die Orchestration ist ein mächtiges Mittel, »um den musikalischen oder stimmungsmäßigen Gehalt oder den generellen Charakter einer Komposition umzusetzen bzw. zu unterstreichen« und »spielt bei der Verdeutlichung der musikalischen Form eine wesentliche Rolle.«⁶ Die systematische Beschäftigung mit der Orchestration begann in der Klassik und wurde

⁵Das kann z.B. bedeuten, dass die gleichen handelnden Personen vor der gleichen Kulisse agieren, aber ihr Verhältnis zueinander sich ändert. In Computerspielen interagiert man oft mit Nicht-Spieler-Charakteren, die abhängig von den Taten des Spielers eine unterschiedliche Gesinnung zu ihm haben, etwa im Spiel *Fable* (Lionhead Studios 2005). Ein anderes Beispiel sind mehrstufige Rätsel. Bis zur vollständigen Lösung bleibt das Setting gleich aber die Spielsituation entwickelt sich mit jeder Teillösung weiter.

⁶Sevsay (2005), Seite 17.

ab der Romantik (1800 bis 1914) zu einer wichtigen kompositorischen Disziplin (Berlioz & Strauss 1844/1904). Sie zählt zum Handwerk des Komponisten, nicht des Interpreten. Da es hier aber darum geht, ein bereits komponiertes Particell⁷ auf ein bestimmtes Instrumentarium zu übertragen, nicht es zu kreieren, ist die Orchestration eher ein interpretatorischer Aspekt, kein kompositorischer. Sie wird deshalb auch oft von Orchestratoren (in gewisser Weise »Kompositionsgehilfen«) durchgeführt, anstatt vom Komponisten selbst – eine gängige Praxis in der Filmmusik (Karlin & Wright 2004).

Hier erweist die Musik-Engine *iMuse* ihre Vielseitigkeit, denn sie hält auch auf diesem Gebiet einige Möglichkeiten bereit. Sie bietet eine Schnittstelle zur Steuerung von Tempo, Intonation, Lautstärke, Instrumentation und Stereo-Panorama (siehe Anhang B). Wingstedt's (2006) Engine *REMUPP* ermöglicht ebenfalls die Beeinflussung von Tempo und Instrumentation⁸ sowie von Artikulation und Hall. Brown & Kerr (2009) gehen noch einen Schritt weiter. Neben Tempo, Dynamik, metrischer Akzentuierung, Artikulation und Klangfarbe beeinflussen sie auch melodische und rhythmische Eigenschaften der Musik. Die Melodiekontur kann entlang der Tonhöhen-Achse gestaucht und gestreckt werden, so dass eine eher enge gesangliche Melodie zu einer weit ausholenden sprunghaften werden kann. Die metrische Prägnanz bzw. rhythmische Dichte kann ausgedünnt werden: Anfangen mit den unbetonten Zählzeiten werden mehr und mehr Events (Töne und rhythmische Impulse) aus der Musik gestrichen und die entstehenden Löcher durch Verlängerung der jeweils vorhergehenden Note ausgefüllt.

Die Arbeit mit interpretatorischen Mitteln ist allerdings nur begrenzt in der Audio-Domäne möglich. Akustische Phänomene wie Hall, Echo, Doppler oder die Dämpfung einiger Frequenzbänder lassen sich noch relativ einfach umsetzen. Niklas Röber (2008) zeigt, dass sich sogar die Graphik-Hardware für das Akustik-Rendering eignet. Auch klangliche Verzerrungen können in Echtzeit auf ein Audio-Signal aufgerechnet werden, wie etwa bei Puckette (2010). In Grenzen ist es sogar möglich, die Rhythmik und Dynamik einer Aufnahme zu verzerren (Grachten, Arcos & de Mántaras 2006, Jehan 2010, Fabiani 2009, 2010). Je differenzierter dies aber geschehen soll, desto unpraktischer erweist sich das Audio-Signal, wie in Abschnitt 1.1 bereits beschrieben.

Hier bietet das MIDI-Format deutlich mehr Flexibilität und Eingriffsmöglichkeiten. Auch *iMuse* und *REMUPP* sind MIDI-basierte Lösungen. Jeder einzelne Ton, jeder einzelne Tonparameter (Datum, Lautstärke, Länge, Ausgabeport und -kanal etc.), das Wiedergabetempo und alle Controller (für Panorama, Abmischung, Hall, zur Steuerung beliebiger Klang- und Effektgeneratoren) können direkt editiert und beliebig verändert werden. Die Palette an Möglichkeiten

⁷die kompositorische Essenz, meist für Klavier zusammengefasst

⁸Ein Wechsel der Instrumentierung lässt sich sehr einfach auch mit parallelem Arrangement (Cross-Fade der Instrumente) umsetzen.

wird nur durch die auf dem jeweiligen System installierte Hard- und Software begrenzt.

Die Arbeit mit MIDI ist allerdings eher Hardware-nah und erfordert viel Detailarbeit. Die interpretatorischen Ausdrucksmittel werden deshalb meist stark vereinfacht. So bleiben Tempo und Lautstärke innerhalb einer Interpretation meist konstant; die Detailarbeit, die bei menschlichen Interpreten beobachtet werden kann, fehlt. Dadurch klingt die Musik eher nüchtern. Desgleichen Artikulation, die nur die Länge der Töne beeinflusst, nicht einmal ihre Lautstärke. Bei einem Wechsel des Interpretationsstiles geschehen Veränderungen sprunghaft ohne Übergang. Wie beim parallelen Arrangement gilt auch hier: An musikalischen Strukturgrenzen klingt das noch akzeptabel, innerhalb der Formteile entsteht jedoch eine deutliche Inkonsistenz. Obwohl die Musik auf kompositorischer Ebene weiterläuft und nur ihre Darbietungsweise sich wandelt, ist ein aus dem Kontext heraus nicht motivierter und unvorbereiteter abrupter Wechsel sehr deutlich als Bruch wahrnehmbar.

Das zu vermeiden, erfordert detailliertes Wissen und Modelle zur Beschreibung ausdrucksvoller Interpretationen. Tatsächlich ist die automatische *Performance-Generierung* ein eigenes Forschungsfeld in der Musikinformatik, das noch immer viele offene, auch grundsätzliche, Fragen bereithält. Die drei etabliertesten Ansätze zur automatischen Generierung solcher Interpretationen sind regelbasiert (Friberg, Bresin & Sundberg 2006), Machine-Learning-basiert (Widmer & Goebel 2004) oder leiten sich aus einer mathematischen Musiktheorie ab (Mazzola, Göller & Müller 2002). Diese Verfahren treffen sozusagen die Entscheidung, wann welche interpretatorischen Mittel zum Einsatz kommen. Grundlagen dafür sind die musikalische Form (Phrasengliederung, melodische Intervallik etc.), eine beabsichtigte emotionale Einfärbung sowie Interpreten-spezifische Muster und Angewohnheiten, die nachgebildet werden sollen.

Die Überführung emotionaler Ausdrucksabsichten in interpretatorische Parameter wird unter anderem von Friberg & Bresin (2000), Friberg (2008), Livingstone (2008), Hoeberechts & Shantz (2009) beschrieben. Dabei werden verschiedene Einstellungen für Tempo, Lautstärke und Artikulationsweise auf Positionen in der *Valenz-Arousal-Ebene*⁹ abgebildet und dazwischen interpoliert.

Verfahren zur Generierung ausdrucksvoller Interpretationen, auch mit sehr unterschiedlichen ausdrucksmäßigen Ausrichtungen, gibt es demnach einige. Dabei sollte nicht vergessen werden, dass der Mensch die Musik auch immer noch manuell produzieren kann. Viele Musiker und Musikproduzenten legen darauf trotz mühevoller Detailarbeit großen Wert, schließlich können sie hier der Interpretation ihr ganz persönliches Gepräge geben. Zudem sind sie oft mit der grundsätzlichen Gestalt der generierten Ausdrucksmittel unzufrieden – ein

⁹ein sehr beliebtes Modell zur Diskriminierung von emotionalen Ausprägungen (Russell 1980)

Crescendo beginnt zu spät eine Temposchwankung wirkt zu neutral etc. Diese grundsätzliche Gestalt interpretatorischer Ausdrucksmittel war bislang kaum Gegenstand genauerer Untersuchungen, auch nicht die Freiheitsgrade, die der Interpret bei ihrer Umsetzung hat. Für Temposchwankungen hat sich eine aus der *Stopping Runners*-Theorie abgeleitete Kurven-Charakteristik etabliert (Friberg & Sundberg 1999). Artikulation wird über die Veränderung der Tonlänge hinaus kaum betrachtet. Die Werkzeuge, mit denen Musikproduzenten hier arbeiten, sind dementsprechend unpraktisch. Diese Lücke soll Kapitel 5 schließen. Die dort entwickelten Modelle erlauben einen höheren Detailgrad bis hinein in subtilste Nuancen. Deren editorischer Aufwand ist dabei deutlich überschaubarer, wenn auch nach wie vor groß.

Darüber hinaus fehlen Verfahren, die verschiedene Interpretationsstile und Orchestrationen während der Wiedergabe musikalisch schlüssig überleiten – genau das, was benötigt wird, um einer interaktiven Szene zu folgen. Spätestens hier, wo die Musik beim Endanwender läuft und durch Interaktion beeinflusst wird, ist der Mensch als Produzent machtlos. Hier können nur noch automatische Verfahren helfen. Auf Grundlage der in Kapitel 5 erarbeiteten Modelle werden in Kapitel 6 entsprechende automatische Überleitungstechniken vorgestellt, die diese Lücke schließen.

3.4 Reharmonisierung

Die Melodieharmonisierung ist ein klassisches Problem der Musikinformatik. Zu einer gegebenen Tonfolge soll eine Akkordfolge gefunden und mehrstimmig ausgesetzt werden. Lösungsansätze hierfür gibt es einige, etwa von Schottstaedt (1989), Ebcioğlu (1992), Hörnel (2000), Pachet & Roy (2001), Verbiest, Cornelis & Saeys (2009). Bei der Reharmonisierung geht es nun darum, die bestehende Aussetzung an eine neue Akkordfolge anzupassen. Auf diese Weise kann sowohl subtile Varianz in Endlosschleifen erzeugt als auch die Stimmung eines Musikstücks verändert werden. Man denke nur an das beliebte Beispiel einer fröhlichen Dur-Melodie, die »vermollt« wird und dadurch eine dunklere pessimistischere Einfärbung erhält. Durch mehr oder weniger unauffällige Anpassungen der Stimmführungen können Geschlecht und Komplexität der zugrunde liegenden Akkorde und sogar die Akkorde als ganzes verändert werden. Das melodische Material des Musikstücks bleibt dabei wiedererkennbar.

Yoon & Lee (2006) stellen einen Planungsansatz für affektive Reharmonisierungen vor. Um einem Musikstück einen anderen Spannungsverlauf zu geben, werden zunächst die Spannungspotentiale der Akkorde berechnet. Verfahren dafür sind etwa von Krumhansl & Shepard (1979), Lerdahl (1988, 1996), Chew (2000)

bekannt. Nun können die Akkorde lokal durch Alternativen mit anderen Spannungspotentialen substituiert werden.

Um auch die Stimmführungen im polyphonen Satz an die neue Harmonisierung anzupassen, entwickelten Livingstone (2008) und Stenzel (2005) je einen Lösungsansatz. Livingstone verändert das Geschlecht der zugrunde liegenden Tonart. Um Dur zu Moll zu wandeln, werden in einer jeden Stimme Terz und Sext der Tonart um einen Halbton nach unten versetzt¹⁰. Das Prinzip lässt sich auch für andere Tonartgeschlechter, etwa Kirchentonarten, anwenden. In einem Prototyp seines Systems *CMERS* hatte Livingstone auch die Veränderung der harmonischen Komplexität durch Hinzunahme oder Wegfall von Akkordtönen vorgesehen, später jedoch darauf verzichtet: Zur Erhöhung der Komplexität müssen neue Akkordtöne eingeführt werden, jedoch welche? Welche der wenigen Stimmen im polyphonen Satz soll die neuen Töne dann spielen und dafür auf andere Töne im Akkord verzichten?

Bei Wingstedt et al. (2005) wurde das Problem vom Komponisten¹¹ selbst gelöst. Mehrere Versionen mit unterschiedlich komplexer Harmonik waren hier komponiert. Dafür wurden zusätzliche Stimmen eingeblendet (paralleles Arrangement). Automatische Ansätze, um das zu generieren, gibt es noch nicht. Selbst die in Livingstone's Prototyp tatsächlich implementierte Harmonievereinfachung (Wegfall von Akkordtönen) führte zu teils extremen Änderungen in den Stimmführungen. Das betraf auch die Melodiestimme. Die Resultate bezeichnete Livingstone selbst als »unmusikalisch«.

Stenzel (2005) ändert sogar die komplette Harmonisierung und unterlegt den Satz mit völlig neuen Akkorden. Zur Anpassung der Stimmführungen werden unpassende Töne um ein kleinstmögliches Intervall verschoben. Dieser Ansatz ist unabhängig von Geschlecht und Komplexität der neuen Akkorde. Da mit jedem veränderten Ton auch alle seine nachfolgenden Töne verschoben werden, nimmt die Melodieverfremdung hier wie ein numerischer Fehler immer weiter zu.

Hier zeigt sich das Kernproblem bei der Reharmonisierung: die unvermeidliche Verzerrung der Stimmführungen. Sie wird insbesondere dann als befremdend empfunden, wenn melodische Figuren ihre charakteristischen Eigenschaften verlieren oder die Reharmonisierung ihnen gestisch entgegen steht.

¹⁰Ein Beispiel: Beim Wandel von C-Dur zu c-Moll würden in jeder Stimme die Töne *e* zu *es* und *a* zu *as* geändert werden. Theoretisch beinhaltet ein solcher Wandel auch die Veränderung der Sept (*h* zu *b*). Die kann aber nicht pauschal erfolgen, weshalb Livingstone generell darauf verzichtet.

¹¹Stefan Lindberg komponierte die Demo-Musiken.

3.5 Komponierte Überleitung

Um zwischen zwei vorgefertigten Musikstücken zu wechseln, wurde bisher hart oder weich geschnitten, was im Allgemeinen keine akzeptable Lösung ist, wie in Kapitel 1 ausgeführt. Die Musiken müssen sauber übergeleitet werden. Das sequentielle Arrangement bietet dazu eine gute Möglichkeit, die aber mit erheblichem Aufwand verbunden ist, denn die Überleitungen müssen alle vorkomponiert werden (siehe Abschnitt 3.1). Die Überlegung liegt nahe, genau diese Aufgabe durch ein automatisches Verfahren ausführen zu lassen. Es bräuchten nicht länger alle nur denkbaren Überleitungen vorbereitet werden, vielmehr könnte in Echtzeit genau die eine gerade benötigte generiert werden. Das Ziel ist also, aus der bei der Wiedergabe gerade erreichten musikalischen Situation heraus möglichst organisch in die Zielmusik überzuleiten.

Ein erster Ansatz, der Tonhöhen und Rhythmus zweier Melodien linear über ein gewisses Zeitintervall interpoliert, wurde von Mathews & Rosler (1968) beschrieben. Die *Music Morphing*-Algorithmen von Wooller & Brown (2005) greifen diese Idee wieder auf und ergänzen sie um zwei weitere Materialvermischungstechniken.

- Unterschiedlich lange Fragmente von Ausgangs- und Zielmusik werden sequentiell miteinander vermischt. Das geschieht in Abhängigkeit von der Position innerhalb der Überleitung. Zu Beginn dominiert die Ausgangsmusik. Zum Ende der Überleitung hin wird das Material der Zielmusik bevorzugt.
- Aus den Melodien von Anfangs- und Zielmusik wird je eine Markov-Kette konstruiert. Mit deren Hilfe wird dann eine Übergangsmelodie generiert. Dabei werden die Markov-Ketten abwechselnd angewendet. Die Dauer der Anwendung ist gewichtet. Von Beginn bis Ende der Überleitung verschiebt sich die Gewichtung linear von der Ausgangsmusik hin zu Gunsten der Zielmusik.

Ein genetisches Verfahren, das mehrere Eingabe-Musiken miteinander (re-)kombiniert, stellt Gartland-Jones (2003) mit *MusicBlox* vor. Eine Fitness-Funktion bewertet den Grad der Verwandtschaft zwischen der Lösung und den originalen Musiken. Ist sie bspw. nahe an Musikstück *A*, sollte aber eher zu Musikstück *B* tendieren, wird sie erneut mit *B* kombiniert oder anderweitig mutiert. Obwohl Gartland-Jones das nicht zum Erzeugen von Überleitungen verwendet, kann es gerade im Kontext der *Music Morphing*-Techniken von Nutzen sein, wie Brown, Wooller & Miranda (2009) zeigen.

Alle diese Ansätze eignen sich vorzugsweise im Kontext serieller Musik, da sie mit Konstruktionsprinzipien arbeiten, die dieser speziellen Stilgattung zueigen sind. Da andere Stile auf anderen ästhetischen Prinzipien beruhen, sind diese Verfahren dort eher ungeeignet. Das stellten schon Mathews & Rosler (1968) fest, die ihre Interpolationstechnik mit zwei Volksliedmelodien ausprobierten und das Ergebnis als »nauseating« und »jarring«¹² bezeichneten. Hier wird deutlich, wie schwer es ist, allgemeingültige Lösungen zu finden.

3.6 Generative Musik

Die letztendliche Abkehr von Arrangement und Bearbeitung vorkomponierter Musik stellt schließlich die automatisch generierte Musik dar. Dies ist heute in Echtzeit, d.h. während ihrer Wiedergabe, möglich. Durch Interaktion kann die Parametrisierung des Kompositionsprozesses beeinflusst werden. Dadurch wird mittelbar Einfluss auf die erzeugten musikalischen Strukturen und Ausdruckscharakter genommen.

Bei Wallis, Ingalls & Campana (2008) sind es die Parameter Tempo, rhythmische Komplexität, Artikulation, Tonartgeschlecht, harmonische Komplexität und Registerlage, durch die der affektive Gehalt der generierten Musik gesteuert wird. Das System *Herman* (de Quincey 1998, Stapleford 1998, Robertson, de Quincey, Stapleford & Wiggins 1998) bezieht auch Melodik und Form der Musik mit ein. Bei Hoeberechts, Demopoulos & Katchabaw (2007), Hoeberechts & Shantz (2009) wird zunächst ein »neutrales« musikalisches Material generiert, das dann gemäß der gewünschten emotionalen Prägung adaptiert wird. Dazu werden Tonart, Tempo, Tonhöhe, Grad der Konsonanz/Dissonanz, Lautstärke und Artikulation entsprechend angepasst. Casella's (2002) Agenten-basiertes System *MAGentA* schaltet zwischen grundverschiedenen Kompositionsalgorithmen um und editiert ihre Parameter.

Keiner dieser Ansätze beinhaltet jedoch Verfahren, um zwischen den verschiedenen Algorithmen bzw. Stilen überzuleiten. Die direkte Änderung der Parameter führt zu einem abrupten Stilwechsel, der nur bei ähnlichen Stilen akzeptabel ist. Graduelle Parameteränderungen funktionieren innerhalb ein und des selben Kompositionsalgorithmus' meist gut. Hoeberechts & Shantz (2009) beschreiben bspw. ein Verfahren, um die Parameterausprägungen für »Mischemotionen« zu errechnen, was dann dafür verwendet wird, Musik mit dem entsprechenden Ausdruckscharakter zu generieren bzw. mit interpretatorischen Mitteln emotional einzufärben. Wooller & Brown (2005) stellen mit dem *Markov Morphing*

¹²Mathews & Rosler (1968), Seite 103.

ein Verfahren vor, mit dem zwischen verschiedenen Markov-Ketten¹³ übergeleitet wird, während die Musik generiert wird. Problematisch wird es, wenn verschiedene Algorithmen ineinander übergeleitet werden sollen, sie aber keine äquivalenten Parameter besitzen, die sich aufeinander abbilden lassen. Hier könnten Konzepte des stilistischen Pluralismus (siehe Anhang C) zu möglichen Lösungsansätzen führen. Noch gibt es aber keine Lösung für dieses Problem.

Auch die direkte Interaktion mit dem Kompositionsprozess ähnlich wie der Umgang mit einem Musikinstrument ist möglich. Diese besondere Form ist als *Live Coding* bekannt (Brown & Sorensen 2009). Hier werden die Programme, welche Musik generieren, in Echtzeit geschrieben und modifiziert. Bei Konzertaufführungen wird der Quelltext, den der Musiker kontinuierlich bearbeitet, oft auch für das Publikum sichtbar auf eine Leinwand projiziert. Diese Art der Interaktion kann vom Spieler eines Videospiele freilich nicht erwartet werden. Eine Spiele-Engine könnte hierfür aber vorgefertigte Scripte bereithalten. Vor allem im Zusammenhang mit Musikspielen besteht hier großes Potential (siehe Kapitel 7).

3.7 Zusammenfassung

Die Möglichkeiten zur Nichtlinearität in Musik sind enorm vielseitig. Sie reichen vom Arrangement vorgefertigter Musik über deren Bearbeitung bis hin zur Generierung von immer neuer Musik. Entlang dieser Achse nehmen auch Flexibilität und Wandelbarkeit der Musik zu. In gleichem Maße nimmt die Bedeutung der eher statischen Audioformate (Wave, MP3, Ogg Vorbis etc.) als Träger der musikalischen Informationen ab. Flexiblere im musikalischen Sinne besser editierbare Formate werden wichtiger, wie etwa der MIDI-Standard (MIDI Manufacturers Association 2001), *cSound* score (Boulanger 2000) oder Pattern-Sprachen (McLean & Wiggins 2010).

Sie setzen jedoch eine Resource voraus, die oft nicht gegeben ist, eine qualitativ einheitliche Tonerzeugung auf allen Systemen. So verschieden, wie die Soundkarten von PC zu PC sind, unterscheiden sich auch die mitgelieferten MIDI-Klänge. Deren Qualität ist im Heimcomputer-Bereich ohnehin unzureichend.¹⁴ Manche Soundkarten verzichten sogar ganz auf eine MIDI-Schnittstelle. Hier sind Software-Synthesizer oder -Sampler erforderlich, die etwa als Teil einer

¹³Markov-Ketten sind ein beliebtes Mittel, um melodische, rhythmische und harmonische Progressionen zu analysieren und beim automatischen Komponieren zu reproduzieren (Ames 1989). Sie kamen schon bei der Entstehung der berühmten *Illiad-Suite* von Hiller & Isaacson (1959) zum Einsatz und haben sich bis heute als ein wichtiges Werkzeug gehalten.

¹⁴Daher kommt das verbreitete Vorurteil, MIDI klinge schlecht. Tatsächlich ist es die Soundkarte, welche schlechte Klänge produziert. MIDI selbst klingt gar nicht, denn es sind nur Steuersignale.

Spieleapplikation mit installiert werden müssten. Im Falle des *cSound*-Synthesizers werden die Klangdefinitionen als ASCII-Daten (die sog. *orchestra files*) mitgeliefert. Auch Sample-Daten müssten zusammen mit dem Sampler installiert werden. Das ist heute nicht gängige Praxis, denn es ist mit zusätzlichen Produktionskosten bzw. Lizenzgebühren verbunden. Gute Sample-Bibliotheken sind zudem nicht nur teuer sondern auch umfangreich und beanspruchen große Speicherressourcen. Vor diesem Hintergrund wundert es nicht, wenn Spieleentwickler die einfachere Lösung als das »geringere Übel« ansehen und bei den qualitativ hochwertigen, wenn auch unflexiblen, Audioaufnahmen bleiben.

In der Welt der Spielekonsolen sieht das anders aus. Hier sind einheitliche Hard- und Software-Ausstattung gewährleistet. Klangsynthese und /oder Sample-Bibliothek sind vorinstalliert und von teils guter Qualität. Hier hat sich der MIDI-Standard erhalten und mit ihm seine flexiblen Möglichkeiten. Collins (2008) macht deutlich, dass die Möglichkeiten von Audioformaten ausgeschöpft und ihre Grenzen erreicht sind. Sie schreibt: »The potential for adaptability is far more inherent in the MIDI format, and there are some functions of MIDI that are not possible with wave files [...]. As MIDI expands capabilities with advanced downloadable sounds [...], MIDI may return as a force to be reckoned with in game music. [...] Now that wavetable samples have reached a high fidelity, MIDI has the advantage over standard Redbook [...].«¹⁵ Auch für Page & Kelly (2007)¹⁶ steht für das lange Zeit als veraltet geltende MIDI-Format eine Renaissance bevor.

3.8 Einordnung der weiteren Beiträge dieser Arbeit

In dieser Arbeit kommen beide, sowohl Audio als auch MIDI, zum Einsatz. Die Audio-Domäne beschränkt sich naturgemäß auf sequentielles und paralleles Arrangement, deren Möglichkeiten oft kaum ausgeschöpft werden (siehe Abschnitte 3.1 und 3.2). Collins (2009) beschreibt, wie einfach die Kontroll-Logik zur Steuerung der Musik in Computerspielen meist gehalten ist. Damit lassen sich komplexere Musikszenarien nur sehr schwer modellieren. Hierfür wird in Kapitel 4 ein Arrangement-Mechanismus entwickelt, der beide Arrangement-Typen miteinander und mit dem interaktiven Szenarium verbindet.

Die MIDI-Domäne muss hingegen den Ruf des »Sterilen« überwinden. Wo Menschen Musik machen, spiele der Computer nur Note für Note mechanisch her-

¹⁵Collins (2008), Seite 147.

¹⁶Jason Page ist Audio Manager bei *Sony Computer Entertainment Europe*.

unter. Kapitel 5 widmet sich deshalb der Modellierung ausdrucksvoller Interpretationen. Auf dieser Grundlage werden in Kapitel 6 Techniken entwickelt, um verschiedene Interpretationsstile interaktiv und musikalisch schlüssig ineinander überzuleiten. Dieser Ansatz beschränkt sich nicht nur auf die Arbeit mit interpretatorischen Ausdrucksmitteln. Er schließt auch das parallele Arrangement mit ein und erlaubt einfache Reharmonisierungen und Verzierungen.

Ein unterschätztes Problemfeld bildet die diegetische Musik, also die Musik, die innerhalb der Szene spielt (z.B. ein Straßenmusikant, ein Radio etc.). Um diese angemessen zu inszenieren, sind einige Besonderheiten zu beachten, die bei der »normalen« nicht-diegetischen Musik keine Rolle spielen. Kapitel 7 zeigt, wie hierfür das ganze Instrumentarium der nichtlinearen Musik genutzt werden kann.

Die Musik-Engine als Echtzeit-Arrangeur

SEQUENTIELLES und paralleles Arrangement werden in diesem Kapitel zu einem gesamtheitlichen mehrdimensionalen Arrangementkonzept verbunden. Dem liegen folgende Veröffentlichungen zu Grunde:

- (Berndt, Hartmann, Röber & Masuch 2006),
- (Berndt & Hartmann 2006),
- (Berndt 2008).

Beim sequentiellen Arrangement wird die zeitliche Abfolge vorgefertigter musikalischer Teilstücke bearbeitet. Welches Teilstück als nächstes folgen soll, kann bis zum Ablauf des aktuell spielenden neu entschieden werden. Die Musik kann bei ihrem derzeitigen Ausdruckscharakter verweilen oder sich nach und nach mit jedem weiteren Teilstück weiter entwickeln, vom Ausgangspunkt entfernen und einen von vielen möglichen Wegen beschreiten. Solche Veränderungen können allerdings nur von Teilstück zu Teilstück geschehen, nicht innerhalb eines solchen. Das macht das sequentielle Arrangement schwerfällig. Erst Sekunden nach der auslösenden Interaktion reagiert die Musik.

Ganz anders das parallele Arrangement. Die Musik unterteilt sich in mehrere synchron spielende Spuren (auch Tracks oder Kanäle). Jede davon steuert ein eigenes musikalisches Material, eine eigene Melodie, Rhythmik oder auch nur Klangfarbe bei. Unterschiedliche Kombinationen können im Ausdruckscharakter sehr verschieden ausfallen. Ein für sich allein heller Flötensatz kann durch massive Akkorde im tiefen Blech angereichert und durch hohe Streicher verstärkt werden. Eine schlichte Melodie kann durch eine filigrane Alternative ersetzt werden. Solche Änderungen lassen sich durch sanfte Überblenden direkt

und musikalisch nahtlos umsetzen. Die Latenz zwischen Interaktion und musikalischer Reaktion ist gleich Null. Im Gegensatz zum sequentiellen Arrangement ist das parallele Arrangement also sehr reaktiv. Damit alle Kombinationen und Überblenden musikalisch überzeugen, müssen die Spuren aber harmonisch und metrisch zueinander passen. Zwar kann der musikalische Rahmen, der dadurch gesteckt ist, trotzdem sehr weit reichen, in der Praxis ist die Tendenz zu formaler Ähnlichkeit trotzdem sehr stark. So weit abschweifen wie das sequentielle Arrangement kann das parallele Arrangement nicht.

Beide Ansätze haben also ihre Vor- und Nachteile, denen sie jeder für sich unterliegen. Sie lassen sich aber derart miteinander verknüpfen, dass die Vorteile des einen die Nachteile des anderen ausgleichen. Das reaktivere parallele Arrangement wird zum Werkzeug für kurzfristige Wandlungsprozesse und überbrückt die Latenz, bis das sequentielle Arrangement greift und zu grundlegend neuer Musik führt. Diese Verbindung beider Arrangement-Prinzipien wird im folgenden detailliert dargestellt.

4.1 Verteilte Musik

Eine derart enge Verknüpfung von sequentiellem und parallelem Arrangement muss mit der Verschiedenheit der beiden Prinzipien umgehen. Das sequentielle Arrangement ist diskret. Nur an bestimmten Zeitpunkten wird über den Fortgang des musikalischen Ablaufes entschieden. Das parallele Arrangement ist hingegen kontinuierlich. Die Abmischung der Kanäle kann und muss ständig aktualisiert werden, sonst verkommen Überblenden zu groben Lautstärke-Stufen. Um dieser Gegensätzlichkeit zu begegnen, lohnt es, zunächst einen Blick auf die Topologie der zu vertonenden interaktiven Szenarien zu werfen. Auch hier finden sich sowohl diskrete als auch kontinuierliche Ordnungen.

Ein typisches Beispiel für eine diskrete Topologie ist das Menü, das sich in verschiedene Untermenüs gliedert. Musik kommt zum Einsatz, um die jeweilige Position im Menübaum zu repräsentieren. Jedes Untermenü hat seine eigene Hintergrundmusik. Gerade in DVD-Menüs wird das exzessiv praktiziert. Ein weiteres Beispiel für diskrete Topologien sind die verzweigten Story-Graphen in Computerspielen, deren Knoten etwa Handlungssituationen repräsentieren und deren Kanten den Fortgang der Handlung.

Kontinuierliche Topologien finden sich zumeist in der Anordnung von Regionen und Objekten in einer virtuellen Szene. Sie kann eindimensional sein, wie in klassischen Jump-and-Run-Spielen, in denen der Spieler sich vom meist links gelegenen Beginn des Levels zum rechten Ende durcharbeiten muss, wie etwa

im Klassiker *Super Mario Bros.* (Nintendo EAD 1985). Kontinuierliche Spielwelten können auch zweidimensional sein, wie in klassischen Strategiespielen und Adventures (Walter 2002). Dreidimensionale Spielwelten sind in 3d-Actionspielen und Flugsimulationen anzutreffen. Musik ist dabei meist Repräsentant für Orte, Objekte und Nicht-Spieler-Charaktere. Wenn sich die Spielfigur ihnen nähert, oder sie sich der Spielfigur nähern (z.B.: feindliche Einheiten umzingeln die Basis des Spielers), soll dies auch mit einer musikalischen Reaktion einhergehen.

In Spielen sind meist beide Topologietypen gleichzeitig präsent, etwa eine dreidimensionale Spielwelt mit ortsgebundener Musik und eine Handlungsebene, die durch Aktionen des Spielers vorangetrieben wird. Musikalisch kann das bedeuten, dass jede Handlungssituation ihre eigenen ortsgebundenen Musiken mitbringt. Ein vormals freundlicher idyllischer Ort kann nach Fortschreiten der Handlung zu einem gefährlichen werden. Ein vormals feindseliger Nicht-Spieler-Charakter kann im weiteren Verlauf zu einem guten Freund werden usw.

Damit eine Musik-Engine all diesen Konstellationen gerecht werden kann, müssen die jeweiligen Topologien auf den Arrangement-Mechanismus übertragen werden. Karen Collins' (2009) Kritik an der extremen Einfachheit der musikalischen Kontroll-Logik und Forderung nach angemesseneren Lösungen findet genau hier ihren Niederschlag. Die Standard-Funktionen eines handelsüblichen CD-Players (Play, Stop, Skip) sind vollkommen ungenügend. Auch die Sequencer-Metapher, die musikalische Objekte (Patterns) entlang der Zeitachse bestimmten Tonspuren zuweist, ist für die Vertonung komplexer Szenarien eher ungeeignet, denn sie repräsentiert nur inner-musikalische Ordnungen aber überträgt sie nicht auf topologische Ordnungen.

Viel intuitiver ist es, die Musiken direkt in der Szene durch die Platzierung von Soundquellen zu repräsentieren (siehe Abbildung 4.1), eine Praxis, die für das Authoring von Geräuschkulissen längst etabliert ist (Röber, Deutschmann & Masuch 2006). Jede Soundquelle repräsentiert eine Tonspur im Sequencer, spielt diese synchron mit allen anderen ab. Ähnlich der physikalischen Realität ist jede Soundquelle in einem jeweiligen Radius hörbar. Bewegt sich die Spielfigur auf eine Soundquelle zu, blendet diese ein und umgekehrt. In dieser Konstellation kann das parallele Arrangement komplett durch gängige Audio-Engines, wie *OpenAL* (Loki Software & Creative Technology 2009), ausgeführt werden.

Diese Anordnung von musikalischen Soundquellen heißt *Musikszene*. Eine Musikszene ist also definiert als eine Menge von Soundquellen, denen wiederum je eine Tonspur zugeordnet ist. Jede Tonspur ist ein Stream von musikalischen Segmenten, der in Echtzeit arrangiert wird. Hier geschieht das sequentielle Arrangement. Um dieses zu steuern, werden zusätzliche Marken, die sog. *Musikwechselmarken*, in die Musikszene eingeführt. Das sind Polygone (ebenso we-

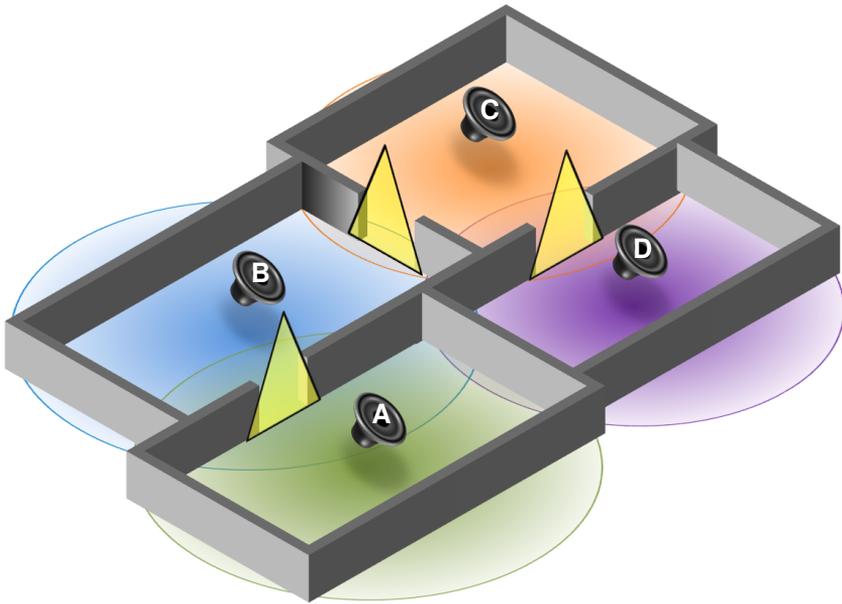


Abbildung 4.1: In einem Gebäudegrundriss sind vier Soundquellen platziert. Jede hat ihren eigenen Hörbarkeits-Radius. Auf den Durchgängen zwischen den Räumen sind Musikwechselmarken platziert.

nig Teil der visuell sichtbaren Szene wie die Soundquellen), die auf den Übergangsstellen in der Szenentopologie platziert werden, bspw. auf einer Tür, die in einen anderen Raum führt. Mittels Kollisionserkennung wird bestimmt, ob und in welcher Richtung der Spieler die Marke durchschreitet. Betritt oder verlässt er den Raum? Je nach Richtung werden dann neue Sequenzen von musikalischen Segmenten in eine jede Tonspur geladen. Sie ersetzen die Sequenzen, die auf die gerade spielenden Segmente ursprünglich folgen sollten.

Die Sequenz läuft als Endlosschleife; ausgenommen sind diejenigen Segmente, die als *OneShot* definiert sind. Sie werden nur beim ersten Durchlauf gespielt. *OneShot*-Segmente können z.B. für Überleitungen verwendet werden, die innerhalb der Schleife nicht mehr nötig sind. Sie können aber auch mitten in den Ablauf eingefügt werden. Der erste Durchlauf kann sehr einprägsame motivische Segmente enthalten, die sich bei mehrfacher Wiederholung im Ohr des Hörers schnell verbrauchen würden. Jede weitere Iteration verzichtet deshalb auf sie und wirkt wie ein Rearrangement.

Eine Musikwechselmarke definiert also nicht nur ihre Position und Form, sondern auch zwei Sequenzen, eine für jede Richtung des Durchschreitens. Von der Spielhandlung getriebene Vorgänge müssen aber nicht zwangsläufig mit Bewegungen der Spielfigur einhergehen. Der Spieler kann still stehen, trotzdem läuft die Handlung weiter. Das sequentielle Arrangement muss deshalb auch durch die Spiele-Engine selbst gesteuert werden können. Hierfür wird der *StoryState* eingeführt. Er definiert je eine Musikszene mit all ihren Eigenschaften und Verhaltensweisen. Ein Wechsel des *StoryStates* ändert nicht nur die Inhalte der Tonspuren in beschriebener Weise. Er platziert Soundquellen neu, setzt ihre Hörbarkeits-Radien neu, erzeugt neue Musikwechselmarken mit neuen Inhalten und kann auch den Spieler/Hörer neu in der Musikszene platzieren.

4.2 Komposition eines interaktiven Musikübergangs

Im Rahmen des beschriebenen Arrangement-Konzepts schreibt man als Komponist nicht mehr in sich abgeschlossene Musikstücke im traditionellen Sinne, sondern Musikszenen, musikalische Landschaften, in denen der Spieler sich frei bewegen kann. Die Musik verteilt sich auf die *StoryStates*, die Musikwechselmarken, und den bis zu dreidimensionalen Raum. Musik hierfür zu planen und zu komponieren, ist zunächst ungewohnt und erfordert etwas Beschäftigung und Eingewöhnung.

Typischerweise wählt man die Soundquellen der topologischen Region (z.B. des Raumes), in welcher sich der Spieler z.Z. befindet, als die primären Soundquellen, ihre Musik als die *Primärmusik* aus. Die Soundquellen anderer Regionen spielen *periphere Musik*. Das sind an die aktuell primäre Musik angepasste Varianten deren eigener Primärmusik. Sie sind klangfarblich und thematisch eigenständig, fügen sich aber in den harmonischen und metrischen Rahmen der derzeitigen Primärmusik ein. Nähert sich der Spieler einer solchen Region, blendet ihr Material ein, kündigt den möglicherweise bevorstehenden Musikwechsel an.

Erst wenn der Spieler die andere Region tatsächlich betritt, die entsprechende Musikwechselmarke auslöst und das sequentielle Arrangement den Musikwechsel vollzieht, wird vom vormals peripheren Material zur neuen Primärmusik umgeschaltet. Die spielt nun aus den entsprechenden Soundquellen der neu betretenen Region. Die Soundquellen der vorherigen Region spielen an die neue Primärmusik angepasste Varianten ihrer nun nicht mehr primären Musik.

Am Beispiel von Abbildung 4.1 verläuft das anschaulich wie folgt. Der Spieler befindet sich in Raum *A*. Damit ist *A* die Primärmusik. Alle anderen Räume

spielen an A angepasste Varianten B_A, C_A, D_A . Da Soundquelle C von Raum A aus überhaupt nicht hörbar ist, kann sie sogar schweigen. Nähert sich der Spieler Raum B , blendet B_A ein. Wenn er Raum B betritt, also die dortige Musikwechselmarke durchschreitet, wird das sequentielle Arrangement ausgeführt. B wird zur Primärmusik. Alle anderen spielen daran angepasste Varianten A_B, C_B, D_B .

Die Latenz, die das sequentielle Arrangement hat, wird vom parallelen Arrangement überbrückt. Der Übergang zur neuen Musik ist mit dem Einblenden bereits in vollem Gange. Wenn schließlich zur neuen Primärmusik gewechselt wird, schließt sich damit der Prozess der Überleitung nur noch in letzter Konsequenz ab. Weiterhin macht es das sequentielle Arrangement möglich, dass der musikalische Wandel nicht in den harmonischen und metrischen Grenzen einer einzigen Primärmusik gefangen bleibt. Damit ist die musikalische Ähnlichkeit, zu der das parallele Arrangement oft zwingt, überwunden.

Als Komponist muss man lernen, mit eben diesen Wechselwirkungen zu spielen. Im Einzelnen mögen musikalische Würfelspiele und Baukastensätze nicht unbekannt sein. In ihrer Kombination stellen sie für die Planung der Musikdramaturgie aber anfänglich ungewohnte Herausforderungen.

4.3 Musik, nicht Physik

Diegetische Musik, also Musik, deren Klangquelle Teil des Szeneninventars ist (z.B. ein Radio oder Straßenmusikanten), kann innerhalb der Diegese wahrgenommen werden. Sie unterliegt den akustischen Gesetzmäßigkeiten der Szene. In einem sehr halligen Raum klingt auch die Musik aus dem Radio sehr verhallt. Ganz anders verhält es sich mit der nicht-diegetischen Musik. Ebenso wie die Erzählerstimme aus dem »Off« ist auch die Musik aus dem »Off« nicht Teil der Diegese und nur vom Publikum wahrnehmbar, nicht von den Charakteren in der Szene. Sie unterliegt nicht ihren akustischen Gegebenheiten. Sie ist kein Soundeffekt, kein Teil der Geräuschkulisse, sondern eine eigene mediale Bedeutungsebene. Im Film wie im Computerspiel präsentiert sich diese Musik auf dem Niveau einer Studioproduktion. Klangliche und akustische Effekte, Stereo- bzw. Surround-Panorama werden allein nach musikalischen und dramaturgischen Erwägungen eingesetzt.

Das bedeutet, dass entsprechende Effekte der Audio-Engine nicht unüberlegt auf die Musik angewendet werden dürfen. Die Positionierung der Soundquellen im Surround-Panorama kann ein hilfreicher Effekt sein, der dem Spieler die Orientierung in einer komplexen Szene erleichtert. Er hört an der Musik, in welcher Richtung sich was befindet. Lässt sich dabei die Position der Soundquelle

aber zu genau orten, wird sie zwangsläufig zu einem diegetischen Gegenstand, zum Objekt in der Szene: »Hier, an dieser Stelle spielt die Musik.« Das darf nicht passieren. Als Orientierungshilfe mag es schon ausreichen, dass der entsprechende Teil der Musik einblendet, wenn der Spieler sich in dessen Richtung bewegt, so dass auf die Lokalisierung im Panorama verzichtet werden kann.

Doch auch das Ein- und Ausblenden an sich folgt im musikalischen Kontext anderen als den physikalischen Gesetzen. Bei zunehmender Entfernung des Hörsers von der Soundquelle verklingt das von ihr emittierte Geräusch. Dieses Abklingen, also die Absorption der Schallintensität im Übertragungsmedium, wird in der Akustik für ebene Wellen wie folgt beschrieben (Kuttruff 2000, Kurzweil, Frenzel & Gebhard 2008).

$$I(d) = I_0 \cdot e^{-md}$$

mit I_0 als Schallintensität direkt an der Soundquelle, m als Dämpfungskoeffizient und d der Entfernung von der Soundquelle. Eine »physikalisch plausible« Einblende würde vorsichtig beginnen, sich stetig schneller steigern und geradezu plötzlich die Ziellautstärke erreichen. So etwas ist nicht nur unmusikalisch sondern auch auffällig. Es katapultiert die Musik auf die Ebene der bewussten Wahrnehmung. Eine Ausblende würde entsprechend umgekehrt verlaufen, plötzlich leiser werden, später nur noch ganz langsam. Statt eines physikalisch plausiblen Verlaufes wäre für musikalische Blenden ein gleichmäßigerer Verlauf wünschenswert. Unter den Abklingmodellen, welche die Audio-Engine *OpenAL* (Hiebert 2006) implementiert, befindet sich eine mögliche Alternative, ein lineares Abklingmodell:

Inverse Distance Model

$$\text{gain}_{\text{inv}}(d) = \frac{d_{\text{ref}}}{d_{\text{ref}} + m(d - d_{\text{ref}})}$$

Exponential Distance Model

$$\text{gain}_{\text{exp}}(d) = \left(\frac{d}{d_{\text{ref}}} \right)^{-m}$$

Linear Distance Model

$$\text{gain}_{\text{lin}}(d) = 1 - m \frac{d - d_{\text{ref}}}{d_{\text{max}} - d_{\text{ref}}}$$

Die Referenzentfernung d_{ref} ist die Entfernung, in welcher eine gewisse Referenzlautstärke erreicht ist. Üblicher Weise werden unterhalb dieser Referenzentfernung d_{ref} und oberhalb einer bestimmten maximalen Entfernung d_{max} nur

noch konstante Lautstärkewerte (g_{min} bzw. g_{max}) gesetzt, um numerischen Problemen und unnötigem Berechnungsaufwand aus dem Weg zu gehen.

Das lineare Abklingmodell ist eine gangbare Alternative für musikalische Ein- und Ausblenden. Es wirkt allerdings sehr mechanisch. Das wird insbesondere durch die Knicke (Undifferenzierbarkeiten) an Beginn und Ende der Blende beim Übergang in den konstanten Verlauf verursacht (siehe in Abbildung 4.2 die Kurve mit Dämpfungskoeffizient $m = 1$). Was es eigentlich nachzubilden gilt, ist die Handbewegung des Tontechnikers am Mischpult, wenn er den Fader eines Kanals sanft und gleichmäßig zieht. In der Computeranimation sind für die Beschreibung solcher Bewegungen sigmoidale Charakteristiken etabliert, bzw. Glockenkurven für die Geschwindigkeits-Gradienten. Sie beschreiben am Beginn eine allmähliche Beschleunigung der Bewegung und zum Ziel hin eine entsprechende Verlangsamung.

Das gilt es, auf die Ein-/Ausblende zu übertragen. Dazu wird zunächst die Funktion $\text{sigmo}(x)$ konstruiert, die im geschlossenen Intervall $x \in [0, 1]$ streng monoton von 1 auf 0 fällt (S-förmig) und außerhalb dieses Intervalls konstante Werte annimmt:

$$\text{sigmo}(x) = \begin{cases} 1 & : x \leq 0 \\ 1 - \frac{(2x)^m}{2} & : 0 < x \leq 0.5 \\ -\frac{|(2x - 2)^m|}{2} & : 0.5 < x < 1 \\ 0 & : 1 \leq x \end{cases}$$

Der Dämpfungskoeffizient $m \in [1, \infty)$ steuert hier die Stärke der S-Krümmung. Bei $m = 1$ entsteht die lineare Charakteristik. Je größer m gewählt wird, desto stärker ist die S-Charakteristik ausgeprägt. Werte zwischen 0 und 1 invertieren die S-Charakteristik und sind in der Praxis nicht sinnvoll, da hier wieder Undifferenzierbarkeiten entstehen. Die Funktion wird schließlich auf den Definitionsbereich (d_{ref}, d_{max}) und den Wertebereich (g_{min}, g_{max}) skaliert (siehe Abbildung 4.2):

Sigmoidal Distance Model

$$\text{gain}_{\text{sigmo}}(d) = \text{sigmo}\left(\frac{d - d_{ref}}{d_{max} - d_{ref}}\right) \cdot (g_{max} - g_{min}) + g_{min}$$

Im Allgemeinen wird die Lautstärke g_{max} größer sein als g_{min} . Das bedeutet, dass die Soundquelle lauter wird, je mehr sich ihr der Hörer nähert. Aber auch

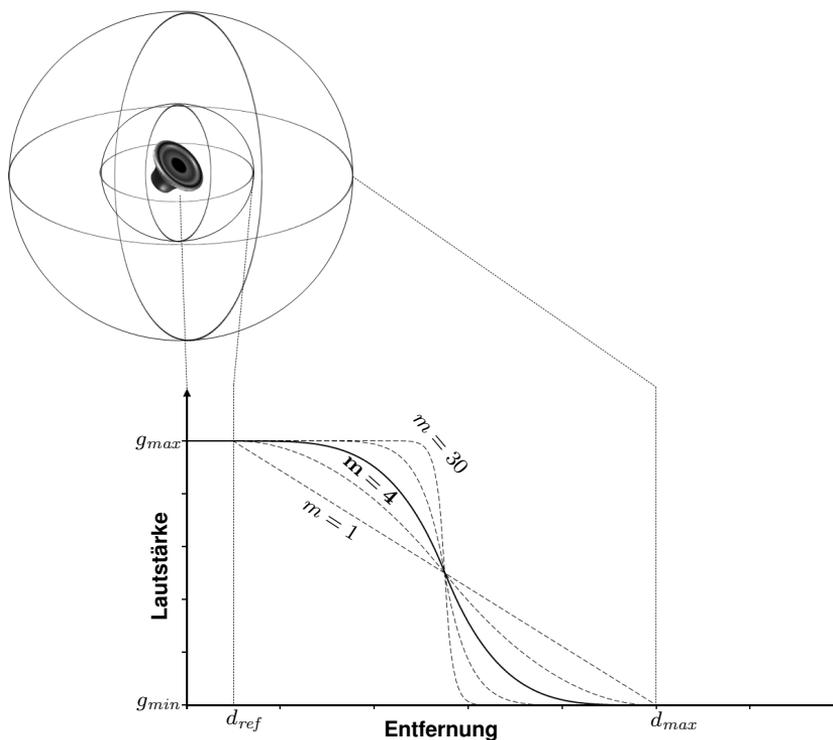


Abbildung 4.2: Sigmoidale Abklingcharakteristik für musikalisches Fading. Die fünf Kurvenverläufe ergeben sich aus den Dämpfungskoeffizienten $m \in \{1, 2, 4, 9, 30\}$.

der inverse Fall ($g_{max} < g_{min}$) kann von Bedeutung sein. Diese Einstellung bewirkt, dass die Soundquelle überall hörbar ist, außer in ihrem direkten Umfeld (in einem Radius von d_{ref}). Auf diese Weise kann quasi ein klangliches Loch erzeugt werden, eine Region, in der diese Soundquelle schweigt. Sie blendet aus, je näher der Hörer kommt.

Nun wird es selten der Fall sein, dass der Spieler sich zielstrebig auf direktem Wege auf eine Soundquelle zubewegt und erst stoppt, wenn sie voll eingeblendet ist. Schließlich kann und wird er sich frei in den von der virtuellen Umgebung gesteckten Grenzen bewegen. Das bedeutet auch, dass Ein- und Ausblenden nicht voll und ohne Unterbrechung ausgeführt werden. Stoppt der Spieler mitten in einer Blende, hält auch diese an. Sie tut dies in der gleichen Weise, wie die Bewegung endet. Abruptes Anhalten führt zu abrupt konstant bleibender

Lautstärke. Geht er abrupt weiter, wird die Blende ebenso unvermittelt fortgesetzt. Hier entstehen neue Knicke in der Lautstärkeregelung.

Sie können verhindert werden, indem die Bewegung der Spielfigur mit einer gewissen Trägheit versehen wird. Das hat allerdings Auswirkungen auf visueller Ebene, nämlich träge Reaktionen auf Benutzerinteraktion, und ist möglicherweise nicht erwünscht. Besser ist es, das Hörer-Objekt¹ losgelöst vom Kamera-Objekt zu bewegen. Hier muss allerdings zwischen zwei verschiedenen Hörer-Objekten unterschieden werden, dem physikalisch in der Szene hörenden, das freilich nicht entkoppelt werden darf, und dem für die nicht-diegetische Musik zuständigen. Dieses Musik-Hörer-Objekt darf entkoppelt werden.

Innerhalb gängiger Audio-Engines ist eine derartige parallele Behandlung von mehr als einem Hörer eigentlich nicht möglich. Nun kann aber eindeutig zwischen Soundquellen unterschieden werden, die nur vom physikalischen Hörer gehört werden, und den Musiksoundquellen, die nur vom Musik-Hörer gehört werden. Keine Soundquelle wird von beiden wahrgenommen. Also ist es möglich, zwei Instanzen der Audio-Szene zu laden, eine für die diegetischen Soundanteile, die zweite für die nicht-diegetischen, und die Ausgabe-Streams beider zu mischen. Das ist eine vergleichsweise unpraktische Lösung.

Alternativ kann auch alles in einer einzigen Audio-Szene umgesetzt werden. Die Lautstärkeregelung der Musik-Soundquellen darf dann nicht von der Audio-Engine automatisch durchgeführt werden, sondern wird von einem separaten Modul übernommen. Dieses simuliert das Musik-Hörer-Objekt und berechnet seine Positionierung inkl. Trägheit bei Bewegungen. Daraus folgt die Lautstärke-Abmischung der Musik-Soundquellen gemäß dem Abklingmodell. Die Platzierung der Musik-Soundquelle im Surround-Panorama, so gewünscht, kann und sollte hingegen ausgehend vom normalen physikalischen Hörer geschehen, denn hier befindet sich der Spieler schließlich. Das Musik-Hörer-Objekt dient nur der musikalischen Abmischung.

Ein weiterer Unterschied zwischen Musik und Physik soll noch genannt werden. In der Akustik werden abhängig vom transportierenden Medium (Luft, trockene Luft, feuchte Luft, Wasser etc.) bestimmte Frequenzbereiche stärker gedämpft als andere. Nicht-diegetische Musik existiert jedoch außerhalb der Szene und wird demzufolge auch nicht von einem darin befindlichen »physikalischen« Medium gefiltert. Ein- und Ausblenden verändern also lediglich die Amplitude des Audiosignals, nicht sein Frequenzspektrum. Dieses verändert sich allerdings in einem anderen Zusammenhang jenseits bloßer Ein- und

¹Das Hörer-Objekt in einer Audio-Szene korrespondiert zum Kamera-Objekt in einer graphischen Szene. Es definiert die Position, Orientierung, einen Beschleunigungsvektor für Hörerbewegungen und die »Master-Hör-Lautstärke«.

Ausblenden und nach anderen Gesetzmäßigkeiten, auf die in Kapitel 5 näher eingegangen wird.

4.4 Bewegliche Soundquellen

Die Region, in welcher eine Soundquelle wahrgenommen werden kann, ist kugelförmig. Die topologische Region der virtuellen Umgebung, die sie abdecken soll, kann im Gegensatz dazu deutlich anders geformt sein. Soll eine Soundquelle etwa einen langen schmalen Gang vollständig abdecken, muss der Radius, in dem sie hörbar ist, so groß gesetzt werden, dass er zwangsläufig auch weit in die umliegende Region hineinragt. Sie vertont dadurch Gebiete, für die sie eigentlich nicht vorgesehen ist. Alternativ wäre denkbar, den Gang mit einer Kette kleinerer Soundquellen zu füllen, die alle die gleiche Musik spielen. Ihre überlappenden Hörbarkeitsregionen führen beim Durchschreiten des Ganges aber zu einem ständigen Lauter- und Leiserwerden. Außerdem entstehen dort, wo sich die Regionen überlagern, Verstärkungs- und Auslöschungseffekte sowie klangliche Artefakte, wenn die Soundquellen leicht phasenverschoben spielen.

Um ungünstig geformte Regionen mit nur einer Soundquelle besser abzudecken, wird für sie ein Bewegungsfreiraum eingeführt. Er wird durch eine Menge von Ebenen beschrieben, die (approximativ) auf die Begrenzungsflächen des abzudeckenden Gebietes gelegt werden (z.B. auf die Wände des langen Ganges). Jede Ebene teilt den dreidimensionalen Raum in zwei Halbräume, einen positiven und einen negativen. Die Schnittmenge aller positiven Halbräume einschl. den Ebenen selbst definiert den Bewegungsfreiraum, innerhalb dessen die Soundquelle nicht mehr an eine feste Position gebunden ist. Hier wird sie so platziert, dass ihre Entfernung zum Hörer minimal ist. Befindet er sich ebenfalls innerhalb, wird sie direkt auf die Hörerposition gesetzt. Ist er außerhalb, folgt sie ihm entlang dem Rand ihres Freiraumes. Das verhindert, dass die Soundquelle zum Hörer springt, wenn der ihren Freiraum betritt, was ein plötzliches Lauterwerden anstelle einer Einblende erzeugen würde. Über Halbraumtests wird bestimmt, ob der Hörer sich innerhalb oder außerhalb des Freiraumes befindet, und an welchen Begrenzungsebenen (und deren Schnittkanten) die Soundquelle gegebenenfalls ausgerichtet werden muss. Dieses Prinzip wurde von Bartz (2010) auch für Klangflächen übernommen.

Aber nicht nur bei der Vertonung der Szenentopologie sind bewegliche Soundquellen nützlich. Sie können auch zur Inszenierung von Nicht-Spieler-Charakteren dienen. Kapitel 2 zeigte bereits, wie wichtig die Rolle der Musik ist, um Charakteren etwa emotionale Tiefe zu geben. Sie kann ihre Einstellung gegenüber dem Spieler (Freund oder Feind?) zum Ausdruck bringen und dabei der

oft hölzernen Mimik und Gestik der virtuellen Puppen unter die Arme greifen. Sie kann auch schlicht dazu dienen, die Präsenz eines Charakters anzuzeigen. Virtuelle Charaktere sind aber im Allgemeinen nicht an einen bestimmten Ort gebunden. Sie können sich mehr oder weniger frei in der Spielwelt bewegen. Die Position der zu ihnen gehörenden Soundquellen muss dementsprechend angepasst werden. Das ist nicht schwer zu erreichen. Die Musik-Engine bietet ein entsprechendes Interface an, über das die Spiele-Engine die aktualisierte Charakter-Position mitteilt, und damit auch ihre Soundquelle neu platziert.

4.5 MIDI-Portierung

Der vorgestellte Arrangement-Mechanismus arbeitet auf Grundlage von Audiodaten. Jedes musikalische Fragment, das von einer Soundquelle abgespielt wird, entspricht einem Audio-Buffer-Objekt, das zu gegebener Zeit an die Wiedergabe-Queue der Soundquelle angehängt wird. Abgesehen von ihrer Abmischung und Positionierung im Panorama findet keine weitere Bearbeitung des musikalischen Materials statt. Die Fragmente bleiben unverändert. Time Stretching, Pitch Shifting, akustische und klangliche Effekte könnten zwar zusätzliche Variabilität einbringen, ändern aber nichts an der musikalischen Substanz.

Richtige Eingriffe in die musikalische Substanz, wie sie etwa von Livingstone (2008), Brown & Kerr (2009) beschrieben werden, erfordern ein symbolisches und im musikalischen Sinne flexibler editierbares Repräsentationsformat. Hierfür entstand eine Portierung aller beschriebenen Funktionalitäten auf Basis des MIDI-Standards (Heinicke 2008). MIDI-Kanäle werden auf Soundquellen abgebildet. Ihre Abmischung wird über ChannelVolume-Controller (Controller Nr. 7) umgesetzt. Lautstärke-Informationen, die in den NoteOn-Events² oder Expression Controllern codiert sind, bleiben dabei erhalten. Die Ein-/Ausblenden beeinflussen also nicht das in sich differenzierte Spiel. Sprungpunkte für das sequentielle Arrangement werden durch Marker in der MIDI-Sequenz angezeigt (vgl. *iMuse* in Anhang B).

Die Stereopositionierung der Soundquelle geschieht über den Panorama-Controller (Controller Nr. 10). Zu beachten ist, dass hier lediglich Pegeldifferenzstereofonie³ stattfindet, die keine interauralen Phasenverschiebungen beinhaltet. Es handelt sich aber, wie schon ausgeführt, nicht um ein physisch in der Szene wahrgenommenes Geräusch. Im musikalischen Zusammenhang spielt solch ein

²entspr. Tastenanschlags-Signal

³Pegeldifferenzstereofonie ist auch bekannt als Intensitätsstereophonie. Hier werden die Pegel von linkem und rechtem Lautsprecherkanal je nach gewünschter Position der Phantomschallquelle eingestellt.

Effekt vorrangig bei der Abmischung von Live-Konzerten eine Rolle, weniger bei Studio-Produktionen. Ein darüber hinaus gehendes (Surround-)Panorama ist im MIDI-Standard nicht vorgesehen und nur über zusätzliche Hard- oder Software-Geräte möglich. Sie belegen hierfür entweder einen oder mehrere der freien MIDI-Controller oder werden über ein gänzlich anderes Interface gesteuert. Es ist auch denkbar, wenn auch nicht mehr sehr praktikabel, einen jeden MIDI-Kanal bei der Wiedergabe nicht direkt auszugeben, sondern erst noch als Audio-Stream in *OpenAL* einzulesen und dort über eine Soundquelle abzuspielen. Abmischung und Panorama geschehen dann in beschriebener Weise.

4.6 Experimente und Diskussion

Die in diesem Kapitel beschriebene Kombination von sequentiellem und parallelem Arrangement wurde in einigen Kompositionsstudien und Demonstrationsszenarien näher untersucht. Die Kompositionsstudien dienten dazu, kompositorische Freiräume und Grenzen zu erforschen. Wie weit geht bspw. die stilistische Flexibilität? Welche Restriktionen erlegt die jederzeitige Umarrangierbarkeit der Musik dem Komponisten auf? Hier entstanden auch autonome (für sich stehende) Musikstücke, von denen einige in praktischen Notenausgaben (Berndt 2008, 2012) zur Veröffentlichung kamen. In den Demonstrationsszenarien wurde der Arrangement-Mechanismus an sich untersucht. Hier sind die Musiken nicht autonom, sondern haben narrative Aufgaben in einer interaktiven Umgebung. Es ging darum, das Verhalten der interaktionsgetriebenen Vorgänge auf evtl. auftretende Seiteneffekte hin zu untersuchen und die Grenzen der musikalischen Reaktionsfähigkeit auszuloten.

4.6.1 Kompositionsstudien

Die vorgestellte Kombination von sequentiellem und parallelem Arrangement arbeitet, ebenso wie die beiden Arrangement-Konzepte jeweils für sich, mit vorkomponierter Musik. Das wird von Komponisten und Musikern einerseits als Vorteil wahrgenommen, denn die musikalische Qualität liegt einzig und allein in ihren Händen. Es bringt für sie aber auch einen Nachteil mit. Der Kompositions- und Produktionsaufwand ist deutlich höher als er es für traditionell in sich abgeschlossene Musikstücke ist. Der Komponist muss die zunächst ungewohnten Wechselwirkungen erst kennenlernen und lernen, mit ihnen zu arbeiten. Die Vielfältigkeit der musikalischen Möglichkeiten hängt allein von dieser Befähigung ab, was Abschnitt 4.2 exemplarisch dargestellt hat.

Wenn die Musik sequentiell und parallel arrangierbar sein soll, ist man zu anderen Herangehensweisen schon bei ihrer Konzeption gezwungen. Das rührt nicht

nur daher, dass die einzelnen Elemente nahtlos aneinander anschließen oder miteinander harmonieren müssen. Dramaturgische Bögen entfalten sich nicht mehr innerhalb der in Endlosschleifen spielenden Sequenzen einzelner Soundquellen, sondern erst durch ihr Arrangement. Bleibt der Spieler stehen, steht auch die Musikdramaturgie. Es blendet nichts neues hinzu, es findet kein musikalischer Wandel statt. Die Musik kreist um den aktuellen Ausdruckscharakter. Erst Interaktion löst die innermusikalischen Entwicklungsprozesse aus. Die Handlungen und Bewegungen des Spielers veranlassen Ein- und Ausblenden und eine Neuorganisation der sequentiellen Abfolge musikalischer Abschnitte.

Um ein lückenloses sequentielles Arrangement in allen Tracks sicherzustellen, ist es notwendig, dass die Segmentgrenzen aller parallel laufenden Tracks zeitgleich übereinander liegen. Bei zeitlich versetzten Segmenten kann es sonst passieren, gerade in Überleitungssituationen, dass ein Track mit seinem nächsten Segment warten muss, bis ein anderer soweit ist, und ein klangliches Loch entsteht. Diese Gleichzeitigkeit der Segmentgrenzen führt jedoch zu formaler Ähnlichkeit (gleicher Gliederungsstruktur) in allen Tracks, schränkt also die Entfaltungsmöglichkeiten des parallelen Arrangements in dieser Hinsicht ein.

Davon abgesehen bleiben der Parallelität der verschiedenen Tracks sehr viele Freiheiten. Verschiedene Instrumentationen eines Musikstücks können problemlos geblendet werden. Dem Spiel mit klangfarblichen Kontrasten, das besonders ab der romantischen Musik zu einem bedeutenden kompositorischen Mittel wurde (Berlioz & Strauss 1844/1904, Sevsay 2005), steht also nichts im Weg.⁴ Dabei muss nicht immer die exakt gleiche Musik aus allen Soundquellen spielen. Eine Soundquelle mag nur die nötigsten Stimmen eines polyphonen Satzes spielen, während andere Soundquellen den Satz mit weiteren optionalen Stimmen anreichern.⁵ Aus einem eher dünnen transparenten Klangbild kann durch Einblenden weiterer Stimmen nach und nach ein dicker komplexer Satz entstehen. Zu beachten ist, dass alle diese Kombinationsmöglichkeiten in sich ausgeglichen sein sollten was Stimmenverteilung, Vollständigkeit der Akkorde, rhythmische und melodische Struktur angeht. Gerade in einer stark ausgedünnten Variante muss mit großen Stimmabständen umgegangen werden. Da kann z.B. ein Kontrabass von den hohen Streichern klanglich zu weit abfallen und sollte vielleicht durch Cello oktaviert oder ersetzt werden.

Dabei müssen die Tracks untereinander nicht einmal stilistisch homogen sein. Schon Dickinson (1989) beschreibt die »Compound Style-Modulation«⁶ (im folgenden *Stilvermischung* genannt) als eine »Herangehensweise an den stilisti-

⁴Viele der Baukastensätze bei Manz & Winter (1976) arbeiten damit. Kontrapunktisch identische Sätze können hier durch ausschließlich hohe oder tiefe Instrumente besetzt werden, oder durch eine gemischte Besetzung.

⁵Das wird in (Berndt 2008, 2012) demonstriert.

⁶Dickinson (1989), Seite 209.

schen Pluralismus«⁷ (siehe auch Kapitel C). Hierbei spielen mehrere Stile nicht nacheinander sondern gleichzeitig. Ein typisches Beispiel: Eine Instrumentalgruppe ist in Stil *A* komponiert. Sie spielt zusammen mit einer anderen Instrumentengruppe, die in Stil *B* komponiert ist. Jede Gruppe ist in sich stilistisch konsistent. Eine enge Verwandtschaft der Stile begünstigt Fusionseffekte. Die Musik wirkt dann in sich homogener. Bei sehr verschiedenen Stilen besteht hingegen die Gefahr, dass der Höreindruck eines zusammenhanglosen Nebeneinanders statt eines gemeinsamen Miteinanders der Stilschichten entsteht. Das kann beabsichtigt sein. Ist es das nicht, müssen Faktoren geschaffen werden, die eine Verschmelzung begünstigen. Neben der rhythmischen ist die tonale Verwandtschaft hierfür sehr wichtig. Ein barocker und ein romantischer Satz können auf den gleichen Grundakkorden basieren, aber der im Barock untypische Gebrauch von Alterationen würde bspw. nur im romantischen Satz auftauchen. Allgemeiner formuliert: Tonal freiere Stile orientieren sich am Tonmaterial der tonal restriktiveren, reichern es mit den eigenen Mitteln an, legen es mit den eigenen Möglichkeiten neu aus. Wenn sich die Gruppen zudem klangfarblich deutlich voneinander absetzen, werden selbst Dissonanzen weniger scharf empfunden als das bei gleichartiger Instrumentierung der Fall wäre (Sevsay 2005). Puristen mögen die Stilvermischung eher skeptisch betrachten. Aber in der heutigen Filmmusikpraxis ist ein, so Kungel (2004), geradezu skrupelloser Umgang mit einer ganzen Palette an Stilen und Stilmitteln gang und gäbe.⁸ Im Dienste der jeweiligen Ausdrucksabsichten (siehe Kapitel 2) ist alles erlaubt. Das parallele Arrangement ist hierfür sogar noch begünstigend.

Selbst für Modulationen ist das parallele Arrangement geeignet. Man stelle sich eine Szene mit zwei Räumen vor, jeder mit seiner eigenen Soundquelle (*A* und *C*). Eine dritte Soundquelle *B* ist in beiden Räumen zu hören. Sie spielt einen harmonisch »unverbindlichen« Satz. Hierbei kann es sich z.B. um hohle Dreiklänge handeln (Grundton und Quint des Akkordes, keine Terz) und um Terzklänge, denen entweder oben die Quinte oder unten der Grundton fehlt, damit ein vollständiger Dreiklang daraus entsteht. Die fehlenden Töne werden von den Soundquellen *A* und *C* gespielt. *A* ergänzt die Dur-Terz einer jeden hohlen Quint, *C* die Moll-Terz. *A* setzt unter die Terzklänge einen Grundton, *C* setzt über sie eine Quint. Damit weist jeder der zwei Räume seine eigene harmonische Ausdeutung auf.

Auch der Eindruck von Temposchwankungen kann über das parallele Arrangement erreicht werden, obwohl die einzelnen Audio-Segmente unverändert bleiben und auch kein Time Stretching ausgeführt wird. Die Soundquellen können aber musikalisches Material spielen, das in seiner metrischen Prägnanz sehr un-

⁷aus dem englischen Titel des Beitrags übersetzt: »approach to stylistic pluralism« Dickinson (1989), Seite 208.

⁸siehe Kungel (2004), Seite 114.

terschiedlich ausgeprägt ist. Lang ausgespielte große Notenwerte wirken langsamer als spitz artikulierte filigrane Bewegungen, trotz unverändertem Grundmetrum. Auch durch Akzentuierung kann der Eindruck unterschiedlicher Tempi entstehen. Die Musik erscheint schneller, wenn auf jedem Schlag im Takt eine Betonung liegt statt etwa nur auf jedem zweiten. Musik der Stilgattung Speed Metal wirkt gerade deshalb so treibend, weil durch das Doublebass-Spiel des Schlagzeugers jede Sechzehntel mit einem Bassdrum-Schlag akzentuiert wird.

Von Künstlerseite zeigten sich weitgehend Offenheit und Bereitwilligkeit, sich das hier vorgestellte Musikkonzept anzueignen, so der Tenor etlicher Gespräche mit Komponisten und Produzenten auf diversen Konferenzen, darunter *Sound, Music and the Moving Image* (eine Konferenz zur Filmmusik), *Audio Mostly* und die *International Computer Music Conference*. Es ist offenkundig, dass die bisher übliche Praxis in interaktiven Medien, nämlich der harte und weiche Schnitt, der musikalischen Qualität schadet. Diese Einbuße kann mit dem vorgestellten Arrangementansatz überwunden und ein durchweg konsistenter musikalischer Ablauf erreicht werden.

4.6.2 Demonstrationsszenarien

Die Demonstrationsszenarien dienen dazu, den Arrangement-Mechanismus an sich zu untersuchen. Zwei interaktive Szenarien wurden entwickelt, eine lineare Bilderfolge in der Tradition klassischer Jump-and-Run-Spiele und eine Geister Schlossgeschichte im Sinne moderner 3d-Adventures. Die graphische Präsentation beschränkte sich auf ein Minimum. Sie half lediglich zur Orientierung, so dass eine sehr ausgiebige, detail- und umfangreiche Arbeit mit der Musik als dramaturgisches und narratives Medium forciert wurde. Die Tracks 1 und 2 auf der beiliegenden Audio-CD stellen zwei klingende Beispiele vor, die bei Live-Durchläufen durch diese Szenarien aufgenommen wurden.



CD-Tracks
1 und 2

Die enge Verknüpfung des Arrangements mit der Topologie der virtuellen Szene kommt der organischen Beiordnung von Musik und interaktivem Geschehen zugute. Doch wie reaktiv ist die Musik nun? Ein- und Ausblenden geschehen zwar sofort und direkt mit den Bewegungen des Spielers in der virtuellen Umgebung. Solche Blenden sollten aber nicht beliebig schnell vollzogen werden, denn das würde einem plötzlichen und unvermittelten Erklingen oder Verstummen nahe kommen – zudem auch an beliebigen Stellen im musikalischen Ablauf und nicht notwendigerweise an musikalisch geeigneten Stellen. Also muss die Abklingkurve bzw. ihr Dämpfungskoeffizient (siehe Abbildung 4.2) so gewählt werden, dass keine allzu steile Steigung entsteht und eine sanfte Blende sichergestellt ist. Die Zeit für eine volle Ein- oder Ausblende sollte sich dabei an der Struktur des zu blendenden Materials orientieren. Feingliedrige Struk-

turen vertragen kurze Blenden eher als große weit ausholende Phrasenbögen. Das leitet sich aus der interpretatorischen Praxis ab, denn auch die dynamische (also lautstärkemäßige) Gestaltung einer Musikinterpretation orientiert sich an musikalischer Struktur (Adorno 1978, 1963, Friberg 1995). Mazzola et al. (2002) prägen dafür die Bezeichnung »Rational Semantics«⁹.

Auch das parallele Arrangement, obwohl sehr reaktiv, kann also nicht beliebig schnell zwischen Musiken überleiten. Dabei sollte aber bedacht werden, dass die Einblende gerade in Kombination mit dem sequentiellen Arrangement bereits läuft, bevor der Spieler die Musikwechselmarke auslöst, die dann zum letztendlichen Musikwechsel führt (siehe Abschnitt 4.2). D.h., die Blende bereitet den Musikwechsel schon lange vor dem auslösenden Ereignis vor. So etwas war bisher bei der Vertonung interaktiver Medien undenkbar. Im Film existiert mit dem *L-Cut*¹⁰ eine ähnliche Konstellation. Bei diesem speziellen Schnitttyp gehen Bild- und Tonspuren zeitlich versetzt zur nächsten Szene über. Meist beginnt die Tonspur vor dem Bildschnitt, damit dieser weniger hart wirkt.

Noch können Musik und Bild aber nicht derart perfekt aufeinander abgestimmt werden, wie das im Film der Fall ist. Kungel (2004) beschreibt, dass nicht nur der Komponist mit seiner Musik den Bildern folgt (oder eben vorseilt). Dafür liegt ihm der Rohschnitt, also eine vorläufige Schnitfassung des Filmes, vor. Die finale Schnitfassung wird dann aber noch einmal an die Tonspuren (einschl. Musik) angepasst. Soll der Bildschnitt betont werden, wird er so verschoben, dass er mit einer musikalischen Akzentuierung zusammenfällt. Soll er hingegen eher maskiert werden, um der Bildfolge mehr Kontinuität zu geben, wird er auf einer unbetonten Zählzeit platziert. Nachdem sich zuerst also die Musik nach den Bildern richtete, werden nun die Bilder noch einmal an die Musik angepasst. Ein solches Hand in Hand der beiden Ebenen findet in interaktiven Medien heute nicht statt. Bild und Musik reagieren unabhängig voneinander auf Interaktionen. Für eine bessere Koordination nach filmischem Vorbild muss von musikalischer Seite signalisiert werden, wann eine Überleitung abgeschlossen und ein sequentieller Musikwechsel vollzogen sein wird. Das ist im hier vorgestellten Arrangement-Konzept problemlos möglich, denn die Länge der musikalischen Segmente, auch des gerade spielenden, ist bekannt. Die Spiele-Engine kann nun selbst entscheiden, ob sie auf das Ende der Überleitung wartet, schon vorher fortfährt, Bildschnitte und andere Ereignisse (Skript-Sequenzen) direkt mit dem Beginn des neuen Segments ausführt oder zeitlich versetzt.

Ein weiterer Aspekt, der bislang nur am Rande Berücksichtigung fand, ist die Variabilität der Musik. OneShots dienen als Überleitungen in eine Musikschleife

⁹Mazzola et al. (2002), Seite 741.

¹⁰Der L-Cut, auch bekannt als *Split Edit* und *akustische Klammer*, wurde erstmals von Fritz Lang (1931) verwendet.

hinein und erklingen auch innerhalb des ersten Schleifendurchlaufs. Alle weiteren Iterationen verzichten auf sie, variieren sonst aber nicht weiter. Ab hier kann die ständige Wiederholung des immer gleichen auffallen. Der Komponist sollte also die Länge der Musik für eine Szene so konzipieren, dass sie die zu erwartende Zeit, die der Spieler durchschnittlich in der Szene verbringen wird, mit den ersten zwei Iterationen großzügig abdeckt.¹¹ Nur extrem langsame Spieler würden mehr als zwei Schleifendurchläufe erleben und feststellen können, dass sich ab der dritten Iteration nichts mehr ändert. Die Varianz, die das Entfallen der OneShot-Segmente bringt, verzögert den negativen Effekt also nur.

Es ist mehr möglich. OneShots können zufallsbehaftet wieder auftauchen. Auch die regulären Segmente können zufallsbehaftet durch Alternativen ersetzt werden (Tobler 2004, Aav 2005). Letzteres ist mit höherem Kompositions- und Produktionsaufwand verbunden, denn hier müssen zu jedem Segment eines oder mehrere weitere geschaffen werden, die dieses nahtlos ersetzen können. Hier bietet die MIDI-Portierung die nötige Grundlage, um auch automatische Variationsverfahren einzusetzen. Melodievariationen werden bspw. von Hörnel's (2000) *MeloNet* und *JazzNet* sowie von Miranda & Biles' (2007) *GenJam* generiert. Eine solche Varianz kann auch in Überleitungen hilfreich sein. Muss der Spieler eine Musikwechselmarke mehrmals im Spielverlauf durchschreiten, kann es auffällig werden, dass die Musik immer gleich reagiert. Variation kann das verhindern und zu einem insgesamt runderen stimmigeren Höreindruck beitragen.

Der wird, zumindest in musikalischer Hinsicht, noch dadurch gestört, dass Ein- bzw. Ausblenden genau dann geschehen, wenn sich der Spieler bewegt. Sie können mitten in einer Phrase beginnen und mitten in der übernächsten Phrase enden. Nach rein inner-musikalischen Gesichtspunkten sind solche »naiven Blenden« problematisch, denn sie haben mit dem musikalischen Material, das ein- oder ausgeblendet wird, nichts zu tun. Besser wäre es, würden die Blenden sich auch an musikalischen Strukturgrenzen orientieren. Hierzu stellt Kapitel 6 einen Ansatz vor. Umgekehrt bedeutet das aber auch den Verzicht auf die direkte Kopplung von Spielerbewegungen und musikalischer Reaktion. Da Benutzerinteraktionen und die visuelle Ebene integrierend auf musikalische Vorgänge wirken (Lissa 1965), werden die naiven Blenden im Gesamtzusammenhang evtl. gar nicht als störend und unmusikalisch wahrgenommen. Letztlich bleibt es den ästhetischen Vorstellungen und Absichten der Entwickler überlassen, welche Lösung den Vorzug erhält (die hier vorgestellte oder die in Kapitel 6).

¹¹Die zeitliche Disposition der Musik richtet sich generell nach der zu erwartenden Länge der zu vertonenden Szene. Diese Praxis ist heute üblich.

Modellierung ausdrucksvoller Interpretationen

IN diesem Kapitel werden Modelle zur Beschreibung von ausdrucksvollen Musikinterpretationen entwickelt und implementiert. Sie bilden die Grundlage für einen Ansatz zur nichtlinearen Interpretation, das *reaktive Orchester*, in Kapitel 6. Folgende Veröffentlichungen liegen diesem Kapitel zu Grunde:

- (Berndt 2006),
- (Berndt & Hähnel 2009a),
- (Berndt & Hähnel 2009b),
- (Berndt & Hähnel 2010),
- (Hähnel & Berndt 2010a),
- (Hähnel & Berndt 2010b),
- (Berndt 2011b),
- (Hähnel & Berndt 2011a),
- (Hähnel & Berndt 2011b).

Kann Musik durch den Computer ausdrucksvoll dargeboten werden? Kann eine solche synthetische Darbietung ebenso ausdrucksstark sein wie eine Darbietung von menschlichen Musikern? Der Gedanke, dass etwas so emotional geprägtes wie das Musizieren auch maschinell machbar sein könnte, polarisiert. Könnten damit reale Musiker ersetzt werden? Das visuelle Pendant zum virtuellen Musiker ist der virtuelle Schauspieler. Lange Zeit galten menschliche Mimik und Gestik als zu komplex, angereichert mit zu vielen subtilen Details, um sie auch

nur annähernd glaubhaft synthetisieren zu können. Zu hölzern wirkten die Bewegungen und Regungen der frühen computergenerierten Akteure. Heute jedoch sind die Grenzen zwischen Mensch und Maschine kaum mehr zu erkennen. Manch ein gefährlicher Film-Stunt und so manche Einstellung wären ohne das virtuelle Double gar nicht inszenierbar. Der virtuelle Schauspieler hat sich etabliert, nicht als Ersatz sondern als Ergänzung, wo reale Menschen nicht verfügbar bzw. bezahlbar sind (Massenszenen) oder die Grenzen der menschlichen Physis überschritten werden.

Wie sieht es in der Musik aus? Würde der virtuelle Musiker den realen ersetzen oder ergänzen? Man denke nur an die Kosten, die entstehen, wenn ein sinfonisches Orchester engagiert wird; Kosten, die einzelne Künstler und Low-Budget-Projekte sich nicht leisten können. Sollten ihnen die orchestralen Klänge also verwehrt bleiben? Man führe sich die Probleme bei der Vertonung interaktiver Medien vor Augen, wo von Menschen eingespielte und produzierte Musik zu statisch ist, um dem raschen Wandel, ausgelöst durch Benutzerinteraktionen, auch nur einigermaßen organisch zu folgen. Ein virtuelles Orchester könnte noch reagieren, wenn kein Mensch mehr dazu in der Lage ist. Ein virtueller Arrangeur könnte noch umarbeiten, was der menschliche Komponist längst nicht mehr beeinflussen kann.

Der Computer hätte also seine Daseinsberechtigung als Interpret wenigstens überall da, wo Menschen nicht wollen oder können. Doch kann er mit dem menschlichen Musiker mithalten? Scheinbar nicht; zu komplex ist die Fülle an interpretatorischen Details, die Menge an großen und kleinen, auffälligen und subtilen Nuancen. Menschen spulen Musik nicht metrisch exakt ab »wie ein Computer«. Sie gestalten das Tempo bzw. Timing, verlangsamen, beschleunigen, zögern, eilen. Menschen spielen nicht alles in einer Lautstärke, sondern nutzen die dynamischen Möglichkeiten ihres Instruments oder ihrer Stimme, um metrische Schwerpunkte, Akzente, melodische Ziele und Spannungsbögen herauszuarbeiten. Damit verbunden sind auch reiche klangfarbliche Differenzierungen. Stimme wie Instrumente verändern den Charakter ihres Klangs, ihr Obertonspektrum, bei unterschiedlicher Spielweise, Lautstärke und in unterschiedlichen Registerlagen. Einer laut schreienden Stimme hört man an, dass sie schreit, selbst wenn man sie nur ganz leise hört. Sogar in der Formung jedes einzelnen Tones, seiner Artikulation, kann man die Spuren menschlicher Gestaltung finden. Wer einmal versucht hat, diese Fülle an Details von Hand in einen Strom von MIDI-Signalen einzuarbeiten, weiß, wie überaus aufwendig es ist, menschliche Interpretation in allen ihren Facetten nachzubilden.

In diesem Kapitel werden Modelle entwickelt, mit deren Hilfe das komplexe Zusammenspiel der vielen interpretatorischen Details approximativ nachgebildet werden kann. Ausdrucksvolle und lebendige Interpretationen lassen sich damit einfacher und intuitiver beschreiben und zu ausdrucksvollen MIDI-Sequenzen

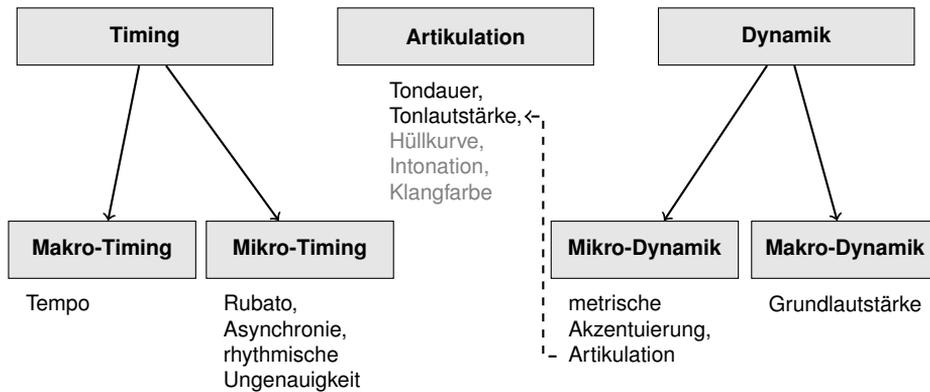


Abbildung 5.1: Ein Überblick über die Merkmalsklassen der ausdrucksvollen Interpretation, zu denen in diesem Kapitel Modelle entwickelt werden. Hüllkurve, Intonation und Klangfarbe werden zwar mit behandelt, aber nicht in Modelle überführt und sind deshalb grau gefärbt.

rendern. Sie erleichtern auch die maschinelle Weiterverarbeitung, z.B. um während der Wiedergabe nahtlos von einer Interpretationsweise in eine andere überzuleiten und damit den Ausdruckscharakter der Musik eng an einem interaktiven Geschehen zu orientieren (siehe Kapitel 6).

Im Vorfeld ist es wichtig, zwischen Interpretations-*Planung* und Interpretations-*Synthese* zu unterscheiden, also zwischen der Entscheidung, wie ein Musikstück mit interpretatorischen Mitteln ausgestaltet wird, und der eigentlichen Umsetzung, dem Zum-Klingen-Bringen des Planes. Zu ersterem gibt es bereits einige Ansätze; siehe dazu Kapitel 3 und den Überblicksartikel von Widmer & Goebel (2004). Dieses Kapitel geht auf letzteres, geht also detaillierter Fragen der Interpretations-Synthese nach. Welche interpretatorischen Merkmale gibt es? In welcher Weise beeinflussen sie das klingende Resultat? Welche Varianzen gibt es in ihrer Ausprägung? Wie lässt sich das durch eine geeignete Parametrisierung in die Modelle einbauen? Die hier entwickelten Modelle beschreiben die Gestalt der interpretatorischen Merkmale (Feature Design), um sie schließlich in MIDI-Sequenzen einzuarbeiten und entsprechende ausdrucksvolle Darbietungen zu erzeugen. Einen Überblick über die hier beschriebenen Merkmalsklassen gibt Abbildung 5.1.

5.1 Allgemeine Bemerkungen zur Interpretationsanalyse

Wie untersucht man Interpretation? Anhören allein ist in seiner Subjektivität und Ungenauigkeit lediglich dafür geeignet, einen eigenen persönlichen Höreindruck zu gewinnen. Der wird stark von der jeweiligen Erwartung, Erfahrung und kulturellen Prägung des Hörers beeinflusst, wie auch von akustischen Gegebenheiten des Aufführungsortes und -mediums. Auch visuelle Reize, wie Bewegungen der Musiker und Dirigat oder gar eine ganze Bühnenshow, können den Höreindruck verfälschen, insbesondere im Hinblick auf die Wahrnehmung von Tempo und Rhythmus; ein Effekt, mit dem in der Filmmusik sogar bewusst gearbeitet wird (Lissa 1965, Schneider 1989, 1990, Bullerjahn 2001, Wingstedt 2008).

Objektiv lässt sich eine konkrete Interpretation erst durch Messung erschließen, *Analysis by Measurement* (Gabrielsson 1999). Die Analyse von Audiodaten zieht dabei meist erheblichen Aufwand von Seiten der Signalverarbeitung nach sich. So gilt es etwa, einzelne Instrumente aus dem Gesamtbild herauszufiltern, Tonanfänge zu ermitteln (*Onset Detection*) oder die Hüllkurven von Tönen nachzuzeichnen. Einfacher gestaltet es sich, wenn der Interpret mit einem MIDI-fähigen Musikinstrument ausgestattet werden kann (Windsor & Clarke 1997, Clarke & Windsor 2000), oder das Instrument mit zusätzlichen Sensoren versehen werden kann (Young 2002). Dann können die Steuersignale bzw. Sensordaten des Instruments direkt ausgelesen und analysiert werden.

Die Messergebnisse mehrerer Aufzeichnungen/Aufführungen können mit dem Instrumentarium des *maschinellen Lernens* (Widmer & Goebel 2004, Saunders, Haroon, Shawe-Taylor & Widmer 2008) zueinander in Bezug gebracht werden. So können Gemeinsamkeiten und Regularien ermittelt, Interpretieren und Interpretationsstile miteinander verglichen und deren Ähnlichkeit oder Verschiedenheit bemessen werden. Das Vorliegen exakter Messergebnisse ermöglicht überdies den Vergleich mit Modellannahmen, die aus musiktheoretischer, -historischer und -praktischer Fachliteratur, wie sie von Bernhard (1657), Quantz (1752), Bach (1753), Mozart (1789), Holzweissig & Holzweissig (1996) vorliegt, gewonnen werden können. Hier relativiert sich so manche Regel.

Nichtsdestotrotz legt die Fachliteratur einen wertvollen Wissensschatz nieder, der einen guten Ausgangspunkt für die Interpretationsanalyse durch Synthese (*Analysis by Synthesis*) ermöglicht (Gabrielsson 1985, Todd 1989, de Poli 2004, Friberg 1995, Friberg et al. 2006). Die einst verfassten Beschreibungen können in formale Modelle überführt werden. Diese lassen sich auf dem Computer implementieren und können verwendet werden, um synthetische Interpreta-

tionen zu erzeugen. Je nach Geschmack und Absichten des Anwenders/Hörers können sie sukzessive weiter adaptiert und verfeinert werden, so dass sich schließlich ein bestimmter Interpretationsstil herausbildet. Dessen Charakteristiken lassen sich mit denen anderer Stile vergleichen. So können bspw. musikgeschichtliche Wandlungsprozesse in der Interpretationspraxis nachvollzogen werden. Der Computer als Interpret kann also sogar musikwissenschaftlich interessant sein, erlaubt er doch hoch aufgelöste Detaileinblicke und die effiziente Bearbeitung großer Datenbestände. Wie gut er menschliches Spiel nachempfindet, ist dabei einzig und allein davon abhängig, wie adäquat die aus den Analysen gewonnenen Modellbeschreibungen die Charakteristiken menschlichen Spieles einfangen.

Die in diesem Kapitel beschriebenen Modelle wurden in einer Kombination aus *Analysis by Synthesis* und *Analysis by Measurement* gewonnen. Den Ausgangspunkt bilden die teils sehr anschaulichen, leider oft auch unterschiedlich interpretierbaren Beschreibungen in musikpraktischer und -pädagogischer Literatur. Sie lassen bereits Rückschlüsse auf die grundsätzliche Gestalt vieler Merkmale zu. Es folgen einige Beispiele.

Holzweissig & Holzweissig (1996): »Legato bedeutet: jeder Ton muss an den nächsten herangebunden werden. Im Notenbild wird dies durch die Bezeichnung *Legato* oder durch Bindebögen angezeigt. Ist ein Bogen zu Ende, wird abgesetzt.« (Seite 4) Die Artikulationsanweisung *Legato* hat also Auswirkung auf die Länge der Töne.

Baur, Eckert & Nonnenmann (2007): »portato (getragen): Die Töne werden leicht abgesetzt. [...] marcato (betont): Die Töne erhalten am Anfang eine Betonung.« (Seite 152) Artikulationsanweisungen beeinflussen nicht nur die Tonlänge, sondern auch ihre Lautstärke. An anderer Stelle wird deutlich, dass solche Veränderungen absolut oder relativ zur Grundlautstärke bzw. -tonlänge sein können.

Bernhard (1657): »Das *Piano* und *Forte*, wie auch die anderen Kunststücke, sollen von uns ins Künftige zur Nachricht, mit sonderbaren Zeichen bemerkt werden, und zwar [...] das *piano* mit dem *p*, und das *forte* mit dem *f*. [...] In gantzen und halben Noten pflegt man das *piano* zum Anfange, das *forte* in der Mitte und zuletzt wiederum das *piano* zu brauchen, [...] wobei denn wohl in acht zu nehmen, daß man nicht plötzlich aus dem *piano* ins *forte* [und aus diesem wieder auf jenes] fallen, sondern allmählich die Stimme wachsen und abnehmen lassen müsse, sonst würde dasjenige, welches ein Kunststück seyn sollen, recht abscheulich lauten.« (Paragrafen 8 bis 10) Hier ist von *Subnotendynamik* die Rede, also davon, dass lautstärkemäßige Vorgänge sich nicht nur von Ton zu Ton vollziehen, sondern bei langen

Tönen auch innerhalb dieser, ja sogar zur Formung einzelner langer Töne dienen. Zudem werden gleichmäßige Lautstärkeübergänge gefordert.

Quantz (1752): »Man muss sich bemühen jede Note nach ihrer gehörigen Geltung zu spielen; und sich sorgfältig hüten, weder zu eilen noch zu zögern. Man muß, zu dem Ende, bey jedem Viertheile auf das Zeitmaaß gedenken; und nicht glauben, es sey schon genug, wenn man nur bey dem Anfange und der Endigung des Tactes mit den übrigen Stimmen zutreffe. Das Übereilen der Passagien kann entstehen, wenn man, besonders bey steigenden Noten, die Finger zu geschwind aufhebt. Um dieses zu vermeiden, muß man die erste Note der geschwinden Figuren, ein wenig markieren, und anhalten; [...] um so vielmehr, da immer die Hauptnoten ein wenig länger, als die durchgehenden, gehört werden müssen.« (Kapitel XII, Paragraph 5) Hier beschreibt Quantz eine besondere Form der Abweichung vom metrisch exakten Spiel, die im Folgenden und in Abgrenzung vom wohlbekannten Begriff des musikalischen *Tempo* als *Rubato* bezeichnet wird. Das *Rubato* ist eine zeitliche Verzerrung, die sich innerhalb eines gewissen Zeitfensters selbst kompensiert und dadurch im Grundmetrum bleibt. Eine gedehnte Note wird durch Kürzung der übrigen Noten im Zeitfenster kompensiert. Damit weist Quantz darauf hin, dass neben dem Tempo noch eine weitere zeitliche Gestaltungsebene in der Musikinterpretation existiert.

Wie stark diese sogenannte Inégalité der Noten ausgeprägt sein kann, beschreibt er noch genauer: »Man muß unter den *Hauptnoten* [...] und unter den *durchgehenden* [...] einen Unterschied im Vortrage zu machen wissen. Die Hauptnoten müssen allezeit, wo es sich thun läßt, mehr erhoben werden, als die durchgehenden. Dieser Regel zu Folge müssen die geschwindesten Noten, in einem jeden Stücke von *mäßigem Tempo*, oder auch im *Adagio*, ungeachtet sie im Gesichte nach einerley Geltung haben, dennoch ein wenig ungleich gespielt werden; so daß man die anschlagenden Noten einer jeden Figur [...] etwas länger anhält, als die durchgehenden [...]: doch muß dieses Anhalten nicht so viel ausmachen, als wenn Punkte dabey stünden.« (Kapitel XI, Paragraph 12)

Auch in wissenschaftlicher Literatur wurden bereits einige, insbesondere zeitliche, Aspekte genauer untersucht. Hefling (1993) und Hudson (2004) beschäftigen sich detailliert mit der Inégalité, Überpunktierung und weiteren Mikro-Timing-Merkmalen. Schlussritardandi¹ werden von Sundberg & Verrillo (1980) und Honing (2003) genauer untersucht. Friberg & Sundberg (1999) leiten aus der Beobachtung motorischer Abläufe (abbremsende Laufbewegungen) die Cha-

¹Das ist die Verlangsamung des Tempos zum Ende des Musikstücks hin.

rakteristik ihrer Ritardandi ab. Weitere Untersuchungen zum Timing in Musikdarbietungen kommen von Todd (1992, 1995), Clarke (1999) und Sundberg, Friberg & Bresin (2003). Einen guten Überblick gibt Gabriëlsson (1999). Mit der Analyse von Dynamik und Artikulation beschäftigen sich u.a. Dixon (2000), Jerkert (2003), Ramirez, Hazan & Maestre (2005), Maestre & Gómez (2005), Maestre Gómez (2006).

Diese initialen Modellannahmen konnten in praktischen Versuchen mit professionellen Musikern und Laien probiert werden.² Hierbei konnten vielfältige Hinweise darauf gewonnen werden, in welcher Weise die Grundcharakteristiken variieren können. So neigen bspw. Laien dazu, Crescendi (lauter werden) erst sehr spät und dann eher plötzlich auszuführen, Decrescendi (leiser werden) entsprechend schon sehr früh. Professionelle Musiker gestalten solche Lautstärkeübergänge dank besser trainierter Motorik am Instrument deutlich ausgewogener und bewusster. Auch der lautstärkemäßige Spielraum ist bei Profimusikern weit größer als bei Laien. Folglich sind Lautstärkedifferenzierungen bei Profimusikern stärker ausgeprägt und für den Hörer deutlicher. Auf dieser Grundlage wurden Modelle entwickelt, die durch teilweise optionale Parameter sehr flexibel eingestellt werden können. Selbst für die Praxis wenig relevante Entartungen werden durch extreme Parametereinstellungen mit abgedeckt. Auf der Basis von Detailmessungen in verschiedenen Aufnahmen konnten schließlich typische Charakteristiken beobachtet und die Modelle überarbeitet werden. Diese Studie umfasste

- zehn professionelle Musik-Produktionen von G. P. Telemanns *Trompetenkoncert in D-Dur* (TWV 51:D7) aus dem Repertoire der Bibliothek des *Zentrums für Telemann-Pflege und -Forschung Magdeburg*,
- Live-Mitschnitte beim *5. internationalen Telemann-Wettbewerb für historische Holzblasinstrumente* 2009 in Magdeburg und
- Aufnahmen mit zehn professionellen Musikern, die einige speziell entworfene Tempo-, Dynamik- und Artikulationsübungen auszuführen hatten. Die Noten zu den Übungen sind in Anhang D einzusehen.

In diesen Aufnahmen wurden die Tonanfänge ermittelt. Vollautomatische Verfahren zur *Onset Detection* können aber gerade bei Orchestermusik keine zuverlässigen Ergebnisse liefern. Im Ensemblespiel verschmieren Tonanfänge oft so stark, dass keine oder zu viele Onsets gefunden werden. Letzteres geschieht, wenn die Schwellwerte für die sog. *Peak Picking*-Phase der *Onset Detection*-Algorithmen zu niedrig eingestellt werden. Stattdessen wurden die Tonanfänge in

²Der Autor übernahm ein dreiviertel Jahr lang die Leitung eines Posaunenchores in Magdeburg und musiziert mit professionellen Musikern und Laien in verschiedenen Ensembles.

einem halbautomatischen Verfahren bestimmt. Zunächst wurde eine Menge von Onset-Hypothesen mit Hilfe zweier etablierter *Onset Detection*-Algorithmen erzeugt (Duxbury, Bello, Davies & Sandler 2003, Bello, Duxbury, Davies & Sandler 2004, Dixon 2006). Beide basieren auf der *Complex Domain Onset Detection*, die mit Energieänderungen im Frequenzspektrum und Phasenverschiebungen im Audiosignal arbeitet. Als nächstes wurden durch manuelles Mittippen beim Hören der Musik Marken in den Hypothesenraum gesetzt. In einer gewissen Umgebung mit einem Radius von 60–90 Millisekunden um eine jede Marke wurde dann die Hypothese mit dem größten Peak in der *Onset Detection*-Funktion als definitiver Onset ausgewählt.

Aus den Inter-Onset-Intervallen (Ton-zu-Ton-Abstände) und der zugrunde liegenden Partitur lassen sich Tempokurven direkt ableiten. Sind die Tonanfänge erst einmal bekannt, kann auch ihre Lautstärke (Dynamik) gemessen werden. Hierfür sind nachbearbeitete Aufnahmen allerdings ungeeignet; die Amplituden entsprechen nach Filterung, Kompression und Abmischung kaum mehr dem Original. Maempel (2010) gibt Überblick und Diskussion über die zahlreichen Veränderungen, die der Produktionsprozesses in eine Musikaufnahme einbringt. Die Trompetenkonzert-Einspielungen wurden für die Analyse von Dynamik und Artikulation folglich nicht verwendet. Weiterhin muss zwischen zwei verschiedenen Interpretationen des Begriffs Lautstärke unterschieden werden, nämlich zwischen der lautstärkemäßigen Formung des Tones (Hüllkurve) und der Lautstärke des Tones insgesamt. Die Hüllkurve beschreibt, wie sich die Amplitude über die Zeit, in welcher der Ton klingt, entwickelt. Das ist für die Messung der Artikulation des Tones von Bedeutung. Es ist aber für die generelle Lautstärke des Tones als ganzes wenig aussagekräftig. Ein Klavierton wird kontinuierlich leiser, solange er klingt. Trotzdem handelt es sich dabei nicht um ein Decrescendo im interpretatorischen Sinne. Es gibt nicht an, wie laut der Ton als musikalisches Ereignis insgesamt gespielt wird. Entscheidend hierfür ist vielmehr die Lautstärke, welche der Ton am Ende seiner Einschwingphase (Attack-Phase) erreicht. Beim Klavier wird diese durch die Stärke des Tastenanschlages bestimmt. Bei Blasinstrumenten wird sie durch den Anblasdruck und bei Seiteninstrumenten durch Anschlagstärke bzw. Bogenführung reguliert. Abschnitt 5.4 geht auf die verschiedenen Erscheinungsformen dynamischer Merkmale näher ein. Hier soll es genügen, dass gezeigt ist, an welchen Stellen Artikulation und Dynamik aus dem Audiosignal abgelesen werden.³

Die in den Messergebnissen auftretenden Merkmalsausprägungen wurden mit denen verglichen, die durch die Modelle erreicht werden können. Die Modelle

³An dieser Stelle einen herzlichen Dank an Tilo Hähnel, der diese Studie im Rahmen seiner Doktorarbeit durchgeführt hat und dessen Auswertung für die Weiterentwicklung der Modelle eine wertvolle Eingabe war.

wurden entsprechend revidiert, angepasst und erweitert zu der in den folgenden Abschnitten dieses Kapitels dargestellten und diskutierten Form.

5.2 Klang

Eine erste Hürde bei der Erzeugung synthetischer Interpretationen stellt schon die Nachbildung der klanglichen Eigenschaften realer Musikinstrumente dar. Einen Überblick über die verschiedenen Arten der Klangsynthese geben u.a. Roads (1996) und Boulanger (2000). Eindrucksvolle Ergebnisse kann hier die auf physikalischen Modellen basierende Klangsynthese erzielen, die verschiedenste Parameter wie Anblasdruck, Bogenführung und Resonanzen des Klangkörpers berücksichtigt (Serra 2007, Demoucron 2008).

Ein weiterer etablierter Ansatz ist das Sampeln realer Instrumente. Dabei werden einzelne Töne in gewissen Intervallabständen (Sekund- und Terzabstand sind üblich) und bei verschiedenartiger und verschieden lauter Spielweise aufgenommen und können nun zu neuen Musikstücken zusammengesetzt werden. Das allein genügt aber nicht. Tonanfang und Tonende können je nach Artikulation und daraus resultierender Verbindung zum Vorgänger- und Folgeton sehr unterschiedlich ausfallen. Eine in dieser Hinsicht sehr akribisch zusammengestellte Datenbank gesampelter Orchesterinstrumente ist die *Vienna Symphonic Library* (Vienna Symphonic Library GmbH 2010b). Neben Standard-Artikulationen, wie Pizzicato, Spiccato, Staccato, Marcato, Legato und Portamento, enthält die Bibliothek auch Samples für spezielle Spieltechniken, bspw. Flatterzunge, mehrere Vibrato-Varianten, Triller und verschiedene gedämpfte Klänge. Allein das artikulierte Spiel der Instrumente lässt sich hier von dem realer Musiker, die diese Samples schließlich eingespielt haben, kaum mehr unterscheiden. Die Samples können in sog. Performance-Matrizen arrangiert und unter anderem über MIDI-Controller direkt angesteuert werden.⁴ Die im weiteren Verlauf dieses Kapitels beschriebenen Techniken zur Interpretations-Synthese wurden in zwei Modi implementiert, einem Standard-MIDI-Modus und einem speziell angepassten Vienna-Modus, der die zusätzlichen Controller-Events für die Wiedergabe über den Sampler *Vienna Instruments* (Vienna Symphonic Library GmbH 2010a) automatisch generiert, um das Potential der Klangbibliothek effektiver zu nutzen.

Für einen authentischen Höreindruck ist außer dem Instrumentenklang an sich auch der Klang des Aufführungsortes wichtig, die Akustik. Dafür geeignete

⁴Umfang und Möglichkeiten der *Vienna Symphonic Library* und des zugehörigen Samplers gehen noch weiter, werden kontinuierlich weiterentwickelt und können auf der VSL-Webseite <http://vsl.co.at/> nachgelesen werden.

VST-PlugIns⁵ gibt es in unterschiedlicher Qualität heute zur Genüge, sowohl kommerziell als auch kostenlos. Zunehmend implementieren auch Soundkarten solche Effekte. Hier kann also auf bestehendes Material zurückgegriffen werden.

Es soll nicht vergessen werden, dass nicht immer klanglicher und akustischer Realismus gewollt sind. Bewusst synthetische Klänge und Effekte haben ebenfalls ihren Platz in der heutigen Musikkultur. Letztlich ist es egal, von wem die MIDI-Signale interpretiert werden, Sampler, Synthesizer oder Effektgenerator. Die dem MIDI-Konzept innewohnende Modularität macht es einfach, die Musik mit ganz unterschiedlichen Instrumentationen und auf ganz unterschiedlichen Geräten abzuspielen, der mit der Soundkarte mitgelieferten Klangbibliothek, dem kostenlosen Synthesizer aus dem Internet oder einem professionellen Hard- oder Software-Sampler oder -Synthesizer. Die klanglichen Möglichkeiten zum ausdrucksvollen Spiel sind also gegeben. Nun gilt es, dieses in seiner Form näher zu untersuchen und zu modellieren.

5.3 Timing

Das Timing einer Interpretation bildet symbolische Zeit (Partiturposition) auf physikalische Zeit (Millisekunden) ab. Musikalische Ereignisse sind in der symbolischen Zeitdomäne organisiert, üblicherweise in einem hierarchischen Raster, das durch Takt, Schlag und Schlagunterteilung aufgespannt wird. Diese Darstellung ist auch unter der Abkürzung *BBU* (Bar, Beat, Unit) bekannt und entspricht dem unter Musikern gebräuchlichen Format zur Angabe von Partiturpositionen (ein Beispiel: »Takt 6 auf Zählzeit 3-und«). Auch der MIDI-Standard kennt eine symbolische Zeit; sie wird durch die MIDI-Uhr repräsentiert. Um *BBU*-Darstellung in MIDI-Darstellung umzurechnen, wird der Umrechnungsfaktor *ticksPerQuarterNote* im Header einer jeden MIDI-Datei definiert, auch bekannt als *PulsesPerQuarter*, kurz: *PPQ*. Er gibt an, wie viele Zeiteinheiten der MIDI-Uhr (MIDI-Ticks, oder kurz Ticks) einer Viertelnote in der Partiturdarstellung entsprechen. Der Wert gilt für das gesamte Musikstück.

Allen symbolischen Zeitdarstellungen ist gemein, dass sie nur ein grobes relatives Timing definieren, nach welchem eine Viertelnote halb so lang ist wie eine Halbe, doppelt so lang wie eine Achtel usw., aber nicht das Tempo, in welchem das Raster vom Musiker abgearbeitet wird! Die Noten werden nicht kürzer notiert, wenn sie schneller gespielt werden sollen. Wie schnell die Musik also gespielt werden soll, ist aus der symbolischen Zeitdarstellung nicht abzulesen.

⁵VST (Virtual Studio Technology) ist eine Kommunikations-Schnittstelle für Audio-Software, entwickelt von Steinberg Media Technologies.

Erst durch das Timing können die musikalischen Ereignisse physikalischen Zeitpunkten zugeordnet werden. Im einfachsten Fall handelt es sich dabei um eine konstante Tempo-Funktion, welche symbolische Zeit linear auf physikalische Zeit abbildet. Die Musik wird hier metrisch exakt in dem jeweiligen Tempo abgespielt. In der Realität sind Tempokurven aber weit komplexer und beinhalten mehr als nur ein konstantes Grundmetrum der Interpretation. Sie sollen deshalb im weiteren mit dem allgemeineren Begriff *Timing-Kurve* bzw. *Timing-Funktion* bezeichnet werden. Phänomenologisch können wenigstens vier konzeptionell verschiedene Klassen von Timing-Merkmalen unterschieden werden, die sich überlagern und zusammen die komplexe Timing-Kurve ergeben.

Tempo: Das Tempo ist die allgemein am besten bekannte Klasse von Timing-Merkmalen. Es definiert das der Interpretation zugrunde liegende musikalische Metrum im Format »Anzahl symbolische Zeiteinheiten pro physikalische Zeiteinheit«. Am gebräuchlichsten ist »Schläge pro Minute« (Beats per Minute, kurz: bpm). Das Tempo kann konstant bleiben, es kann sich aber auch diskret oder kontinuierlich über die symbolische Zeit verändern. Die aktive Arbeit mit Tempoänderungen ist ein wichtiges Mittel, um Höhepunkte, melodische Ziele und musikalische Form (Altmann 2001) herauszuarbeiten. Tempoanweisungen erstrecken sich meist über musikalisch mittel- und langfristige Intervalle (symbolischer Zeit). Sie sind meist länger als ein Takt. Das Tempo sei deshalb auch als *Makro-Timing* bezeichnet.

Rubato: Menschliche Musiker setzen das durch das Tempo definierte Grundmetrum nicht exakt um. Bereits in der barocken Flötenschule von Quantz (1752) werden bewusste und unbewusste Verzerrungen des metrisch exakten Timings besprochen und geschult (siehe Zitate in Abschnitt 5.1). Kleine Streckungen und Stauchungen der Notenlängen dienen dazu, Figuren, Motive, metrische Schwerpunkte und melodisch wichtige Töne herauszuarbeiten. Sie gehen kaum über die Länge eines Taktes hinaus, sind meist deutlich kürzer und zählen deshalb zum *Mikro-Timing*. Heute wird der Begriff Rubato oft auch für zeitlich kurze Temposchwankungen verwendet, die aber ebenso auch durch das Konzept Tempo bereits abgedeckt sind. Deshalb wird der Begriff Rubato hier nach seiner älteren, ursprünglichen Bedeutung verwendet (Hudson 2004, Goebel, Flossmann & Widmer 2010). Das Rubato bezeichnet Timing-Verzerrungen, die sich innerhalb eines gewissen Zeitfensters selbst kompensieren. Anfang und Ende des Zeitfensters sind immer im Grundmetrum. Das Makro-Timing wird vom Rubato also nicht aufgehoben, sondern mit feingliedrigeren Details angereichert. Rubato-Verzerrungen können als singuläre Phänomene auftauchen, z.B. wenn es darum geht, eine bestimmte Melodiestelle herauszuarbeiten. Sie

können aber auch als wiederkehrendes Schema in Erscheinung treten, etwa beim Wiener Walzer taktweise wiederkehrend oder im Swing schlagweise.

Asynchronie: Im Ensemblespiel sind die Musiker nicht immer exakt zusammen. Diese Asynchronie zwischen den Musikern ist teilweise systematisch und leitet sich aus der Hierarchie der Instrumente und ihrer Aufgabe im Musikstück ab (Rasch 1979). Führende Stimmen sind voraus, Begleitstimmen hinterher. Das kann bereits im Spiel eines Pianisten beobachtet werden (Shaffer 1981, Goebel 2001, Goebel et al. 2010). Bei gleich bleibenden Rahmenbedingungen bleibt die Asynchronie ungefähr konstant. Sie unterliegt aber externen Einflüssen, die noch immer nicht alle bekannt sind. Sie wird etwa vom zugrunde liegenden Tempo beeinflusst. Schnelle Tempi lassen weniger Raum für Abweichungen und erzwingen ein präziseres Zusammenspiel.

Rhythmische Ungenauigkeit: Es tauchen auch Abweichungen im Timing auf, die (noch) nicht mit systematischen Ursachen in Verbindung gebracht werden können. Menschen können eben nicht zweimal exakt das Gleiche spielen (Wing & Kristofferson 1973, Miedreich 2000). Selbst wenn motorische Gegebenheiten bei der Bedienung des Instruments (Fingersatz, Bogenführung etc.) mit berücksichtigt werden, bleibt ein Rest an Ungenauigkeit. Sie betrifft die rhythmische Akkuratess des Spiels und bewirkt eine (scheinbar zufällige) zeitliche Versetzung der musikalischen Ereignisse.

Das *Makro-Timing* wird vom Tempo konstituiert. Es ist das Grundmetrum, innerhalb dessen aber noch metrisch exakt gespielt wird. Erst durch Rubato, Asynchronie und rhythmische Ungenauigkeit werden die kleinen meist subtilen Details beigesteuert, das *Mikro-Timing*. Sie füllen die Interpretation mit Lebendigkeit und helfen, den Anschein des Mechanischen zu überwinden. Es folgt die Beschreibung der Modelle, mit denen sie nachgebildet werden. Für die symbolische Zeit wird im folgenden die MIDI-Darstellung verwendet. Es gilt also, für ein jedes Tick-Datum d innerhalb des Musikstücks das entsprechende Millisekunden-Datum $Timing(d)$ zu berechnen.

5.3.1 Makro-Timing

Formalisierung

Der Blick auf ein typisches Notenblatt offenbart eine ganze Reihe von Einzeichnungen, die über das reine Notenmaterial hinaus gehen. Darunter finden sich

Tempoanweisungen wie Allegro, Adagio, Largo, $\downarrow=120$, Ritardando, Rallentando, Accelerando, A Tempo usw. Blendet man alle anderen Informationen aus, zeigt sich, dass die Abfolge all dieser Tempoanweisungen T im Musikstück als eine zeitlich geordnete Liste aufgefasst werden kann, die *Tempomap* M_T . Wer mit modernen Sequenzern arbeitet, ist bereits mit diesem Formalismus vertraut.

$$M_T = (T_0, T_1, \dots, T_n)$$

Im MIDI-Standard wird durch eine Tempoanweisung lediglich ein konstantes Tempo definiert und der Zeitpunkt, ab dem es gilt. Das deckt nicht alle aus der Musik bekannten Anweisungstypen ab. Neben denjenigen Anweisungen, die ein konstantes Tempo setzen, gibt es auch solche, die einen kontinuierlichen Tempoübergang fordern, wie z.B. Ritardando und Accelerando. Deshalb wird die formale Beschreibung von Tempoanweisungen hier nun erweitert. Eine Tempoanweisung T_m ist ein Quintupel der Form

$$T_m = (d_m, tempo_{1,m}, tempo_{2,m}, b_m, i_m) \quad \text{für } 0 \leq m \leq n$$

d_m	Tick-Datum der Anweisung
$tempo_{1,m}$	Anfangstempo in bpm
$tempo_{2,m}$	Endtempo in bpm an d_{m+1}
b_m	Notenwert eines Schlages
i_m	Charakteristik des Tempoüberganges, $i_m \in [0, 1]$

Die Anweisung gilt ab dem symbolischen Datum d_m (in Ticks) und endet mit dem Datum d_{m+1} der Folgeanweisung T_{m+1} , spätestens aber am Ende des Musikstücks. Das wird in MIDI-Dateien durch den EndTrack-Event angezeigt. Innerhalb dieses Intervalls, $[d_m, d_{m+1})$, definiert T_m einen kontinuierlichen Tempoübergang von $tempo_{1,m}$ nach $tempo_{2,m}$. Beide werden in Schlägen pro Minute (bpm) bemessen. Der Notenwert eines Schlages wird von b_m angegeben (Viertelnote $\rightarrow 1/4$, Halbe Note $\rightarrow 1/2$ usw.).

Sind $tempo_{1,m}$ und $tempo_{2,m}$ gleich, handelt es sich um ein konstantes Tempo. Sind sie verschieden, wird ein kontinuierlicher Tempoübergang angezeigt. Dessen Charakteristik wird durch den Parameter i_m gesteuert. Kontinuierliche Tempoübergänge verlaufen bei menschlichen Interpreten in den seltensten Fällen linear (über die symbolische Zeit), was einen eher mechanischen Höreindruck vermitteln würde. Sie zeigen vielmehr ähnliche Charakteristiken wie Potenzfunktionen (Friberg & Sundberg 1999). D.h., dass die Tempoänderung in dem Intervall, in dem sie stattfindet, ungleichmäßig verteilt ist. Der Großteil der Tempoänderung kann in der ersten Hälfte geschehen (*First-Half-Charakteristik*) oder in der zweiten Hälfte (*Second-Half-Charakteristik*). Dies kann unterschiedlich stark ausgeprägt sein, abhängig vom musikalischen Kontext. Der Verlauf

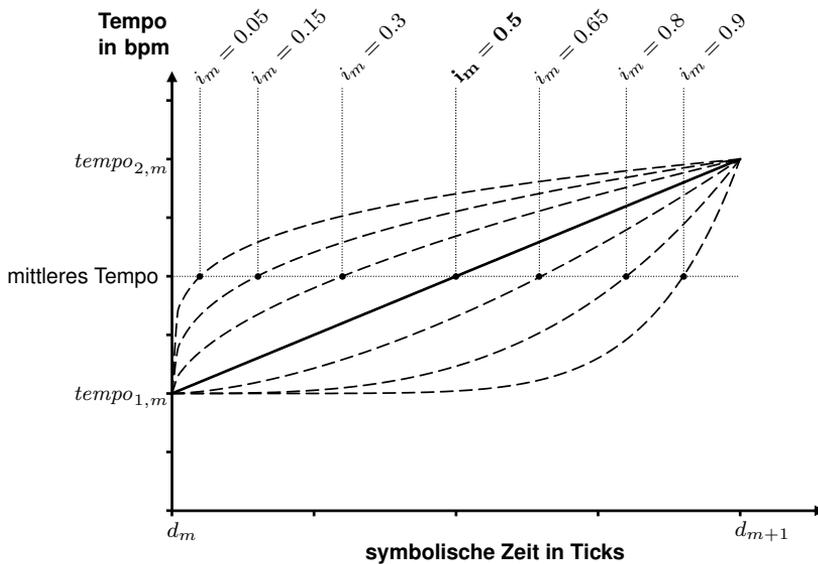


Abbildung 5.2: Sieben Accelerandi mit unterschiedlichen Einstellungen des Parameters i_m und durch Potenzfunktionen umgesetzt (siehe Abschnitt Explizite Tempokurven).

ist tendenziell näher am linearen Verlauf, wenn der Tempoübergang lediglich in ein neues Tempo überleitet, das dann fortgesetzt wird. Spielt der Übergang aber auf einen bestimmten Punkt hin (z.B. das finale Ritardando, das auf den Schlussston hinspielt), sind die Kurvenverläufe tendenziell stärker gekrümmt. Es ist diese Eigenschaft, die durch den Parameter i_m gesteuert wird. Er gibt eine relative Position im Intervall $[d_m, d_{m+1}]$ an. An dieser Stelle zwischen den beiden assoziierten Tempoanweisungen T_m und T_{m+1} soll die Hälfte der Tempoänderung, also das mittlere Tempo $(tempo_{1,m} + tempo_{2,m})/2$ erreicht sein. Dies wird im weiteren auch als *Mean Tempo Condition* bezeichnet. Durch i_m wird damit implizit eine Stützstelle für den Verlauf der Tempokurve vorgegeben (siehe Abbildung 5.2).

$i_m = 0.5$ Das mittlere Tempo ist genau in der Mitte zwischen d_m und d_{m+1} erreicht. Der Tempoübergang ist in sich ausgeglichen. Hierfür kann sogar eine lineare Übergangsfunktion gewählt werden.

$0 < i_m < 0.5$ bewirkt einen zuerst raschen, später nur noch allmählichen Tempoübergang, d.h. *First-Half-Charakteristik*.

$0.5 < i_m < 1$ bewirkt einen zunächst zögernden, dann rascher werdenden Tempoübergang, d.h. *Second-Half-Charakteristik*.

$i_m = 0$ ist ein Spezialfall, bei dem sofort mit $tempo_{2,m}$ begonnen wird.

$i_m = 1$ ist ein Spezialfall, bei dem $tempo_{1,m}$ bis zur nächsten Anweisung beibehalten wird.

Je mehr der Parameter i_m von 0.5 abweicht, desto stärker ist entweder die *First-* oder die *Second-Half-Charakteristik* ausgeprägt und desto ungleichmäßiger ist der Tempoübergang auf die beiden Hälften des Intervalls $[d_m, d_{m+1}]$ verteilt. Näheres zur tatsächlichen Umsetzung solcher kontinuierlicher Tempoverläufe folgt.

Berechnung

Die Tempomap kann auch als eine Folge von Kurvensegmenten interpretiert werden, die entlang der symbolischen Zeitachse jedem Zeitpunkt d vom Beginn des Musikstücks bis zum Ende ein Tempo $Tempo(d)$ zuweist. Diese Tempofunktion ist nicht notwendigerweise stetig. Um nun ein Tick-Datum d , das innerhalb des Definitionsbereichs der Tempomap liegt, in ein Millisekunden-Datum $Ms(d)$ zu konvertieren, muss das Integral der inversen Tempofunktion (die Inter-Tick-Intervall-Funktion oder Timing-Gradient) berechnet werden.

$$Ms(d) = Ms(0) + \frac{60\,000}{4 \cdot PPQ} \left(\sum_{k=0}^{m-1} \int_{d_k}^{d_{k+1}} \frac{1}{Tempo(x) b_k} dx + \int_{d_m}^d \frac{1}{Tempo(x) b_m} dx \right) \quad (5.1)$$

Dabei ist m der Index der letzten Tempoanweisung T_m vor d , also derjenigen Anweisung, für die $d_m < d \leq d_{m+1}$ gilt. $Ms(0)$ ist das Millisekunden-Datum des Anfangs des Musikstücks. Der Rest dieser Berechnung addiert die Integrale eines jeden Kurvensegments bis zum Datum d auf. Dies wird durch die Funktion \widetilde{Ms} approximiert.

$$\widetilde{Ms}(d) = \begin{cases} d & : d \leq d_0 \\ \text{const}(d) + \widetilde{Ms}(d_m) & : tempo_{1,m} = tempo_{2,m} \\ \text{tran}(d) + \widetilde{Ms}(d_m) & : \text{sonst} \end{cases}$$

Die Integrale der einzelnen Kurvensegmente werden von den Funktionen `const` und `tran` berechnet und aufaddiert. Solange keine Tempoanweisung vorliegt

($d \leq d_0$), wird das Tick-Datum identisch auf Millisekunden abgebildet. Wenn die Tempoanweisung ein konstantes Tempo definiert ($tempo_{1,m} = tempo_{2,m}$), berechnet die Funktion `const` das Integral.

$$\text{const}(d) = \frac{60\,000 \cdot (d - \widehat{d}_m)}{tempo_{1,m} \cdot 4 \cdot b_m \cdot PPQ}$$

Wenn durch T_m hingegen ein echter Tempoübergang definiert ist, d.h. $tempo_{1,m}$ und $tempo_{2,m}$ sind verschieden, wird die Berechnung durch die Funktion `tran` ausgeführt. Für diese Funktion werden nun zwei mögliche Lösungsansätze vorgestellt. Im Lösungsansatz *Explizite Tempokurven* wird aus den Parametern der Tempoanweisung die Tempokurve explizit konstruiert und das Integral aus Formel 5.1 durch numerische Integration approximiert. Im Lösungsansatz *Implizite Tempokurven* wird der Kurvenverlauf der Funktion \widehat{M}_s für das fragliche Segment direkt konstruiert, der Weg über die Tempokurve und das Integral quasi umgangen. Die Lösungen sind nicht äquivalent, es ergeben sich also nicht identische Verläufe.

Explizite Tempokurven

Der Verlauf der Tempokurve wird durch eine Potenzfunktion im Einheitsquadrat beschrieben und auf die tatsächlichen Abmessungen, die durch die Tempoanweisung gegeben sind, skaliert (siehe Abbildung 5.2).

$$\text{Tempo}(d) = \left(\frac{d - d_m}{d_{m+1} - d_m} \right)^{p(i_m)} (tempo_{2,m} - tempo_{1,m}) + tempo_{1,m}$$

für $d_m < d \leq d_{m+1}$

Dabei muss der Exponent $p(i_m)$ so gewählt werden, dass die *Mean Tempo Condition*

$$\text{Tempo}((d_{m+1} - d_m)i_m + d_m) = \frac{tempo_{1,m} + tempo_{2,m}}{2} \quad (5.2)$$

erfüllt ist. Das bedeutet, an der durch i_m markierten Tick-Position muss das mittlere Tempo erreicht sein. Diese Bedingung stellt sicher, dass die Charakteristik des Tempoverlaufs durch den Parameter i_m in der vorgesehenen Weise gesteuert wird. Der Exponent lässt sich wie folgt berechnen.

$$p(i_m) = \log_{i_m} 0.5 = \frac{\ln 0.5}{\ln i_m} \quad \text{für } 0 < i_m < 1$$

Die Spezialfälle $i_m = 0$ und $i_m = 1$ werden als konstante Tempi behandelt, wie oben beschrieben. Die Berechnung des Millisekunden-Datums $\text{tran}(d)$ wird mittels numerischer Integration durch $\widetilde{\text{tran}}(d)$ approximiert (Simpson-Summe mit $N/2$ Iterationen, siehe z.B. Quarteroni, Sacco & Saleri 2007, Huckle & Schneider 2002).

$$\widetilde{\text{tran}}(d) = \frac{d - d_m}{3N} \left(\frac{1}{\text{Tempo}(x_0)} + \sum_{k=1}^{\frac{N}{2}-1} \frac{2}{\text{Tempo}(x_{2k})} + \sum_{k=1}^{N/2} \frac{4}{\text{Tempo}(x_{2k-1})} + \frac{1}{\text{Tempo}(x_N)} \right) \frac{60\,000}{4 \cdot b_m \cdot PPQ}$$

$$\text{mit } x_k = d_m + k \cdot \frac{d - d_m}{N} \text{ und geradem } N$$

Die Güte der Approximation kann mit dem Parameter N geregelt werden. Je größer N gewählt wird, desto exakter wird die Timing-Kurve berechnet. Der Fehler der Approximation ist gegeben durch

$$\left| \text{tran}(d) - \widetilde{\text{tran}}(d) \right| \leq \frac{(d - d_m)^5}{180 N^4} \max_{x \in [d_m, d]} \left| \left(\frac{1}{\text{Tempo}(x)} \right)^{(4)} \right|$$

Doch wie groß darf dieser Fehler sein? Die zeitliche Versetzung zweier Klangergebnisse nimmt der Mensch ab einer Versetzung von 2 bis 70 Millisekunden wahr, je nach Akustik und Länge der Einschwingphase der Ereignisse (Fricke & Louven 2008). Bei sehr langen Einschwingphasen kann diese »Verwischungsschwelle« sogar noch deutlich größer sein. Unterschiedliche Charakteristiken einer Temposchwankung kann der Mensch sogar noch schwerer differenzieren, da kein exaktes Signal nebenher läuft und Bezugspunkte liefert. Selbst Unterschiede von mehreren Sekunden können hier nicht weiter auffallen. Im Abschnitt Diskussion wird eine diesbezügliche Studie beschrieben und ausgewertet. Größere Exaktheit ist also nur in Fällen notwendig, in denen es um Fragen der Synchronisation geht. Generell steht es der Implementierung frei, selbst zu entscheiden, welche Genauigkeit, also welches N , gewünscht ist. Je größer N gewählt wird, desto aufwendiger wird allerdings auch die Berechnung.

Fünf Klangbeispiele für unterschiedliche Tempoverläufe mit den in Tabelle 5.1 aufgeführten Einstellungen werden auf der beiliegenden Audio-CD demonstriert. Das Notenmaterial dazu ist in Anhang E abgedruckt.



CD-Track	$tempo_{1,m}$	$tempo_{2,m}$	i_m	N
3 Accel.	40	120	0.7	512 (jede Note)
4 Accel.	40	120	0.3	512
5 Rit.	120	40	0.5	512
6 Rit.	120	40	0.5	2
7 Rit.	120	40	0.8	512

Tabelle 5.1: Die Eckdaten der Tempoverläufe in den Klangbeispielen.

Implizite Tempokurven

Im vorigen Ansatz wurden die Tempokurven explizit modelliert. Aus ihnen wurde dann die Abbildung von symbolischer Zeit auf physikalische Zeit (Millisekunden) berechnet. Diese Abbildung wurde von $\widetilde{M}s$ in der Funktion \widetilde{tran} durch numerische Integration implementiert. Der hier vorgestellte Lösungsansatz *Implizite Tempokurven* geht den umgekehrten Weg. Das fragliche Kurvensegment von $\widetilde{M}s$ wird direkt konstruiert. So ist für die Millisekunden-Berechnungen, die später in Echtzeit zur Wiedergabe geschehen sollen, nicht mehr der Weg über die numerische Integration der inversen Tempofunktion notwendig. Der Ansatz ist insbesondere dann von Nutzen, wenn die Millisekunden-Daten bereits vorliegen (z.B. aus Messungen) und approximiert werden sollen. Die *Mean Tempo Condition* (Bedingung 5.2) soll auch hier als Steuerbedingung für die Charakteristik des Tempoverlaufs dienen. Die Herausforderung besteht nun also darin, dass der Potenzverlauf wenigstens soweit angenähert werden sollte, dass die *Mean Tempo Condition* erfüllt ist.

Der Timing-Gradient, der bei einer kontinuierlichen streng monotonen Tempoänderung entsteht, ist ebenfalls streng monoton. Die Timing-Kurve (sowieso streng monoton wachsend) hat in dem Intervall also auch keine Wendepunkte. Dieser Verlauf wird zunächst durch eine quadratische Bézierkurve approximiert (Farin 2002). Sie wird durch drei Kontrollpunkte P_0 , P_1 und P_2 aufgespannt und ist folgendermaßen definiert:

$$P(t) = (1-t)^2 P_0 + 2t(1-t) P_1 + t^2 P_2 \quad \text{für } 0 \leq t \leq 1$$

Die Position des ersten Kontrollpunktes (der Anfang des Kurvensegmentes) folgt aus dem vorhergehenden Kurvensegment.

$$P_0 = (d_m, \widetilde{M}s(d_m))$$

Das Tick-Datum des dritten Kontrollpunktes ist ebenfalls bekannt. Wenn das Tempomodell dazu eingesetzt wird, empirische Messungen zu approximieren,

kann das Millisekunden-Datum des dritten Kontrollpunktes direkt auf einen der Messpunkte gelegt werden. Liegt kein solches Datum vor, kann es auch nach dem vorigen Lösungsansatz *Explizite Tempokurven* berechnet werden. Da es sich hierbei um einen Vorverarbeitungsschritt handelt, der nur einmal ausgeführt werden muss, kann der Wert sehr genau bestimmt werden.

$$P_2 = (d_{m+1}, \widetilde{Ms}(d_{m+1}))$$

Auch die Anstiege (Timing-Gradienten) an beiden Punkten sind bekannt. Sie ergeben sich aus Anfangs- und Endtempo, $tempo_{1,m}$ und $tempo_{2,m}$.

$$g_{z,m} = \frac{60\,000}{tempo_{z,m} \cdot 4 \cdot b_m \cdot PPQ} \quad \text{für } z \in \{1, 2\}$$

Der mittlere Kontrollpunkt P_1 kann nun konstruiert werden.

$$P_1 = \left(d_{P_1}, g_{1,m}(d_{P_1} - d_m) + \widetilde{Ms}(d_m) \right)$$

mit
$$d_{P_1} = \frac{g_{1,m}d_m - g_{2,m}d_{m+1} - \widetilde{Ms}(d_m) + \widetilde{Ms}(d_{m+1})}{g_{1,m} - g_{2,m}}$$

Die so approximierten Verläufe zeigen tendenziell die gewünschten Charakteristiken, auch wenn sie nicht identisch mit denen des expliziten Modells sind (siehe Abbildung 5.3). Prinzipiell müssen sie auch nicht identisch sein, schließlich sind beide nur Approximationen von beim Menschen beobachteten Charakteristiken, welche durch die *Mean Tempo Condition* präzisiert werden. Gerade diese Steuerbedingung erfüllt das implizite Modell in seiner momentanen Form aber nicht. Das mittlere Tempo ist nicht an dem durch i_m angezeigten symbolischen Zeitpunkt erreicht.

Die Bézierkurve, so wie sie bislang konstruiert ist, stellt vielmehr eine Initialisierung dar. Um davon ausgehend die *Mean Tempo Condition* zu erfüllen, wird ihr Verlauf weiter angepasst. Dafür gibt es zwei Möglichkeiten.

Das Ziel verschieben: P_1 leitet sich aus den beiden äußeren Kontrollpunkten und den Anstiegen an diesen Stellen ab. Von denen steht P_0 fest. Aber P_2 , das Ende des Kurvensegments, kann entlang der Millisekunden-Achse verschoben werden. Dadurch verändert sich auch die Position von P_1 die näher an P_0 oder P_2 heranrückt. Gleichermassen verändert sich auch die korrespondierende Tempokurve. Eine einfache Möglichkeit für eine näherungsweise Lösung ist die Bisektion, die auf die *Mean Tempo Condition* testet, im Intervall:

$$\widetilde{Ms}(d_{m+1}) \in \left((d_{m+1} - d_m)g_{1,m} + \widetilde{Ms}(d_m), (d_{m+1} - d_m)g_{2,m} + \widetilde{Ms}(d_m) \right)$$

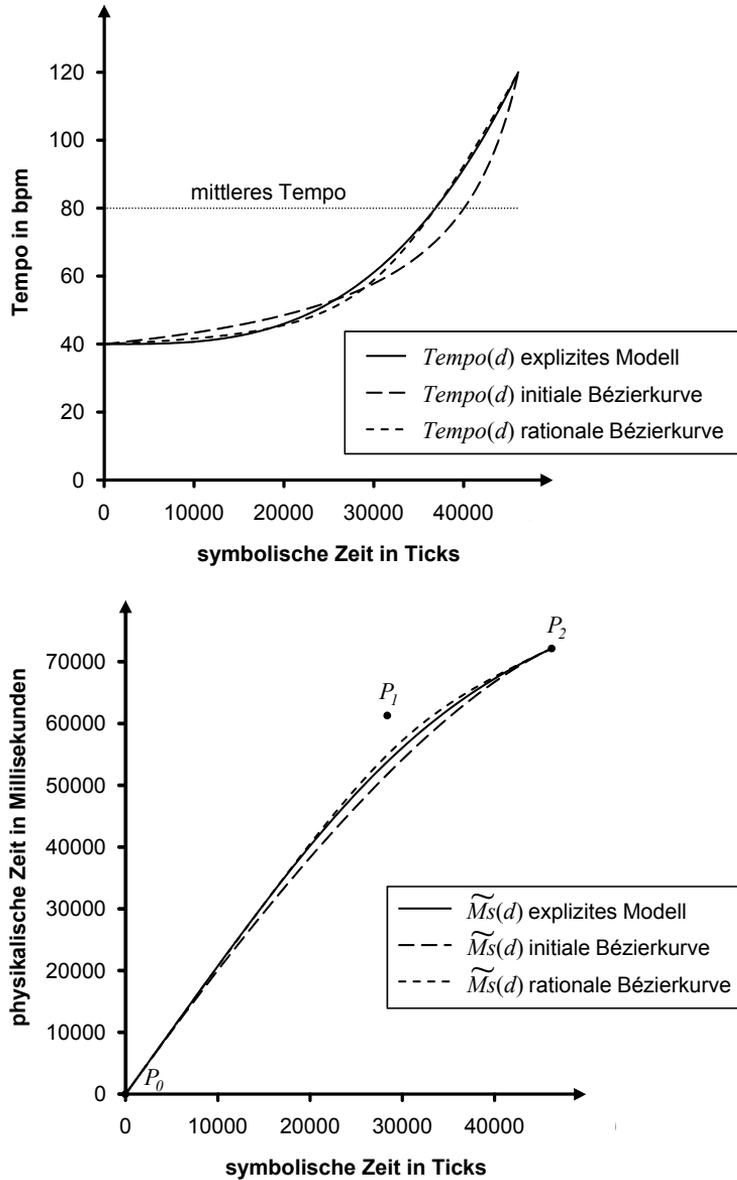


Abbildung 5.3: Die Tempo- und Timing-Kurven des expliziten und impliziten Modells im Vergleich.

Bei diesem Ansatz ist wichtig, zu bedenken, dass die resultierende Timing-Kurve nicht nur lokal von der des expliziten Modells (und auch von der initialen Bézierkurve) abweicht. Innerhalb des Zeitraumes $(d_m, d_{m+1}]$ werden keine Abweichungen kompensiert. Am Ende des Segmentes sind die beiden Modelle folglich nicht mehr synchron. Wenn Synchronie gewollt ist (z.B. bei der Approximation von Messdaten), muss auch P_2 an der initialisierten Position feststehen. Dafür ist die nächste Methode besser geeignet.

Rationale Bézierkurven: Bei einer rationalen Bézierkurve sind die Kontrollpunkte mit Gewichten w_0, w_1, w_2 versehen, durch die der Kurvenverlauf weiter beeinflusst werden kann:

$$P(t) = \frac{w_0(1-t)^2P_0 + w_12t(1-t)P_1 + w_2t^2P_2}{w_0(1-t)^2 + w_12t(1-t) + w_2t^2}$$

Durch Anpassen der drei Gewichte kann das Verhalten des expliziten Modells sehr viel genauer nachgebildet werden. Verschiedene i_m verlangen aber nach unterschiedlichen Gewichts-konstellationen. Eine einzige konstante Voreinstellung der Gewichte kann also allenfalls für ein spezielles i_m geeignet sein, nicht für alle. Um die *Mean Tempo Condition* zu erfüllen, genügt es aber schon, wenn w_0 und w_2 konstant bleiben und lediglich das Gewicht des mittleren Kontrollpunktes w_1 angepasst wird. Alle weiteren Abweichungen, etwa vom expliziten Modell, liegen innerhalb der Varianz, die auch bei menschlichen Interpreten zu beobachten ist und von menschlichen Hörern kaum zu differenzieren ist, wie die Studie im folgenden Abschnitt zeigt.

Diskussion

Von allen Timing-Berechnungen, auch für die noch folgenden Mikro-Timing-Aspekte, ist das Makro-Timing das aufwendigste. Wenn es offline (vor der Wiedergabe) berechnet werden kann, ist das nicht weiter problematisch. Soll es aber in Echtzeit zur Wiedergabe ausgeführt werden, kann die numerische Integration (*explizite Tempokurven*) besonders bei langen Tempoübergängen zu aufwendig werden. Die *impliziten Tempokurven* stellen eine Alternative dar, die direkt in der Tick-Millisekunden-Domäne arbeitet. Sie bietet sich insbesondere dann an, wenn zwischen festen Synchronisationspunkten, etwa empirischen Messungen, approximiert werden sollte, denn hier können Anfang und Ende des Kurvensegmentes direkt in der Domäne platziert werden. Obwohl die beiden Modelle den gleichen Steuerungsbedingungen unterliegen, sind sie nicht äquivalent! Die *impliziten Tempokurven* können an die *expliziten* nur angenähert werden, wenigstens soweit, dass die *Mean Tempo Condition* (Bedingung 5.2) erfüllt ist – die

Typ	i_m -Einstellungen			korrekte Antworten		
	1. Beispiel	2. Beispiel	Differenz	#	%	Signifikanz
Acccel.	0.7	0.3	0.4	16	88.9	0.001
Rit.	0.5	0.4	0.1	5	27.8	0.096
Acccel.	0.6	0.8	0.2	9	50.0	1.0
Rit.	0.4	0.4	0	8	44.4	0.815
Acccel.	0.4	0.3	0.1	9	50.0	1.0
Rit.	0.3	0.6	0.3	14	77.8	0.031

Tabelle 5.2: Die Resultate der Tempostudie. 18 Probanden sollten 6 Paare von Tempoübergängen als identisch oder verschieden klassifizieren. Sie unterschieden sich lediglich in der Einstellung des Parameters i_m . Die Signifikanz wurde mittels Binomialtest bestimmt (Bortz 2005).

Grundlage dafür, dass der Parameter i_m das Modell in der vorgesehenen Weise steuert.

Ist mehr nötig? Welche Genauigkeit ist notwendig? Auch bei menschlichen Interpretieren kann eine recht große Streuung beobachtet werden, so dass es nicht möglich ist, eines der Modelle als das bessere oder adäquatere zu bestimmen. Aber sind sie für den menschlichen Hörer äquivalent? Was kann der menschliche Hörer noch unterscheiden? Um das herauszufinden wurde eine Studie mit 18 Teilnehmern durchgeführt (7 Musiker, 11 Nichtmusiker). Diesen wurden mehrere Tempoverläufe in ein und derselben Musik (16 4/4-Takte) jeweils paarweise vorgespielt. Ritardandi begannen stets mit 40 bpm und endeten mit 120 bpm, Accelerandi entsprechend umgekehrt. Sie wurden alle mit dem Modell *Explizite Tempokurven* mit einer Auflösung von $N = 512$ gerendert; d.h. ein Segment der Simpson-Summe für jede 16tel-Note. Einzig der Parameter i_m wurde variiert. Die Probanden hatten zu entscheiden, ob die gehörten Verläufe gleich (nicht unterscheidbar) oder verschieden sind. Zur leichteren Orientierung wurden auch die Noten des Musikstücks ausgeteilt (siehe Anhang E).

Die Ergebnisse der Studie sind in Tabelle 5.2 zusammengestellt. Die Probanden konnten große Unterschiede, die sogar mit gegensätzlichen Charakteristiken einher gingen, deutlich erkennen. Das war im 1. und 6. Paar von Tempoübergängen der Fall: *First-Half-Charakteristik* ($i_m < 0.5$) gegenüber *Second-Half-Charakteristik* ($i_m > 0.5$). Alle anderen Fälle konnten nicht signifikant klassifiziert werden. Dabei war bei einer i_m -Differenz von 0.2 (3. Paar) der Tempounterschied auf einer Strecke von 10 Takten bereits 10–23 bpm groß. Die resultierende Timing-Differenz betrug 15 Sekunden (Beispiel 2 war 26% länger als Beispiel 1). Auch bei einer i_m -Differenz von 0.1 (2. und 5. Paar) betrug der Tempounterschied 10 Takte lang noch 5–8 bpm, was noch immer 4 Sekunden Unterschied

in der Spielzeit ausmachte (Beispiel 1 war jeweils ca. 7% länger als Beispiel 2). Selbst identische Verläufe (4. Paar) konnten nicht signifikant als identisch identifiziert werden.

Die Studie zeigt, dass andere Tempoverläufe, die sich innerhalb einer gewissen Umgebung um eine Tempokurve T_m bewegen, als Alternativen verwendet werden können. Diese Umgebung kann durch den Parameter $i_m \pm 0.1$ aufgespannt werden. Für den Hörer sind sie nicht unterscheidbar. Speziell im Falle der *impliziten Tempokurven* mittels rationaler Bézierkurven werden auch Timingunterschiede kompensiert, so dass nicht einmal eine zeitliche Differenz entsteht. Für die *expliziten Tempokurven* bedeutet dies, dass bereits eine grobe numerische Integration ($N \in \{2, 4\}$) in den meisten Fällen ausreicht. Dies macht den expliziten Ansatz sehr schnell.

Allerdings gelten diese Beobachtungen und Schlussfolgerungen nur für kontinuierliche Tempoübergänge, bei denen sich das Tempo ständig ändert. Sie sollten nicht auf konstante Tempi und diskrete (stufenweise) Tempoänderungen übertragen werden. Konstante Tempi lassen sich leichter unterscheiden, und plötzliche Tempoänderungen sind leichter wahrzunehmen. Folglich gelten die hier gemachten Schlüsse auch nicht für Tempoübergänge durch mehrere stufenweise Tempoänderungen, was der gängigen Praxis im MIDI-Bereich entspricht.

5.3.2 Mikro-Timing

Im Gegensatz zu den Makro-Timing-/Tempo-Anweisungen sind Mikro-Timing-Anweisungen selten im Notenbild fixiert. Oft werden sie vom Musiker unbewusst gestaltet und treten als Bestandteil eines ganzen Merkmalskomplexes auf. Die lautstärkemäßige Betonung von metrischen Schwerpunkten oder einzelnen Noten kann bspw. mit einer Dehnung der zu betonenden Noten einhergehen, wie schon Quantz (1752) beschrieb. Es sind die subtilen Details, denen sich auch der Hörer meist nicht bewusst wird, die aber seinen Höreindruck mit prägen. Die Musik »groovt« erst durch die systematischen Verzerrungen, die das Rubato in das Timing einbringt, im Blues und Jazz (swingende Metrik, siehe Schütz 2008) ebenso wie im Wiener Walzer oder im barocken Menuett (Perl 1998).

Rubato, Asynchronie im Ensemblespiel und rhythmische Ungenauigkeiten bereichern das Timing um sowohl systematische als auch zufällige Details. Zu den systematischen zählen Rubato und Asynchronie. Die rhythmische Ungenauigkeit ist eine zufallsbehaftete Abweichung. Sie bleiben im Laufe eines Musikstücks nicht gleich, sondern können sich ändern. Eine schwer zu spielende Stelle kann zu größerer rhythmischer Ungenauigkeit führen, kann dazu führen, dass eine Stimme hinter dem Rest des Ensembles »hängt« und erst später

Ex. 2. 11 *Cascata doppia* (Caccini, 1602)

a Written



b Performed



c Performed



Ex 2. 12 Rhythmic alteration for the sake of text (Grimarest, 1707)

a Approximately performed



Et vous...

b Written by Lully, 1676



Et vous me lais - se - rez mou - rir.

b.c.



Abbildung 5.4: Der Versuch, Rubato-Verzerrungen auszunotieren (Hudson 2004, Seite 29).

wieder aufschließt. In Korrespondenz zur *Tempomap* werden auch die Mikro-Timing-Merkmale in zeitlich geordneten Listen organisiert, der *Rubatomap* M_R , der *Asynchronmap* M_S und der *Imprecisionmap* M_I .

$$M_R = (R_0, R_1, \dots, R_r)$$

$$M_S = (S_0, S_1, \dots, S_s)$$

$$M_I = (I_0, I_1, \dots, I_i)$$

Rubato

Das Rubato ist eine systematische Verzerrung der symbolischen Zeit. Es beschreibt Dehnungen und Stauchungen des Metrums, die sowohl als singuläre als auch regelmäßig wiederkehrende Merkmale in Erscheinung treten. Der Musiker mag einen Lauf zögernd beginnen und dann beschleunigen oder die Eins eines jeden Taktes immer etwas mehr dehnen als den Rest. Abbildung 5.4 zeigt, wie Hudson (2004) einmal versuchte, diese Verzerrungen auszunotieren. Eine Rubatoanweisung R_m definiert das *Rubatoschema*, das Muster dieser Verzerrungen (veranschaulicht in Abbildung 5.5). Es hat die Länge f_m (in Ticks) und wird

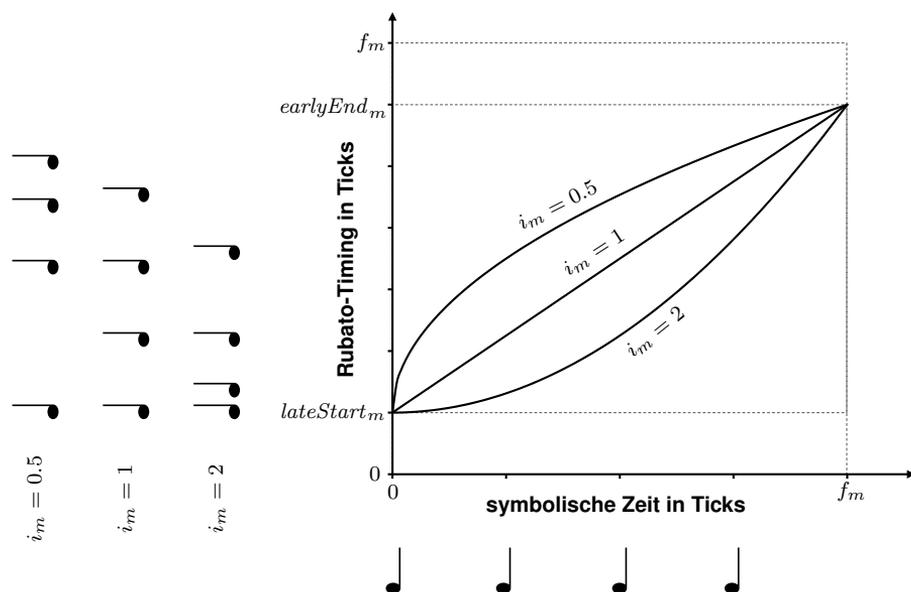


Abbildung 5.5: Drei Rubatoschemata mit jeweils unterschiedlichen Einstellungen des Parameters i_m und einer Framelänge f_m von vier Vierteln.

ab dem Datum d_m wiederholt angewendet, bis die Folgeanweisung R_{m+1} oder das Ende des Musikstücks erreicht ist.

$$R_m = (d_m, f_m, i_m, lateStart_m, earlyEnd_m) \quad \text{für } 0 \leq m \leq r$$

d_m Tick-Datum der Anweisung

f_m Länge eines Zeitfensters (*Rubato Frame*) in Ticks

i_m Charakteristik des Rubatos, $i_m \in \mathbb{R}^+$

$lateStart_m$ zusätzliche Verzögerung am Beginn

$earlyEnd_m$ vorgezogenes Ende

Eine zentrale Eigenschaft des Rubatoschemas ist, dass es die Abweichungen vom metrisch exakten Spiel innerhalb des Zeitfensters wieder kompensiert. Das Rubato verändert also nicht das durchschnittliche Grundmetrum. Damit sind alle Zeitpunkte $d_m + k f_m$ mit $k \in \mathbb{N}_0$ synchron mit dem Grundmetrum. Innerhalb des Zeitfensters wird entweder eine Dehnung mit anschließend kompensierender Stauchung oder aber eine Stauchung mit anschließender Dehnung des Metrums beschrieben. Durch die Potenzfunktion x^{i_m} wird dieses Verhalten innerhalb des Intervalls $[0, 1]$ nachgebildet, das problemlos auf $[0, f_m]$ skaliert werden

kann. Durch den Exponenten i_m wird die Charakteristik der Verzerrung gesteuert.

$i_m = 1$ entspricht der identischen Abbildung von metrischer Zeit auf Rubatozeit ohne jede Verzerrung.

$0 < i_m < 1$ bewirkt eine Dehnung am Beginn des Zeitfensters, die durch Stauchung zum Ende hin kompensiert wird.

$i_m > 1$ bewirkt eine Stauchung zu Beginn mit anschließender Dehnung.

Durch $lateStart_m$ kann eine zusätzliche Verzögerung am Beginn des Fensters gesetzt werden. Der Parameter $earlyEnd_m$ fügt eine zusätzliche Pause am Ende des Fensters ein. Das Timing wird hier also innerhalb des Frames noch zusätzlich zusammengestaucht. Damit die Stauchung nicht genauso groß oder größer als das Zeitfenster wird, gilt für beide Parameter

$$0 \leq lateStart_m < earlyEnd_m \leq 1$$

wobei 0 dem Beginn des Zeitfensters entspricht und 1 dem Ende (siehe Abbildung 5.5). Die Rubato-Verzerrungen werden von der Funktion *Rubato* implementiert. Sie bildet symbolische Zeit auf symbolische Zeit ab.

$$Rubato(d) = frameDate(d) + \left[\left(\frac{(d - d_m) \bmod f_m}{f_m} \right)^{i_m} (earlyEnd_m - lateStart_m) + lateStart_m \right] f_m$$

mit $frameDate(d) = d - [(d - d_m) \bmod f_m]$

und $d_m < d \leq d_{m+1}$



CD-Tracks
8, 9, 10

Hörbeispiele für verschiedene Rubati sind auf der beiliegenden Audio-CD gegeben. Am Beispiel des in Anhang E abgedruckten Notenmaterials werden verschiedene Rubati demonstriert. Dabei werden die Frame-Länge f_m und die Charakteristik i_m variiert. Zwei Musikbeispiele werden mit dem dritten Satz, *Grave*, aus G. P. Telemanns *Trompetenkonzert D-Dur* (TWV 51:D7) gegeben. Während die Version in Track 9 relativ gerade gespielt wird, ist in Track 10 ein fast swingendes Spiel zu hören.

Asynchronie

Die Asynchronie im Ensemblespiel ist eine einfache zeitliche Versetzung einzelner Stimmen. Es lässt sich bspw. beobachten, dass führende Stimmen (Solist,

Melodiestimme etc.) oft etwas voraus spielen. Die Begleitstimmen sind hingegen hinterher. Allgemeiner formuliert, spiegelt sich die Aufgabe der Stimme im Ensemble auch in ihrer zeitlichen Versetzung wider (Rasch 1979, Repp 1996). Auch die Motorik beim Spiel des Instruments kann Ursache für Asynchronie sein (Goebel 2001). Die Modellierung solcher Versetzungen ist denkbar einfach. Die Anweisung

$$S_n = (d_n, s_n) \quad \text{für} \quad 0 \leq n \leq s$$

definiert das Tick-Datum d_n , von dem an alle musikalischen Ereignisse der betreffenden Stimme, d.h. alle MIDI-Events des betreffenden Kanals, um einen Millisekunden-Wert s_n versetzt werden. Die Versetzung endet mit dem Datum d_{n+1} .

Zwei Hörbeispiele sind auf der beiliegenden Audio-CD gegeben – eine Aussetzung des Volksliedes *Es zogen auf sonnigen Wegen*. Die Melodie-führende Bratsche ist in Track 11 um 50 Millisekunden vor die Klavierbegleitung gesetzt. Hier ist sie die führende Stimme. In Track 12 ist dies umgekehrt; das Klavier führt.



CD-Tracks
11 und 12

Rhythmische Ungenauigkeit

Rhythmisch ungenaues Spiel bedeutet, dass die betroffenen Klangereignisse (etwa Töne, rhythmische Impulse, MIDI-Events) nicht zeitlich exakt gespielt werden, sondern mehr oder weniger um den exakten Zeitpunkt streuen. Es gibt eine Vielzahl möglicher Gründe dafür, etwa

- externe Einflüsse, welche die Konzentration des Spielers ablenken,
- Varianzen in der Mechanik von Instrumenten (verschiedene Klaviere spielen sich unterschiedlich),
- Beeinträchtigungen der Motorik des Spielers (Ermüdung bei langen Konzerten, bei Kälte werden die Finger steif – ein typisches Organistenproblem).

Leech-Wilkinson (2010) zeigt am Beispiel des Pianisten Alfred Cortot aber auch, dass Menschen zu überaus exakten Reproduktionen fähig sind. Während also die Varianz in der rhythmischen Exaktheit bei Laien tendenziell groß ist, kann sie bei gut trainierten Profimusikern auch außerordentlich gering sein. In Tapping-Experimenten (siehe auch Abschnitt 5.6) mit 25 Teilnehmern, davon 8 studierten Musikern, waren die studierten Musiker im Mittel 1 Millisekunde

hinter dem exakten Onset (Standardabweichung: 44 Millisekunden). Die übrigen Teilnehmer (Laien- und Nicht-Musiker) waren im Mittel 20 Millisekunden hinter dem exakten Onset (Standardabweichung: 60 Millisekunden).⁶

Die rhythmische Ungenauigkeit wird durch eine Gauß-Verteilung mit der Standardabweichung σ (in Millisekunden) und Erwartungswert 0 modelliert. Extreme Ausreißer können allerdings zu unplausiblen Resultaten führen. Sie werden deshalb an der oberen und unteren Grenze einer gewissen Umgebung (exaktes Millisekundendatum $\pm \delta_j$, ebenfalls in Millisekunden) geclippt.

$$I_j = (d_j, \sigma_j, \delta_j) \quad \text{für } 0 \leq j \leq i$$

Der Parameter d_j gibt das Tick-Datum der Anweisung an. Ab dieser Position im Musikstück wird die Millisekunden-Abweichung jedes NoteOn-Events in der MIDI-Sequenz mit der Funktion $Rand(\sigma_j, \delta_j)$ berechnet. Ihr Wertebereich ist $[-\delta_j, \delta_j]$.

Timing komplett

Um das endgültige Millisekunden-Datum eines Tick-Zeitpunktes d zu berechnen, werden die Makro- und Mikro-Timing-Modelle in der Funktion *Timing* zusammengeführt. Die Abfolge der Konvertierungsschritte wird in Abbildung 5.6 dargestellt. Die Funktion ist folgendermaßen definiert,

$$Timing(d) = \widetilde{Ms}(Rubato(d)) + s_n + Rand(\sigma_j, \delta_j)$$

wobei die Indizes n und j die jeweils letzten Anweisungen S_n und I_j vor d anzeigen.

$$\begin{aligned} d_n < d \leq d_{n+1}, & \quad S_n, S_{n+1} \in M_S \\ d_j < d \leq d_{j+1}, & \quad I_j, I_{j+1} \in M_I \end{aligned}$$

Diskussion

Das Makro-Timing mag der auffälligere Timing-Aspekt sein. Schließlich werden große Temposchwankungen leichter wahrgenommen als die kleinen Abweichungen, die das Mikro-Timing beisteuert. Doch ohne diese subtilen Details wirkt auch ein sehr expressiv gestaltetes Tempo immer noch steril, zu perfekt.

⁶Die Stimuli der Studie waren 8 Takte eines barocken Menuettes, 12 mal wiederholt, in 4 Varianten. Für die Auswertung wurden jeweils die letzten 5 Wiederholungen verwendet; das macht zusammen 9700 Taps.

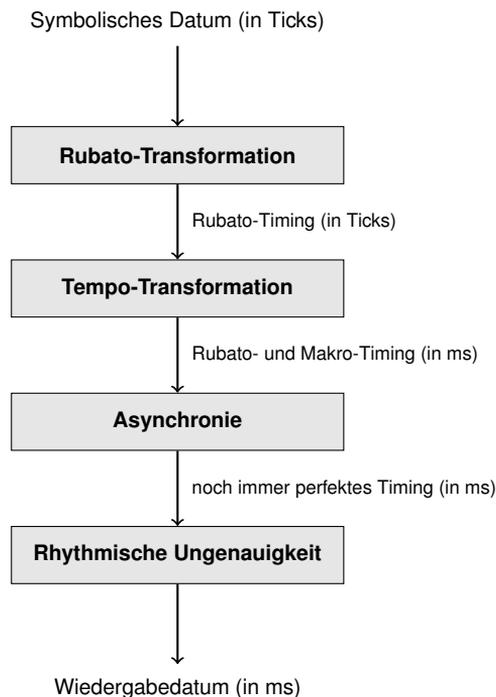


Abbildung 5.6: Die Abfolge der Einzelkonvertierungen bei der Umrechnung von symbolischer Zeit in physikalische Zeit.

Erst Rubati, Asynchronie zwischen den Stimmen und rhythmische Ungenauigkeiten beleben das Spiel.

Dabei sind die zufallsbehafteten Ungenauigkeiten lediglich ein Zugeständnis an die Tatsache, dass auch heute noch nicht alle Einflussfaktoren bekannt sind und nicht alle Abweichungen systematisch erklärt werden können. Sie sind das Grundrauschen, über das auch die Musiker selbst nicht erhaben sind. Auch Leech-Wilkinson (2010) wertet seine Beobachtungen im Falle des Pianisten Alfred Cortot als Ausnahmefall. Cortot schaffte es, im Abstand von einem Jahr mehrere Chopin-Interpretationen nahezu identisch zu reproduzieren. Die Realität im Musikeralltag sieht anders aus – insbesondere im Zusammenspiel mit anderen Musikern. Hier beeinflusst man sich gegenseitig und reagiert auf das, was in den anderen Stimmen vorgeht. Das Zusammenspiel mehrerer Musiker kommt eher einem homeostatischen, d.h. selbstausgleichenden, System gleich. Solche Wechselwirkungen sind in den hier vorgestellten Modellen nicht enthal-

ten. Sie könnten noch jene schwer an Einzelmerkmalen festzumachende Spannung beisteuern, welche die Eigendynamik des Ensemblespiels mit sich bringt.

Eine Schnittstelle dafür kann der Formalismus der Asynchronie sein. Er ermöglicht im Prinzip die Millisekunden-genaue Versetzung jeder einzelnen Note. Das ist manuell sehr aufwendig und deshalb unpraktikabel. Von Hand wird man sich damit begnügen, konstante Versetzungen über längere Strecken zu definieren, also eine grobe Asynchronie. Werden jedoch die Wechselwirkungen zwischen den Stimmen simuliert, kann daraus auch eine detailliertere *Asynchrony-map* generiert werden. In diese kann bspw. die Simulation der Motorik beim Spielen eines Instruments einfließen, etwa die durch einen bestimmten Fingersatz hervorgerufenen Verschiebungen einzelner Noten sowie Arpeggien⁷. Solche Effekte wurden u.a. von Parncutt (1998), Parncutt, Sloboda & Clarke (1999), Mazzola et al. (2002) untersucht und beschrieben.

5.4 Dynamik

Unter dem Begriff Dynamik wird in der Musik alles zusammengefasst, was die Lautstärke von musikalischen Ereignissen anbelangt. Jedoch verweist der Begriff Lautstärke hier nicht nur auf die Amplitude des Audiosignals, wie es intuitiv zu erwarten wäre. Er bezieht auch psychoakustische Parameter mit ein. Dazu gehört die Klangfarbe. Ein lauter Ton *klingt* laut, zumeist schärfer weil obertonreicher, egal welche Amplitude er hat. Dieser Aspekt bleibt beim reinen Blenden (Fading) unberücksichtigt; deshalb ist das Blenden auch kein vollwertiger Ersatz für Dynamik. Es reicht eben nicht aus, nur die Amplitude des Audiosignals zu verändern, wenn eigentlich seine Dynamik verändert werden soll. Fabiani (2009, 2010) stellt ein Verfahren zur Resynthese der Dynamik monophoner Audioaufnahmen vor, das die spektralen Veränderungen mit einrechnet. Auch die Einschwingzeit eines Tones kann das subjektive Urteil über dessen Lautstärke beeinflussen. Bei einem extrem kurzen Attack, also einem sehr plötzlichen Energieanstieg, erscheint der Ton lauter als bei einer sanften längeren Einschwingphase. Selbst die Tonhöhe beeinflusst das Lautstärkeempfinden, denn das menschliche Ohr nimmt nicht alle Frequenzen gleich laut wahr. Die subjektiv wahrgenommene Lautstärke wird in psychoakustischer und musikpsychologischer Literatur auch als *Lautheit* bezeichnet. Ein verbreitetes Lautheitsmodell ist die *Zwicker-Lautheit*. (Kopiez & Auhagen 1998, Fastl & Zwicker 2007, Bradter 2007)

⁷Beim Arpeggieren werden zeitgleich notierte Töne zeitlich versetzt, d.h. kurz nacheinander, angeschlagen.

Glücklicherweise kann heute bei den meisten Synthesizern und Samplern davon ausgegangen werden, dass alle diese Informationen in einen einzigen Lautstärkeparameter codiert sind. Dieser eine Wert steuert sowohl die akustischen als auch die psychoakustischen Toneigenschaften. Wo dies nicht der Fall ist, oder die Feinabstimmung nicht zufrieden stellt, kann meist noch manuell nachjustiert werden, so dass schließlich für jede zu spielende Note nur ein einziger Lautstärkewert angegeben werden muss. Im MIDI-Standard ist es die Velocity, also die Geschwindigkeit des Tastenanschlages. Unterschiedliche Synthesizer und Sampler bilden diesen Wert nicht einheitlich ab, wie Dannenberg (2006) feststellt. Die in den folgenden Abschnitten erarbeiteten Dynamik-Modelle erlauben deshalb auch die Umparametrisierung der Abbildung auf die MIDI-Velocity-Domäne.

Ebenso vielschichtig wie das Verständnis des Begriffs Lautstärke ist auch die Dynamik als musikalische und interpretatorische Gestaltungsdomäne. Wie bereits das Timing ist auch die Dynamik ein Zusammenspiel mehrerer Gestaltungsebenen. Auch hier kann zwischen Makro- und Mikro-Merkmalen unterschieden werden. Hatte das Makro-Timing eine Grundtempokurve definiert, so definiert die Makro-Dynamik eine Grundlautstärkekurve. Ebenso wie das Mikro-Timing Varianzen relativ zu diesem Grundtempo darstellte, steuern die Mikro-Dynamik-Aspekte eine Reihe von Details relativ zur Grundlautstärke bei.

Grundlautstärke: Die Grundlautstärke, oder Makro-Dynamik, ist die prominenteste dynamische Gestaltungsebene. Sie wurde im Laufe der Musikgeschichte zunehmend auch im Notenbild eingezeichnet und durch die Komponisten vorgeschrieben. Am Ende der Renaissance und im Frühbarock war das Vokabular noch wenig ausdifferenziert; meist wurde nur zwischen laut/*forte* und leise/*piano* unterschieden (Bernhard 1657), beispielhaft dafür steht die *Sonata Pian e Forte* aus der Sammlung *Sacrae Symphoniae* von Giovanni Gabrieli (1597). Alle übrigen Details oblagen den Interpreten. Bis zur Romantik entwickelte sich daraus das selbst feinste Nuancen erfassende Vokabular, das in der musikalischen Praxis noch heute gebräuchlich ist. Dazu gehören sowohl diskrete Lautstärkeanweisungen (Stufendynamik) als auch kontinuierliche Lautstärkeübergänge (*Crescendo*, *Decrescendo*). Sie dienen der Gestaltung von Phrasen- und Spannungsbögen, von Echoeffekten, außerdem der Verdeutlichung von ausdrucksmäßigen Gegensätzen (*sanft/hart*, *entspannt/erregt*) und dazu, das Mischungsverhältnis der Stimmen zu regeln, z.B. zum Hervorheben der Melodiestimme gegenüber dem Rest des Ensembles.

Metrische Akzentuierung: (auch Schwerpunktbetonung) Das der symbolischen Zeitrechnung zugrunde liegende BBU-Raster dient den Musikern

nicht nur als visuelle Orientierungshilfe im Notentext und zur Koordination von Rhythmus und Zusammenspiel. Es impliziert auch ein bestimmtes Betonungsschema. So wird in einem 4/4-Takt der erste Schlag am stärksten betont, die nächst schwächere Betonung kommt auf dem dritten Schlag gefolgt vom zweiten und vierten. Will der Komponist dies anders haben, verzeichnet er das durch Akzente im Notenbild. Manche Musikstile bringen auch ihre eigenen Betonungskonventionen mit, etwa der Reggae, der den 4/4-Takt quasi invers betont (Schütz 2008). Für den Hörer sind diese Betonungen einerseits Orientierungshilfe (z.B. beim Tanz) aber auch ein Merkmal, welches das Musikstück eher rhythmusbetont oder melodiös erscheinen lässt. Stark ausgeprägte metrische Akzente stehen der Gestaltung von Takt-übergreifenden Melodielinien eher entgegen. Es ist folglich am Interpretieren, zu entscheiden, worauf er den Schwerpunkt legt.

Artikulation: Durch Artikulation wird beschrieben, wie ein Ton geformt ist. Das schließt auch seine Lautstärke mit ein. So fordert bspw. der Akzent eine Erhöhung der Lautstärke. Artikulation bezieht aber auch eine Reihe weiterer Parameter mit ein, die nichts mit Dynamik zu tun haben. Deswegen ist es heute üblich und angemessen, Artikulation separat als einen eigenständigen Aspekt zu betrachten. Dem wird auch hier Folge geleistet. Abschnitt 5.5 widmet sich exklusiv der Artikulation. An dieser Stelle soll lediglich der Hinweis genügen, dass sie auch in die Mikro-Dynamik einfließt.

5.4.1 Makro-Dynamik

Formalisierung

Die Einzeichnung von Dynamik im Notentext ist in der heutigen Musikpraxis üblich, jedoch oft unvollständig. Crescendi und Decrescendi sind nur in seltenen Fällen mit ihren Anfangs- und Endlautstärken versehen. Den Spielern wird die Freiheit gelassen, selbst das richtige Maß zu bestimmen. Manche Einzeichnungen erscheinen redundant, etwa wenn mehrmals die gleiche Dynamik vorgeschrieben wird, ohne dass dazwischen anders lautende Anweisungen stehen. Hier wird damit gerechnet, dass die Spieler diese Dynamik nicht halten, bewusst oder unbewusst. Durch die scheinbar redundanten Anweisungen werden sie wieder auf die eigentliche Grundlautstärke hingewiesen, gewissermaßen in die Schranken gewiesen. Komponisten, Orchestratoren und Bearbeiter rechnen stets damit, dass Musiker mehr machen als das, was im Notentext steht. Nicht jedes Detail muss extra ausnotiert werden. Den Musikern können Freiheiten bei der Gestaltung gelassen werden. Darüber hinaus muss mit instrumentenspezifi-

schen Unterschieden umgegangen werden (Sevsay 2005). Das Forte einer Blockflöte ist naturgemäß leiser als das einer Trompete.

Ein formales Modell für die Beschreibung der Dynamik muss all dieses implizite Wissen einbeziehen. Eine formal vollständige Dynamikanweisung ist deshalb zwangsläufig ausführlicher als das, was der Musiker von seinem Notenblatt her kennt. Das konnte bereits beim Timing beobachtet werden. Analog zum Timing wird die Abfolge aller Dynamikanweisungen D als *Dynamicsmap* M_D bezeichnet. Sie kann global für alle Stimmen bzw. Kanäle gelten oder lokal für eine einzelne Stimme. Die Dynamicsmap enthält neben der Liste der Anweisungen weiterhin einen Header H , der das Dynamik-Vokabular definiert; dazu später mehr.

$$M_D = (H, (D_0, D_1, \dots, D_n))$$

Die Form von Dynamikanweisungen ähnelt der von Tempoanweisungen aus Abschnitt 5.3.1.

$$D_m = (d_m, v_{1,m}, v_{2,m}, \text{shape}_m) \quad \text{für } 0 \leq m \leq n$$

d_m	Tick-Datum der Anweisung
$v_{1,m}$	Lautstärke am Beginn der Anweisung
$v_{2,m}$	Lautstärke am Ende der Anweisung
shape_m	Charakteristik des Lautstärkeüberganges,

Der Definitionsbereich der Dynamikanweisung D_m ist das halboffene Intervall $[d_m, d_{m+1})$, wobei das Tick-Datum d_{m+1} das Datum der nächsten Anweisung in der Dynamicsmap ist, spätestens jedoch das Ende des Musikstücks. Innerhalb dieses Intervalls definiert die Anweisung einen kontinuierlichen monotonen Lautstärkeübergang von $v_{1,m}$ nach $v_{2,m}$. Für die technische Umsetzung müssen dies numerische Werte sein. Im Falle einer MIDI-Implementierung sind es die Velocity-Werte, die im MIDI-Standard Teil eines jeden NoteOn-Events sind. Sie sind ganzzahlig und reichen von 0 (stummgeschaltet) bis 127 (maximale Lautstärke).

In der Musikpraxis ist eine andere Konvention üblich. Hier wird Lautstärke durch umgangssprachliche Bezeichnungen, wie *piano*, *forte*, *fortissimo*, und deren Abkürzungen, etwa *pp*, *p*, *mf*, *f*, *ff* etc., angezeigt. Um beide Konventionen, die numerische und die umgangssprachliche, in der Implementierung zu unterstützen, kann im Header H der Dynamicsmap eine Lookup-Tabelle frei definiert werden. Hier wird jeder beliebigen Zeichenkette ein Velocity-Wert zugewiesen, z.B. *asLoudAsPossible* = 127. Ein typisches Beispiel für eine solche Lookup-Tabelle ist in Tabelle 5.3 dargestellt. Verwendet die Dynamikanweisung numerische Werte, werden sie direkt als Velocity-Werte interpretiert. Werden

String	<i>ppp</i>	<i>pp</i>	<i>p</i>	<i>mp</i>	<i>mf</i>	<i>f</i>	<i>ff</i>	<i>fff</i>
MIDI Velocity	12	36	48	64	83	97	111	125

Tabelle 5.3: Ein Beispiel für eine Lookup-Tabelle, die in der Musik gebräuchliche Zeichenketten mit MIDI-Velocity-Werten assoziiert. Sie wird im Header der Dynamicsmap definiert.

hingegen Strings verwendet, werden sie durch die assoziierten Velocity-Werte aus der Lookup-Tabelle ersetzt. Da für jeden Kanal eine eigene Dynamicsmap und damit auch ein eigener Header definiert werden kann, ist es nun problemlos möglich, einer Flöte ein leiseres Forte zu geben, als einer Trompete, den dynamischen Ambitus eines Instruments noch zusätzlich zu skalieren und zu verzerrern und auch Unterschiede in der Abbildung von Velocity-Werten, wie sie auf verschiedenen Samplern und Synthesizern beobachtet werden können, zu kompensieren.

Berechnung

Die Parameter $v_{1,m}$ und $v_{2,m}$ sind die Eckdaten des Lautstärkeübergangs, der sich von d_m bis an d_{m+1} heran erstreckt. Die konstante Dynamik kann dabei als Spezialfall angesehen werden, bei dem $v_{1,m}$ und $v_{2,m}$ gleich sind. Anderenfalls handelt es sich um ein echtes Crescendo ($v_{1,m} < v_{2,m}$) oder Decrescendo ($v_{1,m} > v_{2,m}$).

Diese haben bei menschlichen Musikern selten einen perfekt linearen Verlauf. Ein linearer Übergang wirkt mechanisch, neutral, ziellos, wohingegen bei den menschlichen Musikern eine große Bandbreite an unterschiedlichsten Verlaufsscharakteristiken beobachtet werden kann (Berndt & Hähnel 2010). Sie werden von ausdrucksmäßigen Absichten des Interpreten und vom musikalischen Kontext forciert. So war in den Performance-Übungen (Anhang D) in Abwesenheit jeglichen musikalischen Kontextes eine Tendenz zu gleichmäßigen fast linearen Verläufen zu beobachten, die sich bei echten Musikstücken nicht zeigte (siehe Abbildung 5.8). Das zentrale Gestaltungsmerkmal von Lautstärkeübergängen ist die Gleichmäßigkeit. Wann vollzieht sich der Großteil der Lautstärkeänderung? Am Beginn, zum Ende hin oder eher in der Mitte? Wie deutlich wird diese Charakteristik vom Musiker herausgearbeitet?

Dieses Verhalten kann durch sigmoidale Verläufe beschrieben werden. Durch die Stärke der S-Krümmung kann der Lautstärkeübergang unterschiedlich deutlich markiert werden. Die Symmetrie des S-Verlaufs kann entlang der x -Achse verzerrt werden, so dass auch ungleichmäßige Übergänge beschrieben werden.

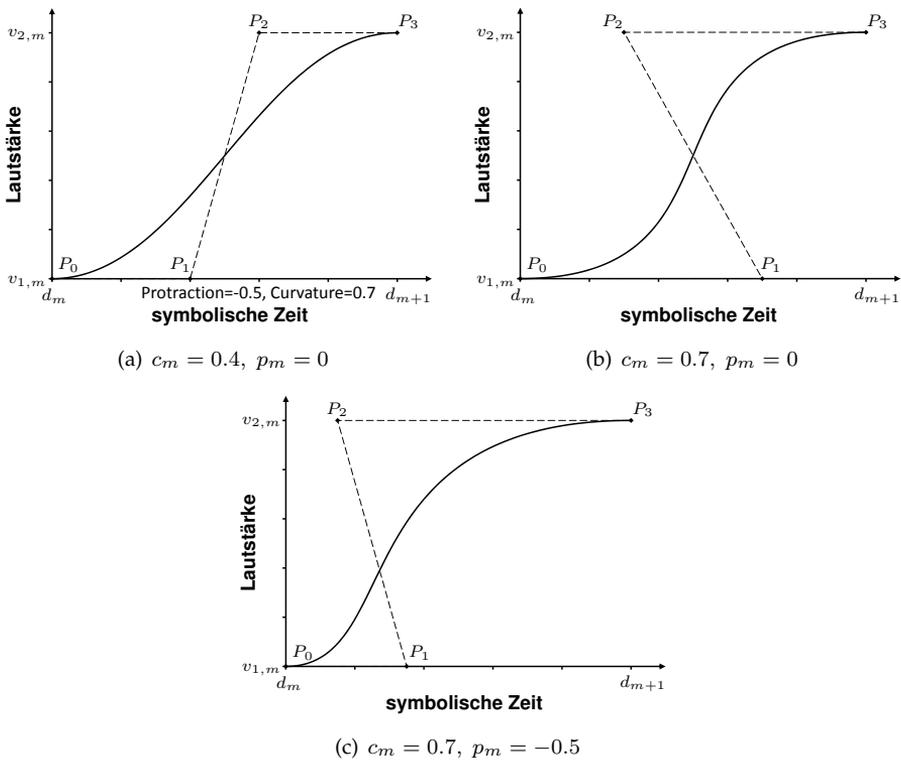


Abbildung 5.7: (De-)Crescendi werden mit kubischen Bézierkurven modelliert. Durch die Punkte P_2 und P_3 wird die Charakteristik des Lautstärkeüberganges bestimmt. Die Parameter hierfür sind *Curvature* c_m und *Protraction* p_m .

Dazu wird eine kubische Bézierkurve verwendet (siehe Abbildung 5.7). Die Kurve wird im Einheitsquadrat definiert und auf die tatsächlichen zeitlichen und dynamischen Abmessungen der Lautstärkeanweisung skaliert. Damit kann die Übergangscharakteristik stets im einheitlichen Bezugssystem des Einheitsquadrates beschrieben werden und ist skalierungsunabhängig. Das Kontrollpolygon der Kurve wird durch die Punkte P_0, P_1, P_2 und P_3 aufgespannt. Anfangs- und Endpunkt sind an die Koordinaten

$$P_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{und} \quad P_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

gebunden; sie markieren Anfang und Ende des Übergangs. Für die inneren Punkte stehen die y -Koordinaten ebenfalls fest:

$$P_1 = \begin{pmatrix} x_1 \\ 0 \end{pmatrix} \quad \text{und} \quad P_2 = \begin{pmatrix} x_2 \\ 1 \end{pmatrix}$$

wohingegen x_1 und x_2 variabel im Intervall $[0, 1]$ sind. Durch sie kann sowohl die Stärke der S-Krümmung als auch ihre Symmetrie gesteuert werden. Diese Bedingungen vereinfachen die Darstellung der Bézierkurve auf:

$$\begin{aligned} x(t) &= (3x_1 - 3x_2 + 1)t^3 + (-6x_1 + 3x_2)t^2 + 3x_1t \\ y(t) &= -2t^3 + 3t^2 \end{aligned}$$

mit $0 \leq t \leq 1$. Für ein Datum $d \in [d_m, d_{m+1})$ und den assoziierten Parameter t_d mit

$$\frac{d - d_m}{d_{m+1} - d_m} = x(t_d)$$

wird der Lautstärkewert $MacroDyn(d)$ durch

$$MacroDyn(d) = y(t_d) \cdot (v_{2,m} - v_{1,m}) + v_{1,m}$$

berechnet. Die Verlaufscharakteristik ist in der Dynamikanweisung im Term $shape_m$ codiert. Hier werden x_1 und x_2 aber nicht explizit angegeben. Statt dessen werden zwei deskriptivere Parameter eingeführt. Die Stärke der S-Krümmung wird durch den Parameter *Curvature* $c_m \in [0, 1]$ gesteuert. Je größer dieser ist, desto deutlicher ist das Crescendo bzw. Decrescendo. Es konzentriert sich dann aber auch auf einen engeren zeitlichen Bereich. Die Symmetrie der Kurve wird durch *Protraction* $p_m \in [-1, 1]$ verzerrt. Ein dritter Parameter, res_m , wird später im Zusammenhang mit Subnotendynamik besprochen.

$$shape_m = (c_m, p_m, res_m)$$

Aus *Curvature* und *Protraction* werden die x -Koordinaten der beiden variablen Kontrollpunkte hergeleitet. Sei $p_m = 0$, dann berechnen sich die Koordinaten wie folgt.

$$x_1 = c_m \quad \text{und} \quad x_2 = 1 - c_m$$

Die Punkte werden entlang der x -Achse gegeneinander verschoben; P_1 von x_0 ausgehend in Richtung x_3 (von vorn nach hinten), P_2 von x_3 ausgehend in Richtung x_0 (von hinten nach vorn). Die S-Krümmung nimmt dabei zu; die Lautstärkeänderung wird markanter. Der lineare Verlauf wird erreicht, wenn $c_m = p_m = 0$ gesetzt werden.

Protraction führt eine weitere Verzerrung ein. Dieser Parameter beschreibt, ob der Lautstärkeübergang eher vorzeitig ($-1 \leq p_m < 0$), relativ spät ($0 < p_m \leq 1$) oder gleichmäßig ausgewogen ($p_m = 0$) stattfindet. Bei diesen Verschiebungen muss sichergestellt werden, dass die inneren Kontrollpunkte (also P_1 und P_2) nicht jenseits der Grenzen x_0 und x_3 verschoben werden, damit Mehrdeutigkeiten im Lautstärkeverlauf vermieden werden. Die Verschiebungen werden deshalb auf den durch *Curvature* eingeschränkten verfügbaren Bereich skaliert. Für $p_m \neq 0$ werden die Koordinaten folgendermaßen berechnet.

$$x_1 = c_m + \frac{p_m}{2} + |p_m| \left(\frac{1}{2} - c_m \right)$$

$$x_2 = 1 - c_m + \frac{p_m}{2} + |p_m| \left(c_m - \frac{1}{2} \right)$$

Subnotendynamik

Der Lautstärkeverlauf kann in die Velocity-Attribute aller betreffenden NoteOn-Events gerendert werden. Das entspricht im Prinzip einer notenweisen Terrassendynamik – jede Note hat ihren Lautstärkewert. Für Instrumente mit perkussiver Klangerzeugung (Klavier, Harfe, Gitarre etc.) genügt das, ebenso für rasche Tonfolgen. Die meisten Blas- und Streichinstrumente wie auch die menschliche Stimme sind darüber hinaus aber in der Lage, die Lautstärke auch während des klingenden Tones zu formen. Dieses Phänomen sei als *Subnotendynamik* bezeichnet. Das kann mit der notenweisen Terrassendynamik allein nicht erreicht werden.

Im MIDI-Format kann das durch den channelVolume-Controller (Controller Nr. 7) und den Expression-Controller (Controller Nr. 11) nachgeahmt werden. Allerdings sind diese oft nur als reine Blenden implementiert, so dass sich zwar die Amplitude des Ausgabesignals ändert, nicht aber die Klangcharakteristik. Moderne Sampler und Synthesizer bieten deshalb alternative Controller-Schnittstellen an, die auch klangfarbliche Änderungen umsetzen. Im Falle des Samplers *Vienna Instruments* ist dies der sog. *Velocity-Cross-Fade-Controller* (Vienna Symphonic Library GmbH 2010a).

Nichtsdestotrotz muss der kontinuierliche Lautstärkeübergang in eine Sequenz von Controller-Events diskretisiert werden. Die Auflösung dieser Diskretisierung wird in shape_m durch den Parameter $\text{res}_m \in (0, 1]$ angegeben. Er definiert die Schrittweite für t bei der Konstruktion der Bézierkurve. Ein Wert von $\text{res}_m = 0.1$ bedeutet, dass die Kurve durch 10 Controller-Events (10 Schritte) approximiert wird, $\text{res}_m = 0.01$ ergibt 100 usw. Eine höhere Dichte von Controller-

Events produziert glattere Ergebnisse, bedeutet aber auch mehr MIDI-Traffic. Eine hohe Auflösung kann weiterhin dazu führen, dass aus numerischen Gründen mehrere Events auf das gleiche Tick-Datum fallen oder den gleichen Lautstärkewert setzen, denn beide sind in MIDI ganzzahlig. Diese unnötige Redundanz wird vermieden, indem die Implementierung aufeinander folgende Controller-Events mit dem gleichen Datum oder Wert automatisch filtert bzw. gar nicht erst erzeugt.

Diskussion



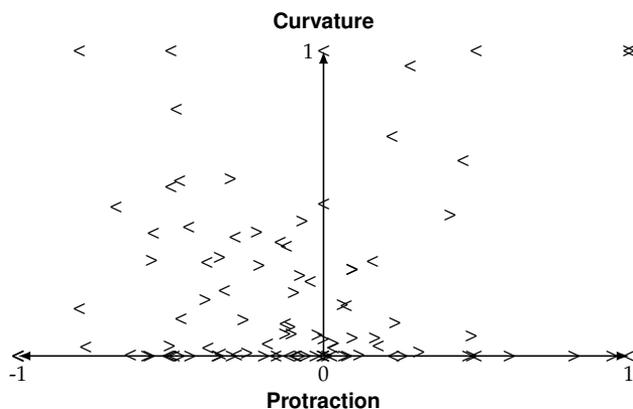
CD-Track
13

Auf der beiliegenden Audio-CD ist ein Musikbeispiel enthalten, in welchem die Dynamik anhand eines Blechbläserstücks demonstriert wird.⁸ Die Subnotendynamik ist hauptsächlich dort auffallend, wo Lautstärkeänderungen auf langen Tönen stattfinden. Tatsächlich sind aber alle Crescendi und Decrescendi mit Subnotendynamik gerendert.

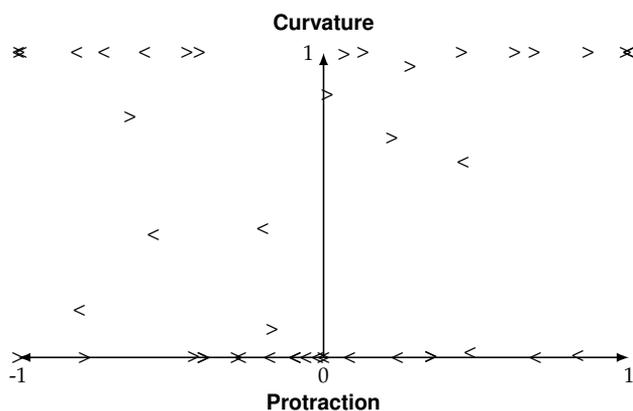
Das Makro-Dynamik-Modell wurde zur Approximation der gemessenen Lautstärkeverläufe bei menschlichen Musikern verwendet, siehe Abschnitt 5.1. Dabei zeigte sich eine große Bandbreite an unterschiedlichen Verläufen. Die aufgetretenen *Curvature-Protraction*-Konstellationen wurden in Abbildung 5.8 geplottet. Hier ist zu erkennen, dass das Modell einen größeren Raum von möglichen Verläufen aufspannt als in der Praxis auftreten. Die *eine* typische Charakteristik gibt es aber nicht; es konnte bei allen Interpreten eine hohe Varianz beobachtet werden.

Es gibt jedoch noch eine Charakteristik, die durch das Modell nicht unterstützt wird, die besonders in zeitlich langen Crescendi und Decrescendi vorkommt, die »inverse S-Charakteristik«. Hierfür müssten $y_1 = 1$ und $y_2 = 0$ gesetzt werden, was im Modell nicht vorgesehen ist. Die geometrische Form des Kontrollpolygons wäre dann nicht Σ , sondern \mathcal{N} . Das bedeutet, die Spieler haben den Dynamikübergang nicht in einem Gang ausgeführt, sondern zunächst nur bis auf ein zwischengelagertes Niveau. Von dort aus wurde dann in die Zieldynamik übergeleitet. Der Grund für dieses besondere Verhalten ist wahrscheinlich darin zu suchen, dass zeitlich lange Lautstärkeverläufe vom Spieler eine feinere Abstufung und Einteilung erfordern, was motorisch nicht immer leicht umzusetzen ist. Also beginnen die Spieler mit dem Dynamikübergang, führen ihn aber nicht vollständig aus, weil sie die Zieldynamik dann zu früh erreichen, sondern sparen sich den Rest bis zum Schluss auf. Tatsächlich lässt sich dieser Verlauf durch zwei Dynamikanweisungen angemessener approximieren als mit nur einer (siehe Abbildung 5.9), was auch eher dem entspricht, was die Musiker hier tatsächlich machen.

⁸Es ist eine Aufnahme des Choralvorspiels zu *Nun sich der Tag geendet hat* (Berndt 2012).



(a) Performance-Übungen (siehe Anhang D)



(b) Live- und Studio-Aufnahmen

Abbildung 5.8: Die aufgetretenen *Curvature-Protraction*-Kombinationen für Crescendi (<) und Decrescendi (>) bei den Messungen (Studie durchgeführt von Tilo Hähnel, siehe Abschnitt 5.1 und Berndt & Hähnel 2010).

Ein gelegentlich geäußelter Kritikpunkt am MIDI-Standard ist die relativ grobe Auflösung des dynamischen Ambitus – nur 128 Stufen (7 Bit). In einem langen gleichmäßigen Crescendo oder Decrescendo können diese Stufen hörbar werden und erzeugen einen eher ungleichmäßigen ruckartigen Höreindruck. Besonders bei stabilen stehenden Klängen können diese Aliasing-Artefakte auffallen. Der MIDI-Standard bietet für die Controller die Möglichkeit einer deutlich höheren Auflösung. Durch Koppelung zweier Controller ergeben sich 16 384

Abstufungen (14 Bit). Das wäre mehr als genug. Trotzdem ist in der Praxis nach wie vor auch die geringere 7-Bit-Auflösung gebräuchlich. Dafür gibt es mehrere Gründe. Die Abstufungen werden erst in genügend langen Verläufen wahrnehmbar. Bei kurzen Crescendi und Decrescendi fallen sie kaum oder gar nicht auf. Je mehr Eigenleben der Klang hat, desto weniger hörbar sind die Artefakte, denn die klangimmanente Amplitudenmodulation macht die Stufe unscharf, ist vielleicht sogar größer als die Stufe. Zudem steht die Musik im Zuge eines Dynamikübergangs selten still (kein stehender Klang). Mit jedem neuen Ton, jedem neuen Onset, wird eine neue Hüllkurve auf das Audiosignal modelliert, die deutlich größere Amplitudenschwankungen mit sich bringt als die Dynamikstufe. Der Hörer kann nur noch die Attack-Pegel vergleichen; für ihn wird das Bezugssystem also diskret. Aus diesem Grund ist bei raschen Tonfolgen keine Subnotendynamik nötig. Jenseits des MIDI-Standards werden aber sehr wohl auch höhere Lautstärkeauflösungen angeboten, etwa vom Software-Synthesizer *cSound* (Boullanger 2000). Das Makro-Dynamik-Modell ist dank seiner Skalierungsunabhängigkeit problemlos auch darauf übertragbar.

5.4.2 Mikro-Dynamik

Die Mikro-Dynamik beschreibt systematische Abweichungen von der Grundlautstärke. Es sind Veränderungen relativ zur Grundlautstärke, wie sie auch im menschlichen Spiel beobachtet werden können. Dazu zählen die Betonung metrischer Schwerpunkte und die Artikulation. Da letztere allerdings mehr als nur ein Dynamik-Merkmal ist, wird sich Abschnitt 5.5 gesondert und detailliert damit beschäftigen. Dieser Abschnitt konzentriert sich auf die Schwerpunktbetonung, also die lautstärkemäßige Akzentuierung metrischer Schwerpunkte.

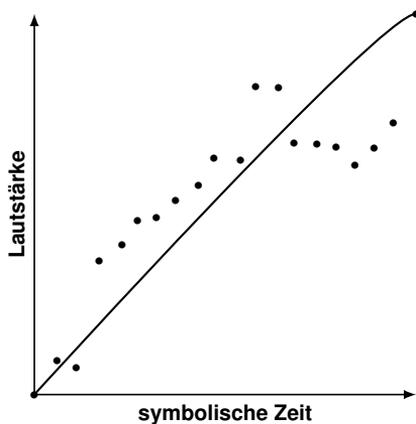
Metrische Akzentuierung

Die lautstärkemäßige Hervorhebung einzelner Taktzeiten ist *das* Mittel, mit dem die der Musik zugrunde liegende metrische Ordnung hörbar gemacht wird. Es folgt ein anschauliches Beispiel. Welches Taktschema liegt dem folgenden Punktmuster zugrunde, 3er- oder 4er-Takt?

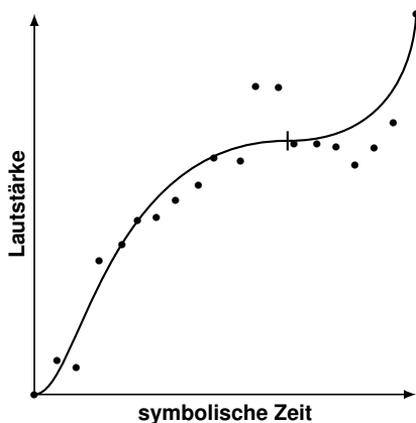
.....

Wenn alle Punkte gleich sind, fehlen Anhaltspunkte, um eine eindeutige Antwort geben zu können. So ist es auch in der Musik. Im schlimmsten Fall verliert der Hörer die Orientierung, vermag nicht, das Gehörte zu größeren Sinneinheiten zu segmentieren. Ganz anders sieht es in den folgenden Fällen aus.

● . . . ● . . . ● . . . ● . . .



(a) Approximation mit einer Dynamikanweisung D_m : $c_m = 0$, $p_m = -0.094$, $r_m = 0.88$.



(b) Approximation mit zwei Dynamikanweisungen D_m : $c_m = 0.404$, $p_m = -0.557$, $r_m = 0.96$ und D_{m+1} : $c_{m+1} = 1$, $p_{m+1} = 1$, $r_{m+1} = 0.94$.

Abbildung 5.9: Lange Lautstärkeübergänge (hier ein Crescendo) lassen sich besser in zwei Phasen approximieren. Die Messdaten sind als Punkte dargestellt; mit r wird der Korrelationskoeffizient jeder Dynamikanweisung angezeigt (Messungen und Approximationen von Tilo Hähnel, siehe auch Berndt & Hähnel 2010).

Ein 3er-Takt.



Ein 4er-Takt. Die Hervorhebung einzelner Punkte macht es möglich, Muster zu erkennen, und sich darin zu orientieren. Ohne dies wäre Tanzmusik nicht denkbar. Selbst Taktwechsel sind nun leicht nachzuvollziehen.



Zwei 4er-Takte gefolgt von drei 3er-Takten. Erst jetzt kann die metrische Ordnung sogar als kompositorischen Gestaltungsebene dienen. Berühmt ist bspw. die *Hemiolle*, die Vergrößerung des 3er-Taktes. Sie ist in der Renaissance- und Barockmusik sehr beliebt gewesen. Hierbei werden zwei normale 3er-Takte zu einem großen zusammenschlossen. Dabei verschieben sich die Betonungen.



In diesem Beispiel wird auch anschaulich, dass Schwerpunktbetonung nicht binär ist (betont und unbetont), sondern mehrere hierarchische Ebenen der metrischen Ordnung ausdifferenzieren kann. Ansätze zur Konstruktion dieser hierarchischen Struktur beschreiben u.a. Lehrdahl & Jackendorf (1983), Miller, Scarborough & Jones (1992). Aus diesen können explizite Betonungsmuster (oder Betonungsschemata) abgeleitet werden. Wenn immer sich Gepflogenheiten und Regeln etablieren, gibt es natürlich auch diejenigen Komponisten und Interpreten, die zeigen, dass es noch mehr gibt, die Regeln brechen und dadurch überraschen. Feste Betonungsmuster können dem nicht gerecht werden, sind also eher kurzfristige Lösungen. Die Schwerpunktbetonung ist sehr flexibel und variiert sogar schon innerhalb ein und des selben Musikstückes, was am Beispiel zweier barocker Menuette gezeigt wurde (Berndt & Hähnel 2010).

Es liegt also nahe, auch die Schwerpunktbetonung in einer zeitlich geordneten Liste zu organisieren, der *Emphasismap* M_E .

$$M_E = (E_0, E_1, \dots, E_n)$$

Eine Anweisung E_m hat die folgende Form.

$$E_m = (d_m, scale_m, (A_m^0, A_m^1, \dots, A_m^k)) \quad \text{für } 0 \leq m \leq n$$

d_m	Tick-Datum der Anweisung
$scale_m$	Skalierung des Betonungsschemas
$(A_m^0, A_m^1, \dots, A_m^k)$	Betonungsschema mit den einzelnen Akzentuierungen

Die Anweisung definiert ein eintaktiges Betonungsschema, das ab dem Tick-Datum d_m auf jeden Takt angewendet wird, bis zum Datum der Folgeanweisung oder dem Ende des Musikstücks. Die Länge eines Taktes ist durch die Taktartdefinition bekannt. Im MIDI-Standard gibt es dafür den TimeSignature-Event. Ein 3/4-Takt enthält drei Schläge; ein Schlag hat dabei die Länge einer Viertelnote. Im 6/8-Takt enthält der Takt sechs Schläge von der Länge einer Achtel. Die Taktart definiert also die Anzahl der Schläge in einem Takt (Zähler) und den Notenwert eines Schlages (Nenner). Die Konvertierung des BBU-Formats in das MIDI-Tick-Format wurde in Abschnitt 5.3 bereits beschrieben. Es kann also für jede Tick-Position d angegeben werden, in welchem Takt ($\text{bar}(d) \in \mathbb{N}$) sie sich befindet und auf welcher Schlagposition innerhalb des Taktes ($\text{beat}(d) \in \mathbb{R}$).

Das Betonungsschema ist eine Sequenz von Akzentuierungen A , durch welche die Lautstärkekontur für einen Takt beschrieben wird, genauer: die Kontur der Abweichungen von der Grundlautstärke. Eine Akzentuierung beschreibt ein Segment dieser Kontur. Es hat folgende Form.

$$A_m^j = (\text{beat}_{j,m}, a_{1,j,m}, a_{2,j,m}, a_{3,j,m}) \quad \text{für } 0 \leq j \leq k$$

mit $\text{beat}_{j,m} \geq 1$
 und $a_{i,j,m} \in [-1, 1]$, $i \in \{1, 2, 3\}$

Das Segment beginnt auf der Schlagposition $\text{beat}_{j,m}$ und wird durch $\text{beat}_{j+1,m}$ terminiert. Der erste Schlag des Taktes ist $\text{beat}_{j,m} = 1$; der zweite Taktschlag $\text{beat}_{j,m} = 2$ usw. Auch Positionen zwischen Schlägen sind möglich; $\text{beat}_{j,m} = 2.5$ entspricht dem Schlag »2-und«. Es ist klar, dass Segmente, die jenseits der Taktgrenzen liegen, formal zwar definiert werden können, auf das klingende Resultat aber keinen Einfluss haben werden. Ein Segment auf $\text{beat}_{j,m} = 5$ findet in einem 2er-, 3er oder 4er-Takt keine Berücksichtigung, erst ab einem 5er-Takt. Auch Beats kleiner als 1 mögen formal möglich sein, bleiben bei der Lautstärkeberechnung aber unberücksichtigt.

Die Akzentuierung der Taktposition $\text{beat}_{j,m}$ wird durch $a_{1,j,m}$ angegeben, wobei der Wert 1 der maximalen Akzentuierung entspricht; 0 bewirkt keine Abweichung von der Grundlautstärke; -1 bewirkt die maximale Zurücknahme. Dieser Wert, wie auch die der anderen a -Parameter, wird durch scale_m auf den tatsächlichen Lautstärkewert skaliert, z.B. auf einen MIDI-Velocity-Wert. Dadurch kann mit einem einzigen Parameter geregelt werden, wie deutlich das Betonungsschema umgesetzt wird.

Doch was passiert zwischen den Akzentuierungen? In den praktischen Versuchen mit Musikern konnte verschiedenes beobachtet werden. Mal wurde versucht, eine konstante unbetonte Lautstärke zu halten. Mal wurde auf die nächste Betonung »hingespelt«, besonders in Auftakten, was durch kurze Crescendi

umgesetzt wurde. Stets handelte es sich um monotone Charakteristiken (steigend, fallend oder konstant). Die Parameter $a_{2,j,m}$ und $a_{3,j,m}$ dienen dazu, das nachzubilden. Den Anfangswert des Übergangs setzt $a_{2,j,m}$, das Ende $a_{3,j,m}$.

$a_{2,j,m} = a_{3,j,m}$ für ein konstantes Niveau,

$a_{2,j,m} > a_{3,j,m}$ für ein Mikro-Decrescendo,

$a_{2,j,m} < a_{3,j,m}$ für ein Mikro-Crescendo (typisch für Auftakte).

Dadurch, dass die Schwerpunktbetonung auf die Grundlautstärke aufgerechnet wird, kann es passieren, dass die sich ergebende Gesamtlautstärke über die technischen Grenzen hinausgeht. Bei einer Grundlautstärke von 110 erlaubt die 7-Bit-Auflösung der MIDI-Velocity nur noch eine maximale Betonung von 17. In der Praxis sind deutlich größere Betonungen nicht selten. Gleiches kann auch in negativer Richtung geschehen: Bei sehr leiser Grundlautstärke ist der Spielraum für die maximale Zurücknahme eingeschränkt. Ein arithmetischer Überlauf und Unterlauf würde zu grundsätzlich falschen Ergebnissen führen. Eine naive Lösung wäre, zu große oder kleine Werte an der jeweiligen Grenze zu clippen. Wenn mehrere, eigentlich verschiedene, Betonungen die Grenze überschreiten, erhalten sie dabei aber alle die gleiche Betonung. Die Struktur des Betonungsschemas geht verloren.

Es ist besser, die Skalierung des Betonungsschemas $scale_m$ zu dämpfen. Wenn ein Überlauf droht ($MacroDyn(d) + scale_m > 127$), werden alle positiven Akzentuierungen nur noch auf den tatsächlich verbleibenden Spielraum skaliert ($127 - MacroDyn(d)$). Droht ein Unterlauf, also Werte kleiner 0, ($MacroDyn(d) < scale_m$), werden alle negativen Akzentuierungen herunterskaliert. Die vollständige Implementierung der Schwerpunktbetonung $EmphDyn$ sieht folgendermaßen aus.

$$EmphDyn(d) = MacroDyn(d) + accent(d) \cdot scale(d)$$

$$accent(d) = \begin{cases} a_{1,j,m} : beat(d) = beat_{j,m} \\ \frac{(beat(d) - beat_{j,m})(a_{3,j,m} - a_{2,j,m})}{beat_{j+1,m} - beat_{j,m}} + a_{2,j,m} : \text{sonst} \end{cases}$$

$$scale(d) = \begin{cases} MacroDyn(d) : MacroDyn(d) + accent(d) \cdot scale_m < 0 \\ 127 - MacroDyn(d) : MacroDyn(d) + accent(d) \cdot scale_m > 127 \\ scale_m : \text{sonst} \end{cases}$$

wobei A_m^j die letzte Akzentuierung vor d ist und $beat_{j+1,m}$ die Taktposition der nächsten auf d folgenden Akzentuierung, spätestens aber das Taktende, so dass folgende Bedingung erfüllt ist.

$$beat_{j,m} \leq beat(d) < beat_{j+1,m}$$

Durch die Dämpfung der Skalierung bleibt das Betonungsschema in sich ausdifferenziert. Zudem ahmt es einen Effekt nach, der auch bei menschlichen Musikern beobachtet werden kann. Bei extremer Dynamik, also an den Grenzen des dynamischen Ambitus, sind Akzentuierungen weniger deutlich. Auch der Mensch skaliert hier herunter.

Diskussion

Die Akzentuierung metrischer Schwerpunkte ist ein außerordentlich schwer zu untersuchender Merkmalskomplex. In Musikaufnahmen wird er von Makro-Dynamik und Artikulation überlagert und durch akustische Effekte, wie Nachhall, verfälscht. Das hier beschriebene Modell entstand deshalb nicht auf Grundlage von Messungen in realen Musikaufnahmen, aus denen sich die Schwerpunktbetonung nur schwer extrahieren lässt. Die Grundlage für das Modell-Design waren statt dessen Gespräche und praktische Versuche mit Musikern. Das Modell bildet nach, was sie in ihrem Spiel bewusst gestalten und im Falle von Profimusikern auch lehren. Das Modell-Design ist also musikalisch plausibel, aber empirische Studien, die real auftretende Betonungscharakteristiken gegenüberstellen könnten, fehlen.

In den durchgeführten Messungen konnten Schwerpunktbetonungen mit einem dynamischen Radius ($scale_m$) von durchschnittlich 2.7 dB beobachtet werden. In einigen Interpretationen traten allerdings Standardabweichungen auf den einzelnen Betonungen von rund 200% des dynamischen Radius auf. Noch ist unklar, ob sich die beobachteten Abweichungen vollständig aus systematischen Ursachen ableiten lassen, so dass die Mikro-Dynamik evtl. auch durch eine Zufallskomponente ergänzt werden müsste, wie im Falle des Mikro-Timings (rhythmische Ungenauigkeit). Eine systematische Erklärung dafür, dass die Betonungen eben nicht in jedem Takt gleich sind, kann auch durch mehrtaktige Betonungsmuster gegeben werden. Diese können sich etwa an der Strukturgliederung der Musik orientieren, lassen sich vielleicht auch aus melodischen und harmonischen Zusammenhängen ableiten, wie im Falle der Hemirole.

Vielleicht ließen sich daraus auch die Verläufe zwischen den Akzentuierungen herleiten. In der vorliegenden Umsetzung sind diese konstant oder linear. Wären gekrümmte Verläufe angemessener? Die Monotonie der Übergänge steht außer Frage. Gäbe es in den Übergängen lokale Maxima oder Minima, wären sie

als Akzentuierung zu werten, nicht als Übergang zwischen Akzentuierungen. Wo komplexere Verläufe gewünscht sind, lassen sie sich aus monotonen Segmenten zusammensetzen. Bedenkt man die geringe, meist einstellige, Anzahl an Tönen, die zwischen zwei akzentuierten Zählzeiten liegen, und den geringen dynamischen Ambitus, den die Übergänge zu überwinden haben, wird jedoch klar, dass gekrümmte Verläufe einen kaum hörbaren Unterschied gegenüber linearen Approximationen ausmachen. Durch empirische Studien konnte bislang weder die eine noch die andere Verlaufscharakteristik als die angemessenere belegt werden, da die Streuung in den Messergebnissen keine solch mikroskopisch genauen Schlüsse erlaubt.

5.5 Artikulation

Die Artikulation beschäftigt sich mit der Formung der Töne. Das beinhaltet nicht nur ihre klangfarblichen Eigenschaften, sondern auch ihr Verhältnis zu den umgebenden Tönen im Musikstück. Ein Ton kann an seinen Folgeton angebunden oder deutlich von ihm getrennt sein. Ein Ton kann im Verhältnis zu anderen zeitgleich erklingenden Tönen und dem zugrunde liegenden Stimmungssystem sauber oder unsauber intoniert sein. Er kann sich der zugrunde liegenden Dynamik unterordnen oder lautstärkemäßig herausstechen.

Im Notentext wird Artikulation nicht für jeden Ton explizit notiert, tatsächlich meist nur für die wenigsten. Das heißt aber nicht, dass alle übrigen Töne »unartikulierte« zu spielen sind, was nicht möglich ist. Jede Tonformung ist eine Artikulation, auch die neutralste. Es wird vielmehr vorausgesetzt, dass die Musiker genug Hör- und Musiziererfahrung mitbringen und selbst eine geeignete Grundartikulation ausführen. Die wird durch die ausdrucksmäßigen Absichten des Musikstücks forciert (gesanglich, tänzerisch, hymnisch, marschartig, getragen, lebhaft usw.); sie kann also nicht problemlos verallgemeinert werden und kann sogar innerhalb eines Musikstücks wechseln. Im Notenbild wird Artikulation erst dann explizit verzeichnet, wenn sie von der erwarteten Grundartikulation abweichen soll.

Die eigentlichen Artikulationsanweisungen sind meist nur anschauliche Umschreibungen (wie etwa gebunden, abgesetzt, kurz, akzentuiert), die von den ausführenden Musikern interpretiert werden müssen. Bei der Umsetzung werden die folgenden fünf Toneigenschaften beeinflusst – mal mehr, mal weniger bewusst.

Hüllkurve: Die Formung des Tones im wortwörtlichen Sinne – die Hüllkurve beschreibt die Amplitudenmodulation des Tones. Sie beginnt mit der Ein-

schwingphase (Attack), geht über die Abfall-Phase (Decay), die in die Haltephase (Sustain) übergeht. In der Release-Phase schwingt der Ton schließlich aus. Die Übergänge in den einzelnen Phasen werden von Synthesizern und Samplern unterschiedlich implementiert, mal durch lineare Funktionen, mal durch gekrümmte Verläufe. Die Länge der Sustain-Phase wird durch die Tondauer geregelt. Erst wenn der Ton beendet wird, geht die Hüllkurve in die Release-Phase über. Aber auch alle übrigen Phasen haben variable Längen. Eine lange Einschwingzeit lässt den Ton sanft erscheinen, macht ihn aber auch schwerfällig. Rasche melodische und rhythmische Figuren erfordern kurze Einschwingzeiten, d.h. perkussivere Tonanfänge. Das Überschwingen am Ende der Einschwingphase wird durch die Decay-Phase nachgebildet. Bei einem Fortepiano würde der Ton bspw. sehr stark überschwingen, das Amplituden-Peak am Ende der Attack-Phase wäre sehr laut und der Sustain-Pegel, in den die Decay-Phase relativ rasch überleitet, deutlich leiser. Die Release-Phase bestimmt, wie lange der Ton ausschwingt. Bei einem sehr dichten gebundenen (legato) Spiel können sich Release- und Attack-Phase überlagern, es entsteht eine Tonverbindung (zumindest bei monophonen Instrumenten).

Detaillierte Hüllkurvenuntersuchungen werden u.a. von Dannenberg, Pellerin & Derenyi (1998), Maestre & Gómez (2005) beschrieben. In den meisten Samplern und Synthesizern können Hüllkurven-Einstellungen über speziell zuzuweisende MIDI-Controller angesteuert werden. Durch viele Sample-Bibliotheken, wie auch durch die *Vienna Symphonic Library*, werden bereits eine Reihe von vorgefertigten Artikulationen samt den entsprechenden Hüllkurvencharakteristiken vorgegeben, die nur noch angesteuert werden müssen. Weitere Einstellungen sind dann kaum mehr nötig und werden deshalb auch hier nicht weiter thematisiert.

Tondauer: Ein Großteil der Artikulationsanweisungen bezieht sich auf die Tondauer. Sie wird auch in vielen Studien und Performance-Systemen als das zentrale Artikulationsmerkmal angesehen, häufig sogar ausschließlich betrachtet (Bresin 2001, Mazzola et al. 2002, Jerkert 2003). Die Tondauer ist die Zeit vom Beginn der Attack-Phase bis zum Ende der Sustain-Phase, also die Zeit, die der Musiker den Ton aktiv spielt (die Taste/Saite angeschlagen lässt, den Bogen bewegt, die schwingende Luftsäule aufrecht erhält). Diese Zeit entspricht selten der im Notenbild notierten Tonlänge. Durch Verkürzung der Tondauer entsteht eine Pause zwischen den Tönen. Der Ton wird von seinem Folgeton deutlich abgesetzt und wirkt eher als ein rhythmischer Impuls denn als Teil einer melodischen Linie. Umgekehrt kann ein Ton auch voll ausgespielt werden und dicht an den Folgeton heranreichen, angebunden werden, wodurch das Spiel eher kantabel (gesanglich, melodios) wird. Die Tondauer prägt also zum einen die

melodisch-rhythmische Qualität eines Tones. Zum anderen dient sie dazu, musikalische Struktur Grenzen zu markieren und das Ende einer Figur oder Phrase vom Beginn der nächsten abzusetzen (Keller 1973). Darüber hinaus ist sie ein gestalterisches Mittel, das der Darstellung ausdrucksmäßiger Kontraste dient.

Tonlautstärke: Artikulation hat auch Auswirkungen auf die Dynamik (siehe Abschnitt 5.4.2 und Anhang D). Viele Artikulationsanweisungen beeinflussen die Lautstärke des zu artikulierenden Tones. Naheliegende Beispiele sind der Akzent und das Sforzato. Beide fordern eine Erhöhung der Lautstärke. Selbst solche Artikulationen, die primär die Tondauer betreffen, werden auch dynamisch differenziert. Das Staccatissimo, das in den Lehrwerken als noch kürzeres Staccato definiert wird (z.B. Michels 2001), war in den Messungen (siehe Abschnitt 5.1) nicht signifikant kürzer, aber lauter. Es wird tatsächlich eher wie ein akzentuiertes Staccato ausgeführt. Bei den Artikulationsübungen konnten Staccato-Dauern von 160 bis 225 Millisekunden beobachtet werden, und Staccatissimo-Dauern von 143 bis 189 Millisekunden. Demgegenüber steht ein signifikanter Lautstärkeunterschied von 4.37 dB. Staccato wurde im Vergleich zur neutralen Artikulation durchschnittlich 2.14 dB leiser gespielt, Staccatissimo um 2.23 dB lauter (Hähnel & Berndt 2010a).

Intonation: Einige Spiel- und Artikulationstechniken beeinflussen auch die Intonation des Gespielten. Ein erhöhter Anblasdruck kann bei Holz- und Blechblasinstrumenten dazu führen, dass die Töne zu hoch intoniert sind. Vergleichbares existiert auch bei Streichinstrumenten im Bezug auf den Druck, der durch den Bogen auf die Saite und den Druck des Fingers am Griffbrett ausgeübt wird. Die Spieler müssen diese Schwankungen manuell ausgleichen. Auch an den Grenzen des melodischen Ambitus eines Instrumentes kann die Intonation aus spieltechnischen Gründen unsauber werden. Jenseits dieser eher unerwünschten Nebeneffekte wird die Intonation aber auch bewusst beeinflusst. Solches leitet sich meist aus dem melodischen und harmonischen Kontext ab. Musikschüler lernen bspw., dass Dur-Terzen tief, Leitöne aber hoch zu intonieren sind. Manche Komponisten benutzen unsaubere Stimmung als klangfarblichen Effekt, etwa Howard Shore (2001, 2002, 2003) im Mordor-Thema seiner *Lord of the Rings*-Soundtracks. Gegenstand wissenschaftlicher Analysen ist die Intonation u.a. bei Fyk (1995) und Maestre, Bonada & Mayor (2006). Mazzola et al. (2002) formalisieren Intonation als ein *Performance Field* in dem System *RUBATO*. Einige Sample-Bibliotheken beinhalten auch bereits intonatorische Schwankungen in den Sample-Daten. In dieser Arbeit soll nicht weiter darauf eingegangen werden. Ein umfassendes Intonationsmodell bleibt zukünftigen Arbeiten vorbehalten.

Klangfarbe: Klangfarbliche Veränderungen sind eine immanente Begleitererscheinung von Artikulationen. Wird ein Ton anders artikuliert, klingt er auch anders. Aus diesem Grund ist Artikulation in der Orchestrationspraxis ein wichtiges Werkzeug (Sevsay 2005). Hier kann, wie schon bei Hüllkurven und Intonation, davon ausgegangen werden, dass professionelle Sampler dies von sich aus mitbringen. Ein Pizzicato-Sample ist eben kein extrem verkürztes Legato-Sample, sondern ein explizit eingespieltes Pizzicato, das nur noch angesteuert werden muss. Klangfarbe losgelöst vom Prozess der Tonerzeugung/Klangsynthese zu betrachten, ist nicht sinnvoll und wird deswegen auch hier nicht weiter thematisiert.

Bislang unberücksichtigt blieb die Onset-Zeit des zu artikulierenden Tones. Sie kann ebenfalls durch die Artikulation beeinflusst werden. Riemann (1884) und Parncutt (2003) beschreiben *agogische Akzente*. Betonung wird hier durch Antizipation oder Verzögerung des Onsets erreicht. In diesem Zusammenhang zeigt sich, dass Artikulation sich nicht nur auf die Dynamik, sondern auch auf das Timing auswirken kann. Durch Rubato und Asynchronie wären dafür bereits mögliche Schnittstellen definiert (siehe Abschnitt 5.3.2).

Betrachtet man all die Toneigenschaften, die durch Artikulation beeinflusst werden, zeigt sich ein vielschichtiges Zusammenspiel. Scheinbar klare Anweisungen, die etwa die Tondauer betreffen, haben eben nicht nur auf die Tondauer Einfluss, was durch Messungen bestätigt werden konnte. Viele dieser Toneigenschaften sind bereits durch die tongebenden Verfahren (Sampler, Synthesizer) weitgehend automatisiert. Deshalb werden im folgenden lediglich für die Aspekte Tondauer und Tonlautstärke Modelle ausgearbeitet.

Jedoch werden gleichlautende Artikulationen nicht immer gleich ausgeführt. Selbst die Grundartikulation muss nicht über das ganze Musikstück einheitlich sein und sogar das Vokabular an Artikulationsanweisungen ist nicht feststehend. In der Barockzeit entstanden einige Basisartikulationen, wie Tenuto, Legato, Portato, Staccato und Pizzicato. Diese wurden in den folgenden Musikepochen bis heute erweitert, durch neue Spieltechniken und durch feinere Differenzierungen ergänzt. Es kann sogar vorkommen, dass mehrere Artikulationsanweisungen miteinander kombiniert werden, etwa ein akzentuiertes Staccato. Diese Flexibilität und Erweiterbarkeit soll auch durch das Modell unterstützt werden. Die Grundidee ist, dass die Definition von Artikulationen von ihrer eigentlichen Anwendung auf einen konkreten musikalischen Kontext formal getrennt wird. Dadurch wird die Definition austauschbar, kann durch alternative Versionen, welche die gleichen Anweisungen anders umsetzen, ersetzt werden. Dafür werden die Formalismen *Artikulationsstil* und *Articulationmap* eingeführt.

Der *Artikulationsstil* definiert die Artikulationen. Diese Definition beschreibt die Manipulationen, die durch die Artikulation an einer Eingabenote vorgenommen

werden. Hier kann jegliche Artikulation vom Anwender frei definiert werden. Weiterhin können mehrere Stile angelegt werden, die gleichlautende Artikulationen unterschiedlich umsetzen.

Die *Articulationmap* ist eine Sequenz von Anweisungen, die einen konkreten musikalischen Kontext artikulieren. Dafür werden nur noch High-Level-Deskriptoren verwendet, welche als Lookup-Referenzen auf die zugehörigen Definitionen im Artikulationsstil dienen. Es reicht also aus, einer Note die Anweisung »Staccato« beizufügen; die Umsetzung von »Staccato« ist im Artikulationsstil definiert. Durch diese Kapselung können der Artikulationsstil problemlos ausgetauscht und verschiedene Stile auf die gleiche Folge von Anweisungen angewendet werden.

Artikulationsstil

Ein Artikulationsstil $AStyle$ ist ein Element aus der Menge der definierten Stile

$$AStyleSet = \{AStyle_0, AStyle_1, \dots, AStyle_t\}$$

und definiert eine Menge von Artikulationsanweisungen A .

$$AStyle_s = \{A_0^s, A_1^s, \dots, A_b^s\} \quad \text{für } 0 \leq s \leq t$$

Eine Artikulationsanweisung ist wiederum eine Liste von Manipulatoren M , die auf die zu artikulierende Eingabernote angewendet werden.

$$A_a^s = (M_0^{s,a}, M_1^{s,a}, \dots, M_n^{s,a}) \quad \text{für } 0 \leq a \leq b$$

Ein solcher Manipulator bezeichnet das Notenattribut *attribute*, das er bearbeitet, sowie die Art *mode* und Wertigkeit *value* der Bearbeitung.

$$M_m^{s,a} = (mode_m^{s,a}, attribute_m^{s,a}, value_m^{s,a}) \quad \text{für } 0 \leq m \leq n$$

Es werden drei Manipulationsmodi unterschieden, die entweder Absolutwerte setzen oder relative Wertänderungen vornehmen (*note* ist die Eingabernote, *Attribute(note)* das zu editierende Notenattribut).

$$Attribute(note) := \begin{cases} value_m^{s,a} & : \text{ set mode} \\ Attribute(note) + value_m^{s,a} & : \text{ add mode} \\ Attribute(note) \cdot value_m^{s,a} & : \text{ relative mode} \end{cases}$$

Mögliche Attribute, die durch den Manipulator bearbeitet werden können, sind die Tonlautstärke (*velocity*), die symbolische Tondauer (*tickDur*) und die physikalische Tondauer (*msDur*). Die entsprechenden MIDI-Event-Attribute sind

$Velocity(note)$ und $Duration(note)$. Da ein Ton im MIDI-Standard immer aus zwei Events besteht, nämlich NoteOn und NoteOff, gibt es das Duration-Attribut eigentlich nicht. Bei der Implementierung wird deshalb ein Vorverarbeitungsschritt ausgeführt, der beide Events zu einem Note-Event mit Duration-Attribut zusammenfasst. Zur Wiedergabe wird er dann wieder aufgespalten.

$$Attribute(note) = \begin{cases} Velocity(note) : attribute_m^{s,a} = \text{“velocity”} \\ Duration(note) : attribute_m^{s,a} \in \{\text{“tickDur”}, \text{“msDur”}\} \end{cases}$$

Die physikalische Tondauer $msDuration$ ist insbesondere bei sehr kurzen Artikulationen, wie Staccato und Staccatissimo, wichtig. Diese sind unabhängig vom Tempo und der notierten Tondauer; sie setzen vielmehr eine absolute Millisekundendauer. Auch für Zäsuren (Atemzeichen) kann es sinnvoll sein, wenn der vorhergehende Ton um einen absoluten Millisekunden-Wert verkürzt wird, so dass die entstehende Atempause unabhängig vom zugrunde liegenden Tempo immer gleich lang ist. Das Duration-Attribut ist allerdings ein Tick-Wert. Deshalb werden bei der Artikulation mit $msDur$ die Millisekunden-Daten in Tick-Daten zurückgerechnet.

Um die hier nicht umgesetzten aber oben beschriebenen übrigen Toneigenschaften (Hüllkurve, Intonation, Klangfarbe) ebenfalls zu unterstützen, können die entsprechenden Attribute im Manipulator (*attribute*-Objekt) mit eingebaut und durch die zugehörigen Funktionalitäten hinterlegt werden. In einer MIDI-Implementierung würden dann entsprechende Controller-Events direkt vor dem Note-Event in die Sequenz eingefügt, welche die entsprechenden Einstellungen am Synthesizer/Sampler vornehmen.

Diese Funktionalität kommt bereits bei der Steuerung des *Vienna Instruments*-Samplers zum Einsatz. Hier kann zwischen verschiedenen Artikulations-Kategorien umgeschaltet werden. Es gibt spezialisierte Sample-Sets u.a. für Portamento-, Legato-, Marcato-, Staccato- und Pizzicato-Artikulation. Einige dieser Kategorien unterscheiden ferner zwischen schneller und langsamer Spielweise. So sind Samples für langsames Legato schwerfälliger als Samples für schnelles Legato oder Triller. Für jeden Ton und seine Artikulation muss also die am besten geeignete Sample-Kategorie automatisch berechnet und per Controller-Steuerung im Sampler angesteuert werden. Dann erst wird der eigentliche NoteOn-Event an den Sampler geschickt.

Auch Namenskonventionen werden in der Implementierung ausgenutzt. Anstelle eines numerischen Index a , der quasi als Zeiger auf die Artikulationsanweisung A_a^s im zugrunde liegenden Stil $AStyle_s$ dient, werden in der Implementierung frei definierbare Zeichenketten verwendet.⁹ Sie erlauben es, jeder Artikulation eine aussagekräftige Bezeichnung zu geben, etwa ein Staccato

⁹Innerhalb des Stiles müssen die Bezeichnungen freilich eindeutig sein.

auch als solches zu bezeichnen $A_{staccato}^s$. Das macht die Articulationmap lesbarer und wird bei der Ansteuerung der Sample-Kategorien in *Vienna Instruments* ausgenutzt. Trägt die Artikulationsanweisung etwa bereits »staccato« im Namen (auch als Substring), wird die entsprechende Sample-Kategorie angesteuert. Das derzeitige Vokabular beinhaltet »portamento«, »legato«, »moltoLegato«, »legatissimo«, »nonlegato«, »portato«, »marcato«, »tenuto«, »staccato«, »staccatissimo«, »spiccato«, und »pizzicato«. Tabelle 5.4 zeigt einen typischen Artikulationsstil.

Articulationmap

Die Articulationmap M_A ist eine zeitlich geordnete Liste mit zwei Typen von Elementen: Stilschalter (*switch*) und Artikulatoren (*articulator*). Ein Stilschalter ist ein Tripel

$$switch = (d, s, a_{standard}) \quad \text{für} \quad AStyle_s \in AStyleSet, A_{a_{standard}}^s \in AStyle_s$$

mit dem Tick-Datum d , von dem an der Artikulationsstil $AStyle_s$ allen folgenden Artikulationsanweisungen zugrunde gelegt wird. Sein Zuständigkeitsbereich endet mit der nächsten *switch*-Anweisung. Damit jeder Artikulationsanweisung ein Stil eindeutig zugewiesen werden kann, muss die erste Anweisung in der Articulationmap eine *switch*-Anweisung sein.

Durch den Parameter $a_{standard}$ wird die Grundartikulation angezeigt, die auf alle Töne angewendet wird, für die keine sonstigen Anweisungen durch Artikulatoren zugewiesen sind. Die Grundartikulation muss selbstverständlich im durch die *switch*-Anweisung gesetzten Artikulationsstil $AStyle_s$ enthalten sein.¹⁰

Die übrigen Elemente in der Articulationmap sind Artikulatoren. Sie weisen einer Note (bzw. einem Note-Event) eine Liste von Artikulationsanweisungen zu.

$$articulator = (note, (a_0, a_1, \dots, a_k)) \quad \text{für} \quad A_{a_0 \dots k}^s \in AStyle_s$$

Anstelle der a -Indizes werden in der Implementierung die Artikulationsnamen verwendet, wie oben beschrieben. Um eine Note mit einer Staccato-Artikulation zu versehen, hat der Artikulator dann die Form $(note, ("Staccato"))$. Das Objekt *note* codiert alle notwendigen Informationen, um den Note-Event in der zu artikulierenden MIDI-Sequenz zu finden: Tick-Datum, Tonhöhe, Dauer, MIDI-Port und -Kanal. Meist genügt das Tick-Datum. Sollte keine eindeutige Zuordnung möglich sein, werden alle in Frage kommenden Note-Events artikuliert.

¹⁰In der Implementierung ist das Attribut $a_{standard}$ optional. Fehlt es, wird »Nonlegato« als Grundartikulation angenommen. Fehlt auch dieses, bleiben alle betroffenen Note-Events unverändert.

Bezeichnung der Artikulation	Modifikatoren		
	<i>mode</i>	<i>attribute</i>	<i>value</i>
Accent	add	velocity	25
Sforzato	set	velocity	127
Cesura	add	msDur	-400
Pizzicato	set	msDur	1
Spiccato	set	msDur	160
	add	velocity	25
Staccatissimo	set	msDur	160
	add	velocity	5
Staccato	set	msDur	180
	add	velocity	-5
Hat	relative	tickDur	0.75
	add	velocity	25
Portato	relative	tickDur	0.8
Marcato	relative	tickDur	0.8
	add	velocity	25
Tenuto	relative	tickDur	0.9
	add	velocity	12
Nonlegato	relative	tickDur	0.95
Legato	relative	tickDur	1
MoltoLegato	add	tickDur	10
Legatissimo	add	tickDur	30
Portamento	add	tickDur	50

Tabelle 5.4: Die Einstellungen der Modifikatoren in einem typischen Artikulationsstil.

Es ist weiterhin möglich, mehrere Artikulationsanweisungen auf eine Note anzuwenden, etwa eine akzentuierte Legatonote (*note*, (“*Legato*”, “*Accent*”). Soll ein Akzent besonders deutlich sein, kann das durch einen doppelten Akzent erreicht werden. Die Anweisungen werden dann in der Reihenfolge eingearbeitet, in der sie aufgelistet sind. Dabei können sich Wechselwirkungen ergeben, die es

zu berücksichtigen gilt. Eine relative Wertänderung gefolgt von einer absoluten ergibt ein anderes Resultat als eine absolute Wertänderung gefolgt von einer relativen.

Diskussion

Dass Artikulationen frei definiert und kombiniert werden können, macht diesen Modellansatz sehr flexibel. Die Kombination von mehreren Artikulationen ist im Prinzip nichts anderes als die Vereinigung ihrer Manipulatoren-Listen. Auch diese Listen können frei mit Manipulatoren gefüllt werden, so dass sich komplexe Verhalten erzeugen lassen, etwa eine Kombination von absoluten und proportionalen Tonverkürzungen, wie sie Hähnel (2010) beschreibt. Nichtsdestotrotz sind die Grenzen der drei derzeit implementierten Manipulatoren-Modi (set, add, relative) schnell erreicht. Die Ergänzung um weitere komplexere Modi wäre eine sinnvolle Weiterentwicklung.

Ferner sind Hüllkurven-Modifikatoren als wertvolle Erweiterung der Modifikatoren-Funktionalitäten zu nennen. Die im menschlichen Spiel beobachteten Hüllkurven-Verläufe sind deutlich variantenreicher als die Grundeinstellungen mancher Sampler und Synthesizer. Wo die Möglichkeit besteht, Hüllkurven zu editieren, sollte dieses Potential genutzt werden, das noch einen hörbaren Qualitätsgewinn einbringt.

Seine Flexibilität ist ein Vorteil des hier beschriebenen Modellansatzes. Genau dieser Aspekt wirft auch einige besonders für die Interpretationsforschung interessante Fragen auf. Wie beeinflussen sich Artikulationen, wenn sie miteinander kombiniert werden? Im hier vorgestellten Modellansatz werden die Manipulatoren einfach nacheinander ausgeführt. Dabei verstärken sich gleich lautende Manipulatoren (Verkürzung + Verkürzung = noch kürzer). Gegensätzliche Manipulatoren heben sich gegenseitig auf (lauter + leiser = weniger oder keine Änderung). Die Set-Manipulatoren ignorieren schlicht alle vorhergehenden Änderungen. Würde ein Mensch dies auch so interpretieren oder würde er eher eine Durchschnittsartikulation bilden (Tenuto + Staccato \approx Portato)?

Hat der Kontext, in dem eine Artikulation gespielt wird, Einfluss darauf, wie sie gespielt wird? Wird etwa die Umsetzung einer Artikulationsanweisung durch die Grundartikulation beeinflusst? Wird ein Staccato breiter gespielt, wenn die Grundartikulation Legato ist, anstatt Nonlegato? Kann die Schwerpunktbetonung vielleicht einen ähnlichen Einfluss haben? Beim Menschen können Varianten beobachtet werden, die sich vielleicht auf solche Wechselwirkungen zurückführen lassen. Im hier beschriebenen Modell fehlen sie. Derzeit sind die Artikulationen kontextfrei definiert und werden immer gleich ausgeführt.

Gleich lautende Artikulationen können nichtsdestotrotz durch einen Wechsel des Artikulationsstiles in ihrer Ausführung verändert werden. Die *switch*-Anweisung ist ein mächtiges Werkzeug, das insbesondere das Gegenüberstellen unterschiedlicher Stile (z.B. in der Interpretationsforschung beim analysierenden Hören) vereinfacht. Stilwechsel kommen aber auch regulär innerhalb von Musikstücken vor. Typische Beispiele lassen sich in der Filmmusik finden, wo sie durch Szenenwechsel motiviert sind, und in sinfonischer Musik, wo unterschiedliche Formteile auch interpretatorisch kontrastiert werden sollen.

Experimente haben weiterhin gezeigt, dass nicht immer eine einzige Grundartikulation genügt. Es kann ganze Artikulationsschemata geben (der Walzer als Beispiel: lang-kurz-kurz). Weil hiervon mehr als nur die Lautstärke der Töne betroffen ist, reichen die Betonungsmuster aus der Mikro-Dynamik (Abschnitt 5.4.2) nicht aus. Es werden Artikulationsgrundschemas benötigt. Deren formale Gestalt kann sich an der von Schwerpunktbetonungen orientieren.

5.6 Diskussion

Alle in diesem Kapitel vorgestellten Modelle können sowohl *lokal* innerhalb einer einzelnen Stimme als auch *global* auf alle Stimmen angewendet werden. Eine typische Konstellation ist die folgende. Eine globale Tempomap sorgt dafür, dass alle Kanäle »zusammen spielen«. Das Zusammenspiel wird aber durch lokal unterschiedliche Mikro-Timing-Charakteristiken belebt. Eine Stimme mag ein Rubato stärker spielen als andere; eine andere Stimme mag immer etwas hinterher sein und rhythmisch sehr unsauber spielen. Die Modelle sind so flexibel gehalten, dass eine große Bandbreite an Charakteristiken und Kombinationen möglich ist, mehr als in der Praxis beobachtet werden kann. Wo sinnvoll, sind Schnittstellen eingebaut, die Neudefinitionen erlauben (eigene Dynamikanweisungen, eigene Artikulationen, eigene Schwerpunktbetonungsmuster usw.).

Für die Interpretationsforschung wird damit ein wertvoller Werkzeugkasten bereitgestellt. Die Modelle können dazu dienen, Musikaufnahmen auf bestimmte Phänomene hin zu analysieren. Für die Tempoanalyse entstehen bspw. durch lange Generalpausen oft Probleme. In Abwesenheit jeglicher metrischer Anhaltspunkte wird der Tempoverlauf sofort mehrdeutig, trotzdem scheinen Hörer und Musiker ein gewisses Tempo weiter zu empfinden und die Länge der Generalpause nicht beliebig zu sein. Mit dem in Abschnitt 5.3.1 vorgestellten Modell können nun plausible Tempoverläufe auch über Pausen hinweg simuliert werden. Zudem sind die Modelle die Grundlage für die Interpretations-Synthese. Interpretationen können nachgebaut und in allen Details feinjustiert werden. Damit wird u.a. für die Analyse durch Synthese (*Analysis by Synthesis*)

und für musikpsychologische Studien ein mächtiges Werkzeug bereitgestellt. Tatsächlich kamen die Modelle bereits in einer Serie von musikwissenschaftlichen und -psychologischen Studien¹¹ in unterschiedlichster Weise zum Einsatz. Es folgt ein kurzer Überblick.

Hörerbewertung: Hierbei wurden den Probanden mehrere Musikausschnitte (alles Barockmusik, schwerpunktmäßig Menuette) vorgespielt, die als reale Einspielungen bzw. Nachbildungen realer Einspielungen bewertet werden sollten. Dabei wurden echte Einspielungen, aufwendig produzierte exakte Nachbildungen, durch die Modelle approximierten Nachbildungen sowie mit den Modellen erzeugte entartete (sprich untypische bis unsinnige) Interpretationen gegenübergestellt.

Interaktion: Die Probanden hatten die Möglichkeit, die Interpretation eines 8-taktigen Musikausschnittes selbst anhand verschiedener Parameter einzustellen. Dabei konnten Timing (speziell Rubato), Dynamik (insbesondere Schwerpunktbetonung) und Artikulation (speziell Tondauer) eingestellt werden. Die Kürze des Musikausschnittes, der in Endlosschleife lief, war für Experimente mit Makro-Merkmalen nicht geeignet. Hier wurde untersucht, welche Präferenzen (typische Einstellungen) es gibt, wie die Teilnehmer den Raum der möglichen Einstellungen explorierten und inwiefern sie in der Lage waren, einzelne Merkmale aus dem Zusammenspiel zu erkennen (Was glauben Sie, was Sie eingestellt haben? Was war Ihre Absicht?).

Tapping-Experiment: Hierbei diente speziell das Rubatomodell dazu, hörpsychologische Effekte zu untersuchen. Wie gut oder schlecht können Menschen dem inégalen Spiel von 3er-Takten folgen (zur Inégalité siehe Hefling 1993)? Welche Lerneffekte gibt es? Erst durch das Rubatomodell war es möglich, die Timing-Verzerrungen beim inégalen Spiel auf Millisekunden genau zu gestalten und zu steuern. Eine solche Experimentiergrundlage können menschliche Einspielungen nicht bieten.

Für die Diskussion der hier vorgestellten Modelle ist insbesondere das erste Experiment (Hörerbewertung) von Bedeutung. Im ersten Test sollten die Probanden die Aussage »Die aktuelle Einspielung ist echt« bewerten (1: trifft zu, 2: trifft wahrscheinlich zu, 3: trifft wahrscheinlich nicht zu, 4: trifft auf keinen Fall zu). Im zweiten Test lautete die Aussage »Diese synthetische Version ist eine Kopie einer realen Interpretation«; hier gab es folglich keine Originalaufnahmen. Die 1:1-Kopien stimmten in Timing, Dynamik und Artikulation mit den Originaleinspielungen überein. Sie unterschieden sich von den Originalen

¹¹Die Studien wurden von Tilo Hähnel geplant und durchgeführt.

durch die klanglich etwas eingeschränkteren Möglichkeiten des Samplers und die Akustik. Da diese auch nicht Untersuchungsgegenstand waren, können die 1:1-Kopien als Referenz angesehen werden. Die mit den Modellen erzeugten Interpretationen entsprachen ebenfalls den Originalinterpretationen, bereinigt um das *Performance-Rauschen*; das sind diejenigen Abweichungen in den Ausgangs- und Messdaten, die durch die Modelle nicht systematisch erklärt und nachgebildet werden konnten. Auch die rhythmische Ungenauigkeit (Abschnitt 5.3.2) kam hier nicht zum Einsatz. Diese Versionen werden im weiteren als *plausible Modelle* bezeichnet. 46 Probanden nahmen an der Studie teil, 23 davon waren Experten für historische Aufführungspraxis. Die Ergebnisse sind in Tabelle 5.5 zusammengefasst.

Eine Signifikanzanalyse der Unterschiede in den Bewertungen ergab, dass die Originale von Experten und Nicht-Experten signifikant besser bewertet wurden. Erstaunlicher Weise schnitten im zweiten Test sogar die flachen Interpretationen, welche sich ausschließlich auf die Artikulation reduzierten, in der Expertenbewertung signifikant besser ab als die 1:1-Kopien (Signifikanz: 0.013). Im Vergleich der Bewertungen von plausiblen Modellen und 1:1-Kopien zeigte sich keine Signifikanz. Die Modelle erweisen sich damit als gleichauf mit den detailgetreuen Kopien. Das deutet darauf hin, dass die Modelle typische Charakteristiken menschlichen Spiels plausibel nachbilden können.

Dass es Performance-Rauschen gibt, zeigt aber auch, dass noch immer nicht alle interpretatorischen Merkmale bekannt und erfasst sind. Eine ganze Reihe von Phänomenen lässt sich aus der Kommunikation im Ensemblespiel herleiten; die Musiker reagieren auf das, was sie vom Rest des Ensembles wahrnehmen. Auch spieltechnische Einflüsse sind bislang nicht berücksichtigt. Manche Passagen im Musikstück mögen schwer zu spielen sein, sind deshalb anfälliger für Abweichungen; manche Akkorde müssen auf dem Klavier arpeggiert¹² werden, weil sie in einer so weiten Lage sind, dass zwei Hände sie nicht auf einmal greifen können. Es bestehen auch Wechselwirkungen zwischen verschiedenen Merkmalen (z.B. zwischen Schwerpunktbetonung und Rubato), die noch nicht vollständig erforscht sind. Außerdem wird die Mehrheit der Musiker schwerlich in der Lage sein, ein und dieselbe Interpretation mehrmals exakt zu reproduzieren, schon gar bei veränderten Rahmenbedingungen (z.B. an unterschiedlichen Stellen im Konzertprogramm, zu unterschiedlichen Tageszeiten, bei unterschiedlichen Temperaturen, mit unterschiedlichen Instrumenten). Dort, wo momentan noch systematische Erklärungen und Modelle fehlen, könnten zufallsbehaftete Einflüsse das »Rauschen« nachbilden. Auf der Timing-Ebene wurde hierfür bereits die rhythmische Ungenauigkeit eingeführt. Aber auch in allen anderen Domänen können solche Phänomene beobachtet werden; auch hier kann eine Zufallskomponente von Nutzen sein.

¹²Die Töne werden zeitversetzt angeschlagen.

Art der Einspielungen	Bewertung der 23 Nicht-Experten		Bewertung der 23 Experten	
	Mittelwert	Standardabw.	Mittelwert	Standardabw.
»Die aktuelle Einspielung ist echt.«				
Originale	1.70	0.470	1.35	0.487
1:1-Kopien	2.61	0.656	3.65	0.573
Modelle plausibel	2.52	0.665	3.35	0.775
Modelle entartet	2.35	0.775	3.26	0.915
»Diese synthetische Version ist eine Kopie einer realen Interpretation.«				
1:1-Kopien	2.22	0.751	2.44	0.609
Kopien invers	2.76	0.705	2.96	0.767
Modelle plausibel	2.26	0.619	2.48	0.846
Modelle entartet	2.94	0.460	3.11	0.656
nur Artikulation	2.22	0.736	1.87	0.757

Tabelle 5.5: Ergebnisse der Hörerbewertung (Kopien invers: die Charakteristiken von Tempo, Makro-Dynamik, Rubato und Schwerpunktbetonung sind invers zu den 1:1-Kopien)

Die Experimente geben weiterhin einen Eindruck von den vielfältigen Anwendungsmöglichkeiten; eine weitere stellt Kapitel 6 vor. Bei der Musik-Produktion hat sich die Parametrisierung der Modelle als besonders vorteilhaft erwiesen, denn sie orientiert sich an den Beschreibungen und Absichten, die man auch als Interpret hat. Soll etwa eine Schwerpunktbetonung stärker sein, wird ihre Skalierung vergrößert. Soll ein Crescendo früher lauter werden, wird der *Protraction*-Wert kleiner gestellt. Der Anwender stellt nicht direkt die Parameter etwa einer Bézierkurve ein, sondern intuitivere Parameter, welche in die eigentlichen Modellparameter transformiert werden. Bedenkt man jedoch, wieviele Details eine Interpretation enthalten kann, ist der Editierungsaufwand immer noch sehr groß. Eine graphische Benutzeroberfläche, die bspw. mit Notations-Metaphern arbeitet, kann die Zeiteffizienz noch deutlich steigern.

Außerdem könnten Funktionalitäten zur vollautomatischen Performance-Planung einen Teil der Arbeit abnehmen, zumindest den Initialisierungsaufwand verringern. Eine »Grundinterpretation« kann sehr schnell automatisch erstellt werden. Der Anwender müsste nur noch Details und Besonderheiten manuell einarbeiten. Wieviel Automation nötig ist, wird sich jedoch von Fall zu Fall entscheiden müssen. Die meisten Künstler werden nur zögernd Aspekte der musikalischen Arbeit aus der Hand geben. In gewissen Phasen des kunstschaftenden Prozesses haben sich Automatismen aber als sinnvolle und willkommene Unterstützung erwiesen (Gelineck & Serafin 2009). Ein virtueller Interpret hilft bspw. in Notensatz-Programmen, einen ersten Höreindruck schon früh im Schaffensprozess zu erhalten. In der Filmmusik ist es üblich, dass vor der Einspielung mit echtem Orchester synthetische Versionen als *Temptracks* aufwendig produziert werden. Nach ihnen wird dann eine vorläufige Schnitfassung des Filmes erstellt. Etliche Musikverlage nutzen synthetische Hörproben in Verbindung mit der angebotenen Notenliteratur.

5.7 Zusammenfassung

Musik kann auch durch den Computer ausdrucksvoll dargeboten werden. Die Fülle an interpretatorischen Details ist groß und selbst Musikern und Musikwissenschaftlern heute noch immer nicht zur Gänze bekannt. Doch die für die Ausdruckskraft einer Interpretation wichtigen Details, welche die Musiker auch bewusst gestalten, lassen sich durch Modelle, wie die in diesem Kapitel vorgestellten, adäquat nachbilden. Das hat die Hörerbewertung klar gezeigt. Defizite gibt es immer noch bei der klanglichen Nachbildung realer Instrumente. In besonderem Maße sind es Streich- und Blasinstrumente, die ein außerordentlich wandlungsfähiges Spiel ermöglichen. Im direkten Vergleich mit echten Einspielungen wirkt das Spiel mancher Sampler und Synthesizer – den virtuellen Doubles –

noch immer ungelenkt. Andere Instrumente, allen voran der Konzertflügel, können bereits beeindruckend authentisch auch virtuell zum Erklängen gebracht werden. Je mehr wir über Tonerzeugung und Spieltechnik anderer Instrumente lernen, desto besser und täuschend echter können auch sie nachgebildet werden. Modelle für Fingersatz und Bogenführung gibt es bereits (Al Kasimi, Nichols & Raphael 2007, Hall & O'Donnell 2009, Hall, O'Donnell & Findlay 2009, Maestre 2009, Tuohy & Potter 2006). Sie kommen bspw. bei der Notenedition und Instrumentalbildung zum Einsatz. Aufbauend auf der hier vorgestellten Grundlagenarbeit zur Interpretations-Synthese gehört zu den künftigen Perspektiven neben der Erweiterung des Modellbestandes auch der Schritt auf die nächsthöhere Abstraktionsebene, die algorithmische Interpretations-Planung.

Der Computer als Interpret hat seine Daseinsberechtigung und emanzipiert sich mehr und mehr gegenüber dem Menschen. Doch wird er allenfalls die »Handwerker« unter den Musikern ersetzen, diejenigen, die lediglich umsetzen, was ihnen das Notenblatt diktiert. Nicht zu ersetzen sind aber diejenigen, die diese Interpretationen erarbeiten, ein Stück immer wieder neu zu interpretieren versuchen, sich nicht mit den gedruckten Einzeichnungen zufrieden geben, sondern immer neue Details kreativ einarbeiten – die »Künstler«. Für sie kann das hier Vorgestellte ein kreatives Werkzeug und Unterstützung im Schaffensprozess sein. Mehr noch; der virtuelle Interpret kann synthetische Instrumente jenseits von Geige, Klavier und Trompete zum Erklängen bringen. Und er kann zum integralen Bestandteil des Werkes werden; ein Werk etwa, mit dem man interagieren kann: ein *reaktives Orchester*, wie es das folgende Kapitel vorstellt.

Das reaktive Orchester

EINE typische Situation bei den Musikaufnahmen für einen Kinofilm: Das Orchester sitzt mit dem Rücken zur Leinwand. Nur der Dirigent sieht die Filmsequenz und dirigiert entsprechend. Das Orchester folgt seinem Dirigat. Damit beeinflusst er nicht nur das Timing der Musik. Er kann auch einen gewünschten Ausdruckscharakter anzeigen, kann Crescendi und Decrescendi von den Musikern einfordern, kann eine gesangliche oder rhythmisch akzentuierte Spielweise signalisieren. In Zeiten des Stummfilms musste sich der Dirigent darauf verlassen können, dass das Orchester zuverlässig auf ihn reagierte. Sonst konnte die Musik nicht nur in Asynchronie zum Film geraten; sie konnte ausdrucksmäßig vollkommen missraten und nicht mehr zum Film passen.

Der Film macht es den Musikern dabei noch relativ leicht. Er steht fest. Das dazugehörige musikalische Geschehen kann vorher einstudiert werden. Nicht so in interaktiven Medien. Hier steht nicht einmal die Reihenfolge der Szenen fest, geschweige denn ihre Dauer. Fertige Musikstücke vorher einzustudieren und aufzunehmen, bringt nichts. Hier enden die Möglichkeiten menschlicher Musiker. Und es beginnen die Möglichkeiten virtueller Interpreten: Angenommen, es wäre kein reales sondern ein virtuelles Orchester, das die Musik live spielt, während etwa der Spieler das Spiel spielt. Dass auch virtuelle Interpreten zu einem verblüffend authentischen ausdrucksvollen Spiel in der Lage sind, hat Kapitel 5 gezeigt. Angenommen, der Dirigent ist kein Mensch sondern eine Musik-Engine, die den virtuellen Musikern während des interaktiven Geschehens die entsprechenden Signale gibt, eine Veränderung der Darbietungsweise einfordert oder sogar einzelnen Stimmen Schweigen gebietet bzw. schweigenden Stimmen Einsätze gibt (paralleles Arrangement, siehe Abschnitt 3.2). Solch ein Orchester ist in der Lage, in Echtzeit auf das interaktive Geschehen zu reagieren.

In diesem Kapitel werden die Techniken für eine derartige nichtlineare Interpretation und paralleles Arrangement vorgestellt. Dem liegen die Modelle aus Kapitel 5 und die folgenden Veröffentlichungen zu Grunde:

- (Berndt & Hartmann 2007b),
- (Berndt & Theisel 2008),
- (Berndt 2010).

Die Mittel der ausdrucksvollen Interpretation sind aus dem vorhergehenden Kapitel bekannt. Sie erlauben es, ein und das selbe Notenmaterial ausdrucksmäßig sehr unterschiedlich einzufärben; rhythmisch schwungvoll, aggressiv akzentuiert, gesanglich, getragen usw. Die Modelle, mit denen diese interpretatorischen Merkmale beschrieben werden, führen eine neue Abstraktionsebene in die Musikrepräsentation ein. Einige aufeinander folgende Noten in einer MIDI-Sequenz, jede mit ihrem eigenen Lautstärkewert, sind nun keine losen Einzelergebnisse mehr, deren Zusammenhang erst aufwendig herausanalysiert werden muss. Vielmehr kann nun leicht abgelesen werden, dass sich ihre Lautstärken aus einem Zusammenspiel von Makro-Dynamik, Schwerpunktbetonung und Artikulation ableiten lassen. Nun ist es problemlos möglich, nur einen einzelnen dieser Aspekte zielgerichtet zu verändern, etwa das Muster der Schwerpunktbetonung, ohne dass andere davon beeinträchtigt werden.

Auf Seiten des parallelen Arrangements wird mit *Orchestration* und *Baukasten-satzweise* gearbeitet, d.h. mit der Parallelmontage von klangfarblichem und musikalischem Material. Die *Orchestration* beschäftigt sich speziell mit den klangfarblichen Mitteln. An dieser Stelle sollte jedoch genauer unterschieden werden zwischen *Orchestration* und *Instrumentation*, die gemäß Ertuğrul Sevsay (2005) folgendermaßen definiert sind.

Instrumentation ist die Lehre von der Kombination gleicher oder verschiedener Instrumente, um einen gewünschten klangfarblichen und tonalen Effekt zu erzielen. »Dabei werden folgende Faktoren dargestellt: dynamische Balance, klangfarbliche Kontraste oder Ähnlichkeiten, Artikulation, die Verwendung von verschiedenen Registern der Instrumente und verschiedenen Lagen des Orchesters, verschiedene Spieltechniken auf demselben Instrument.«¹ Die Instrumentationslehre schließt also auch einige interpretatorische Aspekte mit ein.

Orchestration beschäftigt sich mit den ästhetischen Konzepten, die einer bestimmten Instrumentation zugrunde liegen. »Hier geht es darum, ähnlich

¹Sevsay (2005), Seite 17

oder verschieden ›instrumentierte‹ Abschnitte auszuwählen und miteinander zu kombinieren, Kontraste zwischen diesen Abschnitten zu erzeugen, zu vergrößern bzw. zu verringern, um den musikalischen und stimmungsmäßigen Gehalt oder den generellen Charakter einer Komposition umzusetzen bzw. zu unterstreichen. Orchestration spielt bei der Verdeutlichung der musikalischen Form eine wesentliche Rolle, da die orchestrale Klangfarbe viel leichter wahrnehmbar ist als abstrakte formale Aspekte. So werden also die *Farben (Instrumentation)* nach einer gewissen *Ästhetik (Orchestration)* kombiniert, um die Form zu verdeutlichen.«²

Bei variabler Instrumentation bleibt die Musik inhaltlich doch unverändert. Nur die Besetzung und Spielweise ändert sich. Dieses Prinzip wird bei der *Baukastensetzweise* allgemeiner gefasst. Hier werden nicht nur verschiedene Klangfarben, sondern verschiedenes musikalisches Material miteinander kombiniert. Eine Melodie kann durch eine Alternative ersetzt werden; ein Rhythmus-Pattern kann durch Hinzunahme oder Wegfall von optionalem Material angereichert oder ausgedünnt werden; Harmonien können umgedeutet werden. Einige Beispiele wurden bereits in Kapitel 4 beschrieben. Unterschiedliche Instrumentationen sind, so gesehen, nur ein Spezialfall des Baukastenprinzips.

Gemeinsam spannen ausdrucksvolle Interpretation, Orchestration und Baukastensetzweise ein außerordentlich großes Ausdrucksspektrum auf, das sich als Grundlage für musikalische Nichtlinearität bestens eignet. Eine erste große Herausforderung, nämlich die formale Beschreibung von ausdrucksvollen Interpretationen und ihre klangliche Umsetzung (Rendern ausdrucksvoller MIDI-Sequenzen), ist durch Kapitel 5 bereits bewältigt. Ein weiterer Formalismus speziell für das parallele Arrangement, der beschreibt, welche Stimmen/Instrumente/Kanäle miteinander kombiniert werden, ist nicht notwendig. Einigen Stimmen, die in einer bestimmten Interpretation nicht besetzt sind oder streckenweise aussetzen sollen, Schweigen zu gebieten, soll kein Problem sein, wenn man ohnehin die Dynamik einer jeden Stimme steuern kann. Das parallele Arrangement geht gänzlich in Interpretation auf.

Die zweite große Herausforderung, der sich dieses Kapitel im Speziellen widmet, ist der eigentliche Wandlungsprozess, wenn von einer Interpretation zu einer anderen gewechselt werden soll. Die unterschiedlichen Versionen müssen musikalisch plausibel und vollautomatisch ineinander übergeleitet werden. Doch wann ist eine Überleitung plausibel? Dazu lohnt es, sich noch einmal zu vergegenwärtigen, dass Interpretation und Instrumentation prominent an die musikalische Form gekoppelt sind.

Ein beliebig platzierter harter Wechsel der Interpretation (*Performance Cut*) ist ebenso wie ein naives Interpolieren von bspw. Tempo- und Dynamikkurven

²Sevsay (2005), Seite 17

(*Performance Cross-Fade*) keine adäquate Lösung, denn beides trägt eben nicht dem formalen Kontext Rechnung, in dem der Wechsel geschieht. Jegliche Veränderung, die im Zuge einer Überleitung vorgenommen wird, muss an der kompositorischen Struktur des Musikstücks ausgerichtet sein. Eine Veränderung, die etwa mit dem Beginn einer neuen melodischen Figur einhergeht, ist plausibler als eine Veränderung mitten in einer Figur, weil sie sich für den Hörer leichter als interpretatorische Absicht erschließt. »Hier beginnt etwas neues, also wird es auch anders gespielt«, die Interpretation ist nachvollziehbare Konsequenz aus der musikalischen Form.

Weiterhin ist unmittelbar einzusehen, dass unnötige Veränderungen für den Hörer auch schwer nachzuvollziehen sind. Wenn das Ausgangstempo und das Zieltempo identisch sind, ist eine zusätzliche Temposchwankung nicht notwendig. Das erscheint trivial, impliziert aber eine wichtige Erkenntnis. Die Überleitung hat nicht die Aufgabe, neues zu generieren! Vielmehr soll sie die aktuell spielende Interpretation nur so wenig wie nötig adaptieren, um den Anschluss an die Zielinterpretation herzustellen.

Ferner darf nicht vergessen werden, dass die Musik ein interaktives Medium vertont. Hier ist die Zeit zwischen Benutzereingabe und Reaktion begrenzt. Diese Latenz kann je nach Anwendungsfall verschieden sein. Es gilt also, innerhalb dieses von der Anwendung gesetzten Zeitfensters eine geeignete Überleitung zu finden. Extrem kurze Latenzvorgaben können den Spielraum dafür stark eingrenzen; auch damit muss umgegangen werden. Nichtsdestotrotz kann, noch während eine Überleitung ausgeführt wird, schon wieder die nächste Benutzereingabe anfallen und einen neuerlichen Wechsel der Interpretation fordern. Das Verfahren muss also so flexibel sein, auch aus einer laufenden Überleitung heraus eine neue Überleitung zu berechnen.

An die Überleitung bzw. das Verfahren, das sie erzeugt, werden zusammenfassend die folgenden vier Anforderungen gestellt.

1. Ausrichtung an der musikalischen Form
2. Effizienz (so wenig Änderung wie nötig)
3. Einhalten von Latenzen, die von der Applikation gegeben werden
4. Flexibilität (Überleitung auch aus einer Überleitung heraus)

In den weiteren Abschnitten dieses Kapitels wird eine Lösung hierfür gearbeitet. Die Grundidee dabei gestaltet sich folgendermaßen. Zu einem Musikstück (im MIDI-Format) gibt es eine Menge von Interpretationen (im XML-Format, basierend auf den in Kapitel 5 eingeführten Modellen). Damit können ausdrucksvolle MIDI-Sequenzen gerendert werden. Der Dirigent hat quasi

mehrere Interpretationen mit dem Orchester vorbereitet und muss nun nur noch signalisieren, wenn gewechselt werden soll. Den Übergang führt das Orchester jedoch nicht homophon aus (alle machen das gleiche), sondern jeder Musiker für sich entsprechend der jeweiligen musikalischen Situation in seiner Stimme. Diese Autonomie wird durch einen Agenten-basierten Ansatz nachempfunden. Jeder Musiker berechnet seinen eigenen Übergang von der Anfangs- zur Zielinterpretation. Daraus wird eine neue Interpretation, die Übergangsinterpretation, erzeugt. Mit dieser kann die Musik erneut gerendert und die Wiedergabe nahtlos darauf umgeschaltet werden, vergleichbar mit dem *Double Buffering* in der Computergraphik (Foley, van Dam, Feiner & Hughes 1997).

6.1 Engine-Architektur

Trotz des für alle Musiker im Orchester gleichen Dirigates ist jeder autonom. Diese Eigenständigkeit belebt nicht nur das Ensemblespiel, wie in Abschnitt 5.6 am Beispiel des Timings beschrieben wurde. Spätestens in polyphoner Musik hat auch jede Einzelstimme ihre eigene Form. Wo die eine Stimme gerade eine Phrase beendet, kann eine andere mitten in einer Phrase sein. Ebenso polyphon versetzt reagieren die Musiker auch auf dirigentische Anweisungen, jeder an der für ihn am sinnvollsten erscheinenden Stelle. Diese Autonomie führt zu einer wichtigen Design-Entscheidung: Auch die virtuellen Interpreten sollen autonom agieren.

Dafür ist ein Agenten-basierter Systemansatz prädestiniert. Jeder MIDI-Kanal wird durch einen Software-Agenten repräsentiert, eine unabhängig von den anderen agierende selbständige Einheit mit den folgenden Eigenschaften.

Autonomie: Jeder Agent ist ausschließlich für seinen Kanal (seine Stimme) zuständig. Er rendert seine eigenen expressiven MIDI-Sequenzen, spielt sie in seinem eigenen Echtzeit-Thread ab und berechnet die aus seiner lokalen Sicht bestmögliche Überleitung.

Reaktivität/Adaptivität: Die Agenten unterliegen einem globalen Kontext, dem »Dirigat«, über welchen das globale Verhalten des Systems koordiniert wird. Das bezieht sich speziell auf die gerade geforderte Interpretation und die Latenzvorgaben im Falle eines Wechsels.

Kommunikativität: Jeder Agent plant und rendert seine eigene, aus lokaler Sicht bestmögliche, Überleitung. Damit bei Tempoänderungen aber keine unerwünschte Asynchronie entsteht, müssen die Agenten hier eine gemeinsame Lösung finden. Hierfür werden die Überleitungspläne aller zu einer gemeinsamen Tempoüberleitung vereinigt.

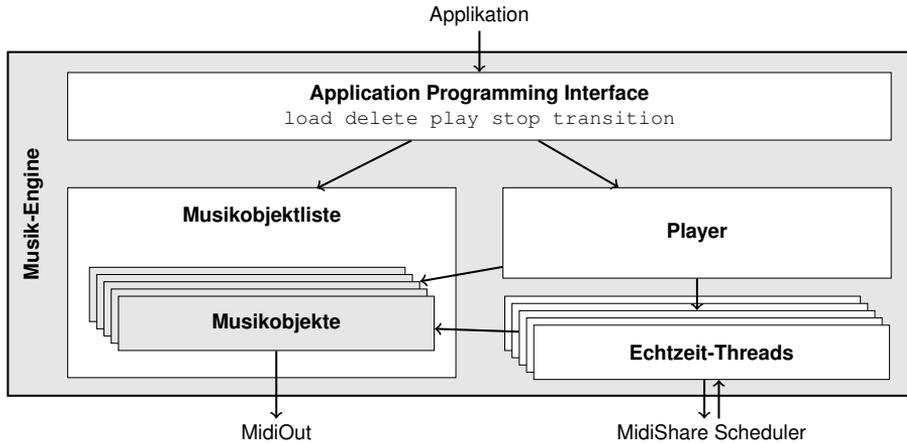


Abbildung 6.1: Der Grundaufbau der Musik-Engine.

Weitere Eigenschaften, die über die konkrete Aufgabenstellung des Überleitens von Interpretationen hinausgehen, sind in Abschnitt 6.3 beschrieben. Die Agenten werden von der Musik-Engine beim Einlesen und der Vorverarbeitung der Musikdaten instantiiert und sind Teil der sog. *Musikobjekte*. Eine Übersicht über den Aufbau der Musik-Engine wird in Abbildung 6.1 gegeben. Die grundsätzlichen Elemente der Musik-Engine sind das Application Programming Interface (API), die Musikobjektliste und der Player.

Application Programming Interface

Das API ist denkbar schlicht gehalten und verbirgt komplexere Abläufe vor dem Anwender. Für ihn gestaltet sich die Steuerung der Musik-Engine ähnlich einfach wie die Bedienung eines CD-Players.

```
musicObjPtr* load(std::string filename);
void delete(musicObjPtr* m);
void play(musicObjPtr* m, long startDate, std::string perfStyle);
void stop();
long transition(musicObjPtr* m, std::string perfStyle, long maxLatency);
```

Musikobjekte können in die Musikobjektliste geladen und wieder aus ihr gelöscht werden (`load`, `delete`). Geladene Musikobjekte können wiedergegeben und ihre Wiedergabe beendet werden (`play`, `stop`). Der Aufruf der Funktion `play` verlangt nach drei Parametern, deren erster die Referenz auf das wiederzugebende Musikobjekt ist. Desweiteren wird die Startposition innerhalb des

Musikstücks angegeben (in MIDI-Ticks) sowie die Interpretation (`perfStyle`), in der die Musik dargeboten werden soll. Auch die letzte Funktion im Befehlsatz des API, `transition`, gestaltet sich ähnlich einfach: Das durch `m` angezeigte Musikobjekt soll innerhalb einer maximalen Latenz (`maxLatency`) zur geforderten Interpretation (`perfStyle`) überleiten. Die Funktion gibt eine ganze Zahl vom Typ `long` zurück. Dies ist die Zeitdauer (in Millisekunden), bis die Überleitung ausgeführt und die Zielinterpretation erreicht ist.

Nachdem der Musiküberleitung durch `maxLatency` ein gewisser Rahmen gesteckt wurde, ermöglicht der Rückgabewert von `transition` eine noch feinere Abstimmung von Seiten der Applikation. Das entspricht dem im Film üblichen Produktionsprozess. Hier arbeitet der Komponist zunächst auf Basis des Rohschnitts. Nach den Musikaufnahmen wird dann der finale Schnitt noch einmal daran angepasst. Im Computerspiel kann das wie in dem folgenden Beispiel aussehen. Zusammen mit einem Skript-Ereignis³ soll ein musikalischer Wandel einhergehen. Das Ereignis kann aber noch um bis zu 6 Sekunden verzögert werden; daraus leitet sich die maximale Latenz `maxLatency` für die Musiküberleitung ab. Nachdem die Überleitung berechnet ist, wird signalisiert, dass die Zielinterpretation schon nach 3 Sekunden erreicht ist. Also wird das Skript in 3 Sekunden gestartet. Eine derartige Feinabstimmung kann der Qualität der Inszenierung außerordentlich zuträglich sein, denn erst so können alle Teilmedien (Bild, Ton, Musik) gemeinsam »auf dem Punkt kommen«.

Player

Der Player ist quasi der Dirigent. Er steuert das globale Verhalten des Agentensystems (das Musikobjekt). Er ist selbst kein wiedergebendes sondern ein steuerndes Medium. Ebenso bringt nicht der Dirigent die Musik zum Klingen – das machen die Musiker –, aber er veranlasst sie dazu. Der Player fungiert als Schnittstelle zwischen den High-Level-Befehlen des API und den eigentlichen Funktionalitäten, die in den Musikobjekten implementiert sind. Er startet und verwaltet die Echtzeit-Threads, in denen die Agenten ihre MIDI-Sequenzen abspielen, und veranlasst Interpretationswechsel.

Für jeden Agenten, der in der momentanen Interpretation zu spielen hat, wird ein eigener Echtzeit-Thread erzeugt. Innerhalb seines Echtzeit-Threads verarbeitet der Agent anstehende MIDI-Events und schickt sie zur Ausgabe. Die MIDI-Events werden also nicht direkt an den Scheduler geschickt, der sie zur

³Skript-Sequenzen sind ein beliebtes Mittel in Computerspielen. Es sind vorgefertigte Ereignisse im Spielablauf, wie etwa das Auftauchen von Gegnern. Sie werden entweder zeitabhängig (z.B. alle 3 Minuten) oder Spieler-abhängig (z.B. erst dann, wenn der Spieler den Raum betritt) ausgelöst.

rechten Zeit an den Ausgabe-Port weitergibt (die herkömmliche Art der MIDI-Wiedergabe). Vielmehr wird der jeweilige Agent zur entsprechenden Zeit aufgerufen, der die Events noch filtern oder vor der Ausgabe verarbeiten kann (mehr dazu im Abschnitt Musikobjekte). Für die rechtzeitige Ausführung dieser Aufgaben sorgt der Echtzeit-Scheduler von *MidiShare* (Orlarey 1990, Fober, Letz & Orlarey 1996, 1999). *MidiShare* ist ein »real-time multi-tasking MIDI operating system specially devised for the development of musical applications«⁴.

Die Threads sind unabhängig voneinander; so kann auch jeder Agent sein eigenes Timing haben. D.h., dass die Reihenfolge der MIDI-Events, wie sie in der MIDI-Datei ursprünglich definiert war, für die ausdrucksvolle Wiedergabe nicht mehr gelten muss. Wenn in einer Stimme ein Rubato vorgeschrieben ist, verschieben sich alle ihre Events innerhalb des Rubato Frames, während andere Stimmen, für die keine solche Timing-Verzerrung vorgeschrieben ist, unverändert bleiben oder vielleicht in ganz anderer Weise verzerrt werden. Prinzipiell kann sogar jeder Agent in einem eigenen Tempo spielen, was zu einer zunehmenden Asynchronie führt. In aleatorischer Musik wird bspw. mit solchen Mitteln gearbeitet.

Musikobjekte

Für jedes geladene Musikstück wird ein Musikobjekt erzeugt und in die Musikobjektliste eingetragen. Das Musikobjekt ist der Container für das Agentensystem. Es enthält die Musikdaten, implementiert Funktionalitäten zu deren Vorverarbeitung und Aufbereitung, um daraus dann die Agenten zu instantiieren. Die Musikdaten setzen sich aus dreierlei zusammen.

MIDI-Daten: Wenn eine MIDI-Datei eingelesen wird, kommen zunächst alle Events in eine einzige Sequenz. Daraus wird im Zuge der Vorverarbeitung für jeden (nicht-leeren) Kanal eine einzelne Sequenz erzeugt und ein Agent instantiiert. Weiterhin werden NoteOn- und NoteOff-Events zu Note-Events zusammengefasst; das erleichtert später das Einrechnen der Artikulation. Die Agenten werden ihre Roh-MIDI-Sequenzen im weiteren nicht mehr verändern, denn sie sind das Ausgangsmaterial, aus dem expressive MIDI-Sequenzen gerendert werden. Sie sind sozusagen das Pendant zur Szenen-Geometrie in der Computergraphik.

Liste von Interpretationen: Parallel zur MIDI-Datei liefert eine sogenannte Performance-Datei eine Liste von Interpretationen. Sie beschreiben gemäß

⁴GRAME (2006), Seite 3

den in Kapitel 5 entwickelten Modellen, wie das Musikstück ausdrucksvoll dargeboten wird. Alle Interpretationsanweisungen sind in Maps organisiert (Tempomap, Rubatomap, Dynamicsmap, Articulationmap usw.). Diese können sowohl lokal für einen einzelnen Kanal definiert werden, als auch global für alle Kanäle (der Vollständigkeit halber: das gilt auch für den Artikulationsstil). Jeder Agent hat Zugriff auf seine lokalen und die globalen Maps. Gibt es sowohl eine lokale als auch eine globale Map vom selben Typ (z.B. eine lokale und eine globale Dynamicsmap), wird die lokale bevorzugt. Bei einer typischen Solist-mit-Begleitung-Situation würde das Orchester durch globale Maps abgedeckt (sie müssen nicht für jede Stimme extra definiert werden) und nur der Solist erhält seine eigenen lokalen Anweisungen. Lediglich im Falle des Artikulationsstiles wird ein wenig anders verfahren. Ist eine Artikulation nicht im lokalen Stil definiert, wird im globalen Stil danach gesucht. So kann ein für alle gemeinsamer Artikulationsstil definiert werden, der nur in einzelnen Fällen lokal abweicht.

Formgliederung: Die MIDI-Daten werden durch sog. Meta-Daten ergänzt. Sie beschreiben die formale Gliederung des Musikstücks. Dabei werden drei Gliederungsebenen unterschieden: Abschnittsgliederung, Satz- oder Phrasengliederung und Figurengliederung. Letzteres entspricht der typischen zeitlichen Ausdehnung von Motiven. Die Formeinheiten, also die Figuren, Phrasen und Abschnitte, werden im Format

(Beginn der Einheit in Ticks, Ende der Einheit in Ticks)

definiert, wobei die Musik nicht lückenlos durch Formeinheiten abgedeckt werden muss und Einheiten sich auch überlappen können (z.B. wenn der letzte Ton der vorigen Phrase gleichzeitig der erste der nächsten ist). Wie schon die Interpretationen, so kann auch die Formgliederung globale und lokale Elemente definieren. Es kann bspw. die Grobstruktur (Abschnittsgliederung) global für alle gleich sein, aber die Phrasen- und Figurengliederung polyphon versetzt (in jeder Stimme anders). Und ebenso wie bei den Interpretationen, gilt auch hier, lokale Informationen dominieren über globale. Gibt es für einen Kanal eine lokale Figurengliederung, ignoriert er eine evtl. vorhandene globale. Die Informationen zur Formgliederung des Musikstücks werden derzeit manuell vom Komponisten oder Bearbeiter eingegeben. Prinzipiell ist es aber auch denkbar, dass sie von einem Verfahren zur Strukturanalyse und Musiksegmentierung ermittelt werden (Rohrmeier 2007, de Haas, Rohrmeier, Veltkamp & Wiering 2009).

Im Zuge der Vorverarbeitung werden die Daten so aufbereitet, dass sie im weiteren schneller verarbeitet werden können. Wenn bspw. in der Dynamicsmap

Strings als Anweisungen verwendet wurden (*mf*, *pp*, *ff*, etc.), werden sie nun durch die im Header assoziierten numerischen Werte ersetzt, so dass dieser Schritt bei allen späteren Verarbeitungen nicht mehr nötig ist. Ferner werden die Daten zu den Interpretationen und zur Formgliederung auf Konsistenz überprüft. Leere Maps werden gelöscht. Auch die Vorberechnungen für die impliziten Tempokurven werden jetzt ausgeführt (siehe Abschnitt 5.3.1). Sämtliche Controller-Events werden aus den MIDI-Sequenzen gelöscht, da sie potentiell in Konflikt mit denjenigen Controllern stehen, die beim Rendern ausdrucksvoller Sequenzen erzeugt werden. Auch MIDI-Tempo-Events werden gelöscht; sie spielen bei der Wiedergabe innerhalb der Musik-Engine keine Rolle. Für jeden Kanal, den die MIDI-Datei bedient⁵, wird ein Agent erstellt und die Verlinkungsstruktur zu seinen lokalen und globalen Daten hergestellt.

Damit sind alle notwendigen Informationen gegeben, um ausdrucksvolle MIDI-Sequenzen zu rendern. Dieser Schritt geht der eigentlichen Musikwiedergabe immer voraus. Die Funktionsweise der Agenten wird in Abbildung 6.2 schematisch zusammengefasst. Zunächst erzeugt jeder Agent eine Kopie seiner originalen Sequenz (sie selbst bleibt unverändert). In dieser werden alle Transformationen vorgenommen, wie in Kapitel 5 beschrieben, mit einer Ausnahme. Das Timing wird in Echtzeit zur Wiedergabe umgesetzt. D.h., wann der nächste Event fällig ist, wird berechnet, nachdem der vorherige abgearbeitet ist. Die ausdrucksvolle MIDI-Sequenz weist folglich immer noch die ursprüngliche metrische Ordnung der originalen Sequenz (symbolische Zeit) auf.

In den Interpretationen können weiterhin für jeden Kanal ein Mute- und ein Loop-Flag gesetzt werden. Ist das Mute-Flag gesetzt, soll der Kanal in der Interpretation schweigen. Der zugehörige Agent wird folglich gar nicht erst aktiviert und rendert auch keine ausdrucksvolle MIDI-Sequenz. An alle anderen Agenten (für die das Mute-Flag nicht gesetzt ist) ergeht der Aufruf `play`. Ist für einen von ihnen das Loop-Flag gesetzt, wird er seine Sequenz in Endlosschleife spielen. Ist es nicht gesetzt, wird die Wiedergabe beendet, wenn der `EndTrack`-Event erreicht ist.

Hat der Agent seine ausdrucksvolle MIDI-Sequenz erstellt, berechnet er das Fälligkeitsdatum (Millisekunden-Datum) des ersten zu spielenden Events und übergibt es an den ihm zugewiesenen Echtzeit-Thread. Alle Agenten besitzen ein Bezugsdatum (der Startzeitpunkt der Wiedergabe), das benutzt wird, um ihr Zusammenspiel zu synchronisieren und numerisch bedingte Timing-Drifts zu kompensieren. Durch Rundungsfehler bei den Millisekunden-Berechnungen sind Kanäle mit einer hohen Event-Dichte tendenziell langsamer als solche mit nur wenigen Events. Für jede Tempoanweisung in den Tempomaps

⁵Zwar definiert der MIDI-Standard nur 16 Kanäle für einen Port, dafür können aber bis zu 256 Ports angesteuert werden, mehr als genug auch für größere Projekte.

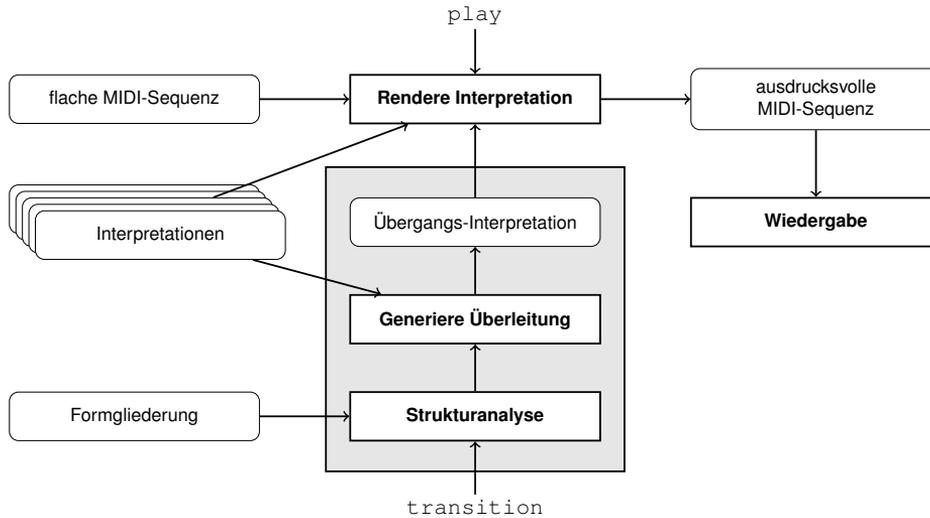


Abbildung 6.2: Die Funktionsweise eines Agenten beim Start der Wiedergabe und bei Überleitungen.

wurde in der Vorverarbeitung ihr Millisekunden-Datum berechnet. Diese werden bei der Wiedergabe mit dem Bezugsdatum addiert. Alle weiteren Timing-Berechnungen geschehen ausgehend von diesen festen Eckpunkten, die bei einer globalen Tempomap für alle Agenten gleich sind. Die Modelle sind so implementiert, dass für diese Eckpunkte Synchronität garantiert werden kann (siehe Abschnitt 5.3.1).

Der Echtzeit-Scheduler ruft den Agenten zum Zeitpunkt der Fälligkeit seines nächsten Events wieder auf, welcher dann die Event-Verarbeitung ausführt. Das muss nicht zwangsläufig zur Ausgabe des Events führen. Diejenigen Events, die von den Ausgabegeräten typischerweise ignoriert werden (wie Tempo, Lyrics und Marker), werden gar nicht erst verschickt. Auch Note-Events können nicht direkt ausgegeben werden, da sie nicht Teil des MIDI-Standards sind. An deren statt wird ein NoteOn-Event generiert und verschickt, sowie der zugehörige NoteOff-Event erzeugt und in die Sequenz eingefügt. Wenn er im weiteren Verlauf fällig wird, wird er verschickt und aus der Sequenz wieder entfernt. Ändert sich die Artikulation der Note, etwa im Rahmen eines Interpretationswechsels, wird der NoteOff-Event neu in der Sequenz platziert.

Diese der eigentlichen Ausgabe vorgelagerte Event-Verarbeitung wird ferner für die Umsetzung unterschiedlicher *Wiedergabemodi* genutzt. Verschiedene Samp-

ler und Synthesizer werten MIDI-Events oft ganz unterschiedlich aus. Sie sollten für eine saubere Ausgabe auf bestimmten Geräten erst entsprechend konvertiert werden. Im Falle des Samplers *Vienna Instruments* werden bspw. vor jedem Ton zwei Controller gesendet, durch die das für die geforderte Artikulation passende Sample-Set angesteuert wird. Wer mit *Vienna Instruments* bereits Musik produziert hat, weiß, dass die Ansteuerung der Performance-Matrizen gerade bei häufig wechselnden Artikulationen manuell sehr mühselig sein kann aber das Klangergebnis entscheidend prägt. In der Musik-Engine ist diese Arbeit durch einen spezialisierten Wiedergabemodus automatisiert. Jeder Agent kann unabhängig von den anderen in seinem eigenen Wiedergabemodus laufen. So können die verschiedensten Ausgabegeräte gleichzeitig bedient werden, ohne dass die originalen MIDI-Daten vom Anwender extra vorbereitet werden müssen.

Der Player kann die laufende Wiedergabe beenden (Funktionsaufruf `stop`). Dabei werden alle Echtzeit-Threads beendet und alle spielenden Agenten aufgerufen, damit sie an ihren jeweiligen Kanal einen `AllNotesOff`-Event (Controller Nr. 123) senden, der alle klingenden Töne beendet. Mit dem Befehl `transition` kann der Player die Agenten aber auch anweisen, zu einer anderen Interpretation überzuleiten. Dabei werden die Parameter `perfStyle` und `maxLatency`, die der Befehl `transition` in der API entgegennimmt an die Agenten weitergegeben.

6.2 Überleitung

Wenn die Interpretation geändert werden soll, wird die Funktion

```
long transition(std::string perfStyle, long maxLatency);
```

eines jeden betroffenen Agenten aufgerufen.⁶ Betroffen sind diejenigen Agenten, die in der Ausgangs- oder Zielinterpretation zu spielen haben. Die momentan spielenden Agenten setzen die Wiedergabe ihrer derzeitigen ausdrucksvollen MIDI-Sequenz aber zunächst ohne Unterbrechung fort. Währenddessen analysiert jeder Agent den musikalischen Kontext seiner derzeitigen Wiedergabeposition (Wo sind die nächsten Strukturgrenzen?). Er erstellt daraus einen Plan für die Überleitung. Auf Grundlage dieses Planes wird eine neue Interpretation, die Übergangsinterpretation, erstellt und in eine ausdrucksvolle MIDI-Sequenz gerendert. Jetzt erst wird die Wiedergabe unterbrochen und auf die neue MIDI-Sequenz umgeschaltet. Da klingende Töne dabei nicht unterbrochen werden und das Umschalten der MIDI-Sequenz sehr schnell vonstatten geht,

⁶Der Zeiger auf das Musikobjekt, den der Befehl noch in der API hat, ist hier nicht mehr nötig.

wird für den Hörer keine Unterbrechung oder Verzögerung wahrnehmbar. Die Musik geht nahtlos weiter.

6.2.1 Strukturanalyse

Ziel der Strukturanalyse ist es, innerhalb der durch maxLatency vorgeschriebenen Latenz den am besten geeigneten Zielzeitpunkt für die Überleitung zu ermitteln. An diesem Punkt soll die Überleitung beendet und die Zielinterpretation erreicht sein. Auf diesen Punkt sollen alle Änderungen, die im Zuge der Überleitung an der Ausgangsinterpretation vorgenommen werden, hinspielen. Dieser Punkt muss die stärkste Strukturgrenze im Latenzintervall sein. Je stärker diese Grenze ist, desto plausibler wird der Wandel der Interpretation erscheinen und desto weniger Aufwand ist später nötig, um diesen Wandel vorzubereiten.

Die stärksten Strukturgrenzen werden durch die Abschnittsgliederung gesetzt. Ein neuer Abschnitt geht meist mit einem neuen kompositorischen Gedanken und neuem musikalischen Material (sei es auf der Ebene der Melodik, Harmonik, Rhythmik oder auch der instrumentalen Klangfarbe) einher und differenziert sich so deutlich vom vorhergehenden Abschnitt. Kaum ein Musiker lässt sich die Gelegenheit entgehen, unterschiedliche Abschnitte auch interpretatorisch zu kontrastieren. Hier kann auch der virtuelle Interpret, der Agent, die Interpretation direkt und ohne vorbereitende Adaptionen im Vorfeld umschalten.

Auf die Abschnittsgrenzen folgen mit absteigender Wertigkeit Phrasen- und Figurengrenzen. Phrasen (Sätze und Halbsätze) und Figuren sind zumeist Teil eines über sie selbst hinaus reichenden Gesamtgebildes, das eher durch Homogenität geprägt wird denn durch Kontraste. Je stärker die Homogenität ausgeprägt ist desto schwächer wird die Strukturgrenze gewertet. Auch die Aussage »Hier beginnt etwas neues« trifft immer weniger zu. Änderungen in der Interpretation müssen infolgedessen besser vorbereitet werden, denn sie können durch den musikalischen Kontext immer weniger gerechtfertigt werden.

Deckt das Latenzintervall mehrere gleichwertige Strukturgrenzen ab, wird die erste von ihnen ausgewählt. Alle folgenden würden das Latenzintervall unnötig ausreizen. Für jedes Formelement sind Anfangs- und Endzeitpunkt definiert. Endpunkte sind nur für diejenigen Agenten interessant, die in der Zielinterpretation nicht mehr spielen sollen, die also in ihrem Spiel »zu einem Ende kommen müssen«. Sie wählen folglich das stärkste Ende innerhalb des Latenzintervalls. Alle anderen – also Agenten, die in der Zielinterpretation weiterspielen oder neu einsetzen – wählen den stärksten Anfangspunkt als Zielpunkt der Überleitung bzw. als Einsatzpunkt.

Typ des Zielpunktes	Wertigkeit
Abschnittsgrenze	5
Phrasengrenze	4
Figurengrenze	3
Taktstrich (der erste Schlag eines Taktes)	2
Beat (ein Schlag innerhalb eines Taktes)	1
eine Position zwischen den Beats	0

Tabelle 6.1: Abhängig von der Wertigkeit ihres Zielpunktes wird die Überleitung unterschiedlich umgesetzt.

Es kann jedoch auch vorkommen, dass innerhalb des Latenzintervalls keine Strukturgrenze liegt. maxLatency könnte z.B. sehr klein gewählt worden sein. Die Formgliederung könnte sehr grobmaschig sein oder schlicht nicht vorliegen. Trotzdem liegt der Musik auch in diesem Fall immer noch ein allgemeines Ordnungsprinzip zu Grunde, die metrische Ordnung. Auf dieser Basis kann weiter nach Zielpunkten für die Überleitung gesucht werden. Hier wären die Gliederung in Takte und Taktschläge zu nennen. Es sind aber bereits sehr schwache Zielpunkte. Die Überleitung sollte hier so ausführlich wie möglich sein, damit der Wandel der Interpretation nicht zu abrupt geschieht. Deshalb wird der letzte Taktstrich (die letzte Eins) im Latenzintervall gewählt. Wenn keine Taktgrenzen innerhalb des Latenzintervalls liegen, wird der letzte Taktschlag ausgewählt.

Wenn schließlich nicht einmal Taktartinformationen definiert sind, wird das Latenzintervall voll ausgeschöpft, der Zielpunkt auf das Ende der Latenz gelegt. Dies ist der am schwächsten zu bewertende Zielpunkt, denn ihm liegen überhaupt keine Informationen zur kompositorischen oder metrischen Ordnung des Musikstücks zu Grunde.

Die momentane Wiedergabeposition sowie der Zielpunkt der Überleitung und seine Wertigkeit (noch einmal zusammengefasst in Tabelle 6.1) bilden den *Überleitungsplan*. Auf dessen Grundlage wird die eigentliche Überleitung erzeugt.

6.2.2 Generiere Überleitung

Die Veränderungen im ausdrucksvollen Spiel werden nicht direkt in der laufenden MIDI-Sequenz vorgenommen. Statt dessen wird eine *Übergangsinterpretation* erzeugt. Das ist eine vollwertige Interpretation, aus der wieder eine expressive MIDI-Sequenz gerendert wird. Sie importiert die Inhalte aller Maps der Aus-

gangsinterpretation bis zum Zielpunkt der Überleitung. Von da an werden die Anweisungen aus den Maps der Zielinterpretation importiert. Die Detailarbeit der eigentlichen Überleitung reduziert sich auf die Umgebung um die Nahtstelle.

Damit jeder Agent seine eigene Überleitung erzeugen kann, werden alle globalen Maps lokal gemacht, außer die Tempomap. Bei einer globalen Tempomap spielen die Agenten synchron (zumindest soweit es das Makro-Timing betrifft). Würde jeder seine eigene Tempoüberleitung erzeugen, entstünde Asynchronie. Um das zu vermeiden, muss die Tempomap auch weiterhin global bleiben. Deren Überleitung wird von allen Agenten gemeinsam berechnet. Aber auch die anderen Maps, die jeder Agent für sich überleitet, können nicht alle gleich behandelt werden, schließlich beschreiben sie sehr unterschiedliche Erscheinungen im ausdrucksvollen Spiel. Generell gilt das Prinzip: Makro-Merkmale (Tempo und Makro-Dynamik) werden übergeleitet. Mikro-Merkmale (also die kleinsten interpretatorischen Sinneinheiten) werden an einer jeweils geeigneten Stelle umgeschaltet, denn sie geschehen innerhalb des Zeitfensters, in dem auch der Mensch eine getroffene Entscheidung und ein angefangenes motorisches Programm nicht mehr (sinnvoll) revidieren kann (Illert 1994, Hähnel 2008). Die folgenden Abschnitte erläutern die Verfahren im einzelnen.

Überleiten der Mikro-Merkmalen

Artikulation: Die Artikulation schaltet direkt am Zielpunkt der Überleitung um. Zunächst werden alle Artikulationsstile aus Ausgangs- und Zielinterpretation in die Stilmenge der Übergangsinterpretation kopiert.

$$AStyleSet_{\text{Übergang}} = AStyleSet_{\text{Ausgang}} \cup AStyleSet_{\text{Ziel}}$$

Ihre Indizes werden entsprechend angepasst. Die Implementierung arbeitet nicht mit Indizes sondern mit aussagekräftigen Strings. Statt der Indizes werden diese also entsprechend geändert, so dass keine zwei Artikulationsstile die gleiche Bezeichnung führen.

Aus der Articulationmap der Ausgangsinterpretation werden alle Einträge, die vor dem Zielpunkt der Überleitung stehen, in die Articulationmap der Übergangsinterpretation kopiert. An diesem Punkt wird eine *switch*-Anweisung eingefügt. Sie schaltet auf Artikulationsstil und Grundartikulation der Zielinterpretation um. Danach werden alle Artikulations- und *switch*-Anweisungen der Zielinterpretation ab dem Zieldatum angehängt. Die *switch*-Anweisungen werden schließlich noch an die geänderten Indizes/Strings angepasst.

Asynchronie und rhythmische Ungenauigkeit: Auch die Mikro-Timing-Merkmale Asynchronie und rhythmische Ungenauigkeit werden am Zielpunkt der Überleitung direkt umgeschaltet. Wenn in der Zielinterpretation an dieser Stelle keine Anweisungen definiert sind, werden entsprechende generiert, die den Wechsel umsetzen.

Hier wären auch allmähliche Übergänge denkbar, etwa indem mehrere abgestufte Anweisungen in die Maps eingefügt werden. Doch können auch beim Menschen plötzliche Änderungen beobachtet werden. Es kann sein, dass nur ein einzelner schwer zu spielender Ton ungenau ist und der Rest sehr genau. Der Aufwand und Effekt, den eine überleitende Interpolation mit sich bringt, lässt sich aus praktischer Sicht also nicht unbedingt rechtfertigen.

Schwerpunktbetonung: Die Schwerpunktbetonung wird ebenfalls direkt am Zielpunkt der Überleitung auf die Zielinterpretation umgeschaltet. D.h., eine entsprechende Betonungsanweisung wird an dem Punkt in die Emphasismap der Übergangsinterpretation eingefügt.

Eine überleitende Interpolation von Betonungsmustern findet nicht statt, denn sie kann zu unerwünschten Ergebnissen führen. Hierbei können ausdrucksmäßig flache und sogar falsche Betonungsverhältnisse entstehen, die weder in der Ausgangs- noch in der Zielinterpretation vorkommen. Zu überlegen bleibt aber, ob ein angefangenes Betonungsmuster erst noch zu Ende durchgeführt werden sollte. Liegt das Ziel der Überleitung auf einer Taktgrenze, ist das von vornherein gegeben. Ist das nicht der Fall, kann von einer Auftakt-Situation ausgegangen werden. Es kann sich bspw. um eine Figur handeln, die auftaktig beginnt. Wenn die Schwerpunktbetonung dies widerspiegeln soll, muss der Wechsel auch an der entsprechenden Stelle im Betonungsmuster geschehen, also nicht zwangsläufig auf einem Taktstrich.

Rubato: Die Änderung des Rubato kann nicht bedenkenlos auf dem Zieldatum der Überleitung platziert werden. Wenn dieses in der Zielinterpretation nämlich mitten in einem Rubato Frame liegt, würden die dann folgenden Rubato-Verzerrungen nach dem Ende der Überleitung phasenverschoben sein. Also muss der Wechsel zum Rubato der Zielinterpretation dort geschehen, wo in der Zielinterpretation ein Rubato Frame beginnt. Das kann auch hinter dem Zielpunkt der Überleitung liegen, solange die Latenzvorgaben eingehalten werden.

Es kann jedoch auch dieser Punkt problematisch sein. Liegt er mitten in einem Rubato Frame der Ausgangsinterpretation, kann die dort angefangene Timing-Verzerrung nun nicht mehr vollständig durchgeführt werden. Es entsteht eine Diskontinuität in der Timing-Kurve. Verzögerte Events

würden vorzeitig fällig und im Zweifelsfall alle auf einmal abgearbeitet. Vorgezogene Events müssten hingegen noch einmal abgearbeitet werden. Um das zu vermeiden, wird der betroffene Rubato Frame in der Ausgangsinterpretation ohne Verzerrung gespielt ($i_m = 1$).

Wurde `maxLatency` sehr klein festgesetzt, kann es sein, dass kein sauberer Rubato-Übergang möglich ist. Interpretationen, die für einen solchen Anwendungskontext geschaffen werden, sollten folglich auf allzu stark ausgeprägte Rubati verzichten oder sich auf entsprechend kurze Rubato Frames beschränken.

Von diesen Verfahren ausgenommen sind diejenigen Agenten, die neu einsetzen. Sie fangen direkt im Stil der Zielinterpretation an. Die Verfahren müssen auch nicht für Agenten angewendet werden, welche ihr Spiel beenden, denn sie spielen über den Zielpunkt der Überleitung nicht hinaus.

Überleiten der Makro-Dynamik

Bei der Berechnung der Dynamiküberleitung wird zwischen dem *Neueinsatz* eines schweigenden Agenten, dem *Beenden* eines spielenden Agenten und dem *Überleiten* eines weiterspielenden Agenten unterschieden. Die unterschiedlichen Vorgehensweisen hierfür werden in Tabelle 6.2 zusammengefasst und nachfolgend im einzelnen beschrieben.

Am einfachsten gestaltet sich der Neueinsatz von schweigenden Agenten, die in der Zielinterpretation spielen sollen. Hat die Strukturanalyse einen starken Zielpunkt für die Überleitung gefunden, also den Anfang eines Abschnitts, einer Phrase oder einer Figur, wird der Agent direkt dort einsetzen. Dazu sucht er in der Zielinterpretation die letzte Dynamikanweisung vor oder auf dem Zieldatum und berechnet daraus die am Zieldatum geltende Lautstärke. Das ist die Einstiegslautstärke. Alle weiteren Parameter werden von der Dynamikanweisung übernommen. Daraus wird die Einstiegsanweisung erzeugt und in die Dynamicsmap der Übergangsinterpretation eingefügt. Die komplette Dynamik davor wird auf 0 gesetzt. Alle darauf folgenden Anweisungen werden aus der Zielinterpretation kopiert.

Wurden nur schwache Zielpunkte für den Einstieg gefunden (Wertigkeiten kleiner als 3), wird eine Einblende im Vorfeld des Zielpunktes, genauer: im Intervall

[momentane Wiedergabeposition, Zielpunkt der Überleitung]

ausgeführt. Je schwächer der Zielpunkt ist, desto neutraler wird die Einblende gehalten, wird symmetrischer und linearer in ihrem Verlauf. Dieses Prinzip

Wertigkeit des Zielpunktes	Neueinsatz eines schweigenden Agenten	Beenden eines spielenden Agenten	Überleiten eines weiterspielenden Agenten
5 Abschnittsgrenze	Direkteinsatz am Zielpunkt	direktes Ende am Zielpunkt	direkt umschalten am Zielpunkt
4 Phrasengrenze	Direkteinsatz am Zielpunkt	Ende am Zielpunkt mit auf 5/6 reduzierter Dynamik $c = 0.3, p = 0.75$	kontextabhängiger Dynamikübergang $c = 0.3, p = 0.75$
3 Figurengrenze	Direkteinsatz am Zielpunkt	Ende am Zielpunkt mit auf 2/3 reduzierter Dynamik $c = 0.3, p = 0.5$	kontextabhängiger Dynamikübergang $c = 0.3, p = 0.5$
2 Taktgrenze	Einblende $c = 0.3, p = 0.25$	Ausblende $c = 0.3, p = 0$	stetiger Dynamikübergang $c = 0.3, p = 0$
1 Beat	Einblende $c = 0.3, p = 0$	Ausblende $c = 0.2, p = 0$	stetiger Dynamikübergang $c = 0.2, p = 0$
0 Ende des Latenzintervalls	Einblende $c = 0.15, p = 0$	Ausblende $c = 0.125, p = 0$	stetiger Dynamikübergang $c = 0.15, p = 0$

Tabelle 6.2: Zusammenfassung der unterschiedlichen Makro-Dynamik-Überleitungen in Abhängigkeit von der Wertigkeit des Überleitungszielpunktes und der Art der Überleitung. Die Einstellungen von *Curvature* c und *Protraction* p des jeweiligen Dynamikübergangs sind aus den Beobachtungen in Abbildung 5.8 abgeleitet.

konnte bei den Analysen in Kapitel 5 (siehe Abbildung 5.8) beobachtet werden. Je neutraler der musikalische Kontext war (im Extremfall waren es nur Tonrepetitionen, wie in Anhang D), desto neutraler waren auch die von den menschlichen Interpreten eingespielten Dynamikverläufe. Das Prinzip findet auch beim Beenden und beim regulären Überleiten Anwendung.

Wurden zum Beenden eines noch spielenden Agenten nur schwache Zielpunkte gefunden, wird vom momentanen Wiedergabedatum aus eine vollständige Ausblende (*decrescendo al niente*) bis zum Zielpunkt durchgeführt. Für stärkere Zielpunkte ist das hingegen weder nötig noch angebracht. Hier ist es sogar wünschenswert, zu hören, dass der Agent eine Figur, eine Phrase oder einen Abschnitt noch zu Ende spielt und dann erst aufhört. Das verdeutlicht die Verbindung von musikalischer Form und Interpretation, aus welcher der Eindruck von interpretatorischer Absicht erwächst. Am Ende eines Abschnitts wird also direkt ausgestiegen. Das Ende einer Phrase oder Figur muss hingegen nicht so zwingend erscheinen. Hier kann es hilfreich sein, wenn der Agent zum Ende hin *abphrasiert*. D.h., er nimmt sich dynamisch ein wenig zurück, bei Figuren- etwas mehr als bei Phrasenenden, umgesetzt durch ein *Decrescendo* im Intervall

[momentane Wiedergabeposition, Zielpunkt der Überleitung].

Ein Agent, der sowohl in der Ausgangs- als auch in der Zielinterpretation zu spielen hat, muss seine laufende Dynamicsmap in die Dynamicsmap der Zielinterpretation überleiten. Am einfachsten umzusetzen ist wieder der Fall, dass eine Abschnittsgrenze als Zielpunkt der Überleitung gefunden wurde. Hier kann der Agent direkt auf die Zieldynamicsmap umschalten; die Abschnittsgrenze rechtfertigt den direkten Wechsel. Sofern die Zielinterpretation keine Anweisung an diesem Punkt definiert, wird eine generiert, welche die Dynamik hier aufnimmt. Alles folgende kann direkt aus der Zielinterpretation kopiert werden. Alles vorausgehende wird aus der Ausgangsinterpretation kopiert.

Bei Phrasen- und Figurengrenzen wird ein *kontextabhängiger Dynamikübergang* durchgeführt. Er orientiert sich stärker am dynamischen Geschehen in Ausgangs- und Zielinterpretation und nur grob am Zielpunkt der Überleitung. Soweit möglich, werden keine zusätzlichen Anweisungen erzeugt, sondern die aus der Zusammenführung von Ausgangs- und Zielinterpretation bestehenden editiert. Die Topologie der Dynamicsmaps – das Grundkonzept der dynamischen Gestaltung – bleibt dadurch erhalten. Zunächst wird die erste Dynamikanweisung, welche die Zielinterpretation im Intervall

[Zielpunkt der Überleitung, Ende des Latenzintervalls]

definiert, gesucht. Wird keine gefunden, wird eine auf dem Zielpunkt erzeugt. Das wird die Einstiegsanweisung D_{m+1} in die Zielinterpretation. Sie muss nicht exakt auf dem Zielpunkt liegen.

Als nächstes wird in der Ausgangsinterpretation die letzte Anweisung vor dem Zeitpunkt der Einstiegsanweisung gesucht. Liegt sie vor der momentanen Wiedergabeposition, wird auf der Wiedergabeposition eine Anweisung erzeugt. Das ist die Übergangsanweisung D_m . Der Zielpunkt der Überleitung wird von den beiden Anweisungen, die in der Übergangsinterpretation direkt aufeinander folgen, umschlossen.

Wiedergabeposition $\leq d_m <$ Zielpunkt der Überleitung $\leq d_{m+1} \leq$ Latenzende

Die Ziellautstärke der Übergangsanweisung $v_{2,m}$ muss nun so festgesetzt werden, dass der Anschluss an die Einstiegsanweisung hergestellt wird. Das muss nicht zwangsläufig ein glatter Übergang sein ($v_{2,m} = v_{1,m+1}$). Wenn in der Zielinterpretation an dieser Stelle eine Dynamikstufe definiert ist, kann diese auch von der Überleitung wieder aufgegriffen werden. Bei einer stufenweisen, d.h. diskreten, Lautstärkeänderung stehen als Ziellautstärke der Überleitung tatsächlich zwei Möglichkeiten zur Auswahl, die obere und die untere Ecke der Stufe. Eine von beiden ist durch den Wert $v_{1,m+1}$ gegeben. Die andere ist gegeben durch die Ziellautstärke $v_{2,m}^*$ derjenigen Dynamikanweisung, die in der Zielinterpretation vor der Einstiegsanweisung liegt; sie sei bezeichnet mit D_m^* . Auf einen der beiden v -Werte muss die Übergangsanweisung »umgespannt« werden. Dem Effizienz-Kriterium (so wenig Änderung wie nötig) folgend, wird derjenige Wert ausgewählt, welcher näher am ursprünglichen Wert von $v_{2,m}$ liegt.

$$v_{2,m} := \begin{cases} v_{2,m}^* & : |v_{2,m}^* - v_{2,m}| < |v_{1,m+1} - v_{2,m}| \\ v_{1,m+1} & : \text{sonst} \end{cases}$$

Wenn nur schwache Zeilpunkte für die Überleitung gefunden wurden (Wertigkeiten unter 3) wird ein *stetiger Dynamikübergang* ausgeführt. Der gleicht dem kontextabhängigen Dynamikübergang, wird aber immer auf $v_{1,m+1}$ blenden und lässt keine Stufendynamik zu.

Tempoüberleitung

Gibt es eine globale Tempomap, unterliegen ihr alle Agenten, sofern für sie keine lokale Tempomap definiert ist. Das bedeutet auch, dass alle diese Agenten die gleiche Tempoüberleitung ausführen müssen, damit sie auch weiterhin synchron zueinander bleiben (zumindest was das Makro-Timing angeht). Um dies zu erreichen, müssen die Agenten gemeinsam eine Tempoüberleitung erzeugen. Dazu werden die Überleitungspläne der Agenten zu einem globalen Plan vereinigt. Damit die Tempoüberleitung für alle Agenten plausibel ist, richtet sich der globale Plan nach dem schwächsten Glied, d.h. nach dem Agenten, der am längsten für die Überleitung braucht, und nach der geringsten Wertigkeit.

Wertigkeit des Zielpunktes	Art des Tempoüberganges	Einstellung Parameter i_m	
		Ritardando	Accelerando
5 Abschnittsgrenze	direkt umschalten am Zielpunkt	/	/
4 Phrasengrenze	kontextabhängiger Tempoübergang	0.65	0.35
3 Figurengrenze	kontextabhängiger Tempoübergang	0.55	0.45
2 Taktgrenze	stetiger Tempoübergang	0.52	0.48
1 Beat	stetiger Tempoübergang	0.51	0.49
0 Ende des Latenzintervalls	stetiger Tempoübergang	0.5	0.5

Tabelle 6.3: Eine Zusammenfassung der unterschiedlichen Tempoüberleitungen in Abhängigkeit von der Wertigkeit des Überleitungszielpunktes. Der Parameter i_m steuert die Charakteristik des Tempoüberganges.

Dazu erzeugt der erste Agent, der die Strukturanalyse abgeschlossen hat, einen globalen Überleitungsplan. Dieser enthält die momentane Wiedergabeposition, den Zielpunkt der Überleitung und dessen Wertigkeit. Jeder weitere Agent vergleicht die beiden letzteren Werte mit dem Ergebnis seiner eigenen Strukturanalyse. Liegt sein Zielpunkt zeitlich weiter hinten, überschreibt er den bisherigen Wert. Hat sein Zielpunkt eine geringere Wertigkeit, überschreibt er auch diesen Wert im globalen Plan.

Sind alle Agenten durch, wird einer von ihnen zufällig ausgewählt. Er erzeugt die globale Tempoüberleitung. Das Verfahren entspricht dem der Dynamiküberleitung beim Überleiten eines weiterspielenden Agenten. An Abschnittsgrenzen (Wertigkeit 5) wird das Tempo direkt auf die Zielinterpretation umgeschaltet. Bei Wertigkeiten von 4 oder 3 wird eine *kontextabhängige Tempoüberleitung* durchgeführt, die auch Tempostufen erlaubt. Bei Wertigkeiten unter 3 wird eine *stetige Tempoüberleitung* durchgeführt. Der Wert des Parameters i_m der Übergangsangweisung T_m wird umso neutraler gewählt, je schwächer die Wertigkeit des Zielpunktes ist. Tabelle 6.3 fasst dies noch einmal zusammen.

6.3 Ergänzungen

Es wurde ein Agenten-basierter Ansatz beschrieben für eine Musik-Engine, die ausdrucksvolle Musik rendert, abspielt und verschiedene Interpretationen zur Wiedergabe automatisch ineinander überleiten kann. Die bisherigen Ausführungen fokussierten dabei auf den Aspekt des Überleitens. An dieser Stelle sollen nun einige weitere Perspektiven und Möglichkeiten vorgestellt werden, die der Lösungsansatz darüber hinaus mit sich bringt.

Melodische und harmonische Varianz: Es ist möglich, eine Artikulation zu definieren, die den Ton, auf den sie angewendet wird, stumm schaltet. Der Modifikator hätte die Form (*set velocity 0*). Dies sei als *Mute-Artikulation* bezeichnet. Ferner ist es möglich, dass für eine Stimme (einen MIDI-Kanal/Agenten) mehrere zeitgleich erklingende Töne definiert werden.⁷ Das seien alternative Töne, von denen je Interpretation nur einer gespielt wird. Alle anderen werden durch die Mute-Artikulation ausgeschaltet.

Verschiedene Interpretationen können damit unterschiedliches Tonmaterial darbieten. Eine Interpretation mag ausschließlich Dur-Terzen spielen, eine andere ersetzt sie durch Moll-Terzen. Eine Interpretation enthält Verzierungen (z.B. Triller, Durchgänge, Rhythmisierungen oder sogar freiere improvisatorische Figuren), eine andere spielt die Musik schlicht.

Varianz in Wiederholungen: Die melodische und harmonische Varianz kann innerhalb von Interpretationen dazu verwendet werden, Endlosschleifen abwechslungsreicher zu gestalten. Bevor ein Agent in die nächste Wiederholung geht, kann er die Interpretation neu rendern. Dabei können Töne durch Alternativen ersetzt werden und auch andere Parameter mehr oder weniger frei umgesetzt werden. Ein Crescendo kann bspw. nicht ganz so laut aufgezogen werden wie beim letzten mal. Die Schwerpunktbetonung könnte dafür stärker ausgeprägt sein (größerer dynamischer Radius $scale_m$). Prinzipiell lassen sich solche zufälligen Einflüsse auf alle Parameter aufrechnen, so dass der Hörer niemals exakt die gleiche Interpretation mehrmals hintereinander erlebt (siehe dazu auch die Diskussion in Abschnitt 5.6).

Schlussbildung: Es muss bei der Vertonung interaktiver Medien nicht ständig Musik spielen. Auch Ruhephasen sind ein dramaturgisch wichtiges Element. Um in eine Ruhephase überzugehen, muss die laufende Musik beendet werden. Das wird heutzutage bestenfalls durch eine Ausblende umgesetzt; schlechtestenfalls wird die Musik einfach ausgeschaltet, ein harter

⁷Bei Klaviermusik ist das sogar üblich.

Schnitt ins Nichts. Mit dem folgenden Verfahren kann auch durch interpretatorische Mittel eine Schlussbildung herbeigeführt werden.

Es kann eine Interpretation definiert werden, die ein sehr langsames Tempo vorschreibt, sowie eine »zurückhaltende« Dynamik, und für alle Kanäle das Mute-Flag (siehe Abschnitt 6.1 unter Musikobjekte) setzt. Eine Überleitung in diese Interpretation bewirkt, dass jeder Agent seine Wiedergabe beendet, dabei langsamer und leiser wird, quasi ein finales Ritardando ausführt. Durch die Strukturanalyse wird dafür gesorgt, dass der angefangene musikalische Gedanke noch zu Ende gespielt wird, so weit die Latenzvorgaben es erlauben. Lediglich im Falle von schwachen Zielpunkten würde eine komplette Ausblende vorgenommen.

Das Verfahren wird nicht in jeder Situation den Gestus des Beendens plausibel nachbilden. An Abschnittsgrenzen wird weder ritardiert noch decrescendiert; hier wirkt das Ende meist sehr abrupt. Wenn die musikalische Form sehr polyphon ist, beendet jeder Agent an einer anderen Stelle; der Schluss »zerfasert« und wirkt verunglückt. Die Schlussbildung wirkt überzeugender, wenn alle Agenten das gleiche machen, gemeinsam am Schluss ankommen und den Schlusston gleichzeitig beenden. Ist das gegeben, kann der Überleitungsmechanismus auch zur Schlussbildung verwendet werden.

Ensemblespiel: Bislang ist die Funktionsweise des Agentensystems relativ einfach gehalten. Jeder Agent steht für sich. Kommunikation nach außen findet nur bei der Tempoüberleitung statt und selbst hier nur mit dem globalen Kontext, nicht direkt mit anderen Agenten. Gerade dies birgt aber noch großes Potential für die Simulation des Zusammenspiels der Musiker.

Jeder Agent führt sein eigenes Bezugsdatum für die Timing-Berechnungen (siehe Abschnitt 6.1 unter Musikobjekte). Wird dieses durch innere oder äußere Einflüsse verändert, gerät der Agent in Asynchronie zum Rest des Ensembles. Wenn andere Agenten dies bemerken, können sie ihr eigenes Bezugsdatum allmählich annähern, bis wieder ein synchroner Zustand erreicht ist. Auf diese Art und Weise lässt sich direkte Interaktion mit den Musikern simulieren, wie in folgendem Beispiel: Der Spieler lauscht einem Ensemble, lenkt die Konzentration eines der Musiker ab, der Rest des Ensembles reagiert.

6.4 Diskussion

Zu Demonstrations- und Testzwecken entstand eine Reihe von Musikstücken mit unterschiedlichen Interpretationen. Anhand derer konnte das entwickelte

Verfahren zur Überleitung von Interpretationen systematisch ausgetestet werden. Im Fokus waren sowohl die Überleitungen an sich, als auch die musikalische Vielfalt und Variabilität, die der Lösungsansatz ermöglicht. Die gemachten Beobachtungen werden in diesem Abschnitt zusammengefasst und diskutiert. Es folgt zunächst eine Übersicht über die Musikbeispiele. Auf der beiliegenden Audio-CD werden jeweils einige Durchläufe mit wechselnden Interpretationen demonstriert. In den Beispielen *Grave*, *Menuett* und *Spielstück* wurde immer dann, wenn eine Interaktion einen Interpretationswechsel auslöst, ein Signalton (»Ding«) mit aufgenommen.



CD-Tracks
14 und 15

Grave: Zum dritten Satz, *Grave*, aus G. P. Telemanns *Trompetenkonzert D-Dur* (TWV 51:D7) wurden drei Interpretationen ausgearbeitet. Sie unterscheiden sich hinsichtlich Tempo, Rubato, Dynamik und Grundartikulation. Die durchschnittlichen Tempi liegen jeweils bei 58, 62 und 123 bpm. Letztere, die schnelle Variante, wird abgesetzter artikuliert. Die 62-bpm-Variante wird sehr dicht gespielt, enthält stärkere Schwankungen der Makro-Dynamik und ein deutliches geradezu swingendes Mikro-Timing (einer Interpretation von Ludwig Güttler, Pommer & Neues Bachisches Collegium Musicum Leipzig (2000) nachempfunden). Im Gegensatz dazu wird in den beiden anderen Interpretationen die jeweils zweite von zwei Vierteln ein wenig vorgezogen (quasi invers zu Güttler) und die Dynamik schlichter gehalten. Die Besetzung mit drei Solo-Streichern bleibt bei allen Interpretationen gleich. Die Reihenfolge, in der die Interpretationen aufgerufen werden, ist in Tabelle 6.4 aufgeführt.

CD-Track	Abfolge der Interpretationen (bpm-Variante)
14	58, 123, 62, 123, 58, 62, 123, 58
15	123, 62, 123, 62, 123, 62, 58, 123, 58, 123, 58

Tabelle 6.4: Die Abfolge der Interpretationen in den Klangbeispielen 14 und 15, *Grave*. Die Interpretationen unterscheiden sich nicht nur im Tempo; bezugnehmend auf die zugehörigen Erläuterungen werden die Tempi hier lediglich als Bezeichner verwendet.



CD-Tracks
16 und 17

Menuett: Für das *Menuett I* aus der Suite TWV 55:a2 wurden vier Interpretationen und eine End-Interpretation (siehe Abschnitt 6.3 zur Schlussbildung) erarbeitet. Die Interpretationen unterscheiden sich vorrangig in Tempo (85–140 bpm) und Rubato (gerade und swingende Metrik sowie eine Variante, bei welcher der Dreiertakt insgesamt zu einem »Quasi-Vierertakt« verzerrt wird). Außerdem variiert die Besetzung (Flöte, Fagott, Cembalo). Der Continuo-Satz im

Cembalo wird ausgedünnt, wenn die anderen Instrumente hinzukommen. Die Abfolge der Interpretationen zeigt Tabelle 6.5. Die Interpretationen heißen:

Voll1: volle Besetzung, Continuo etwas ausgedünnt, 85 bpm, Swing-Timing

Voll2: volle Besetzung, Continuo stark ausgedünnt, 140 bpm, gerades Timing

Cemb1: Cembalo solo, 100 bpm, normaler Dreier-Takt

Cemb2: Cembalo solo, 85 bpm, quasi Vierer-Takt

End: alle beenden

CD-Track	Abfolge der Interpretationen
16	Voll1, Voll2, Cemb1, Voll2, Cemb2, Cemb1, Voll1, Voll2, Cemb1, End
17	Cemb1, Cemb2, Cemb1, Voll1, Voll2, Cemb1, Voll1, Cemb2, Voll2, Cemb1, End

Tabelle 6.5: Die Abfolge der Interpretationen in den Klangbeispielen 16 und 17, *Menuett*.

Spielstück: Das *Spielstück*, ebenfalls eine Komposition im barocken Stil, bedient ein größeres Instrumentarium und eine größere ausdrucksmäßige Bandbreite. Die Interpretationen heißen

Flat: Flöten, Fagott, Perkussion, 100 bpm, sonst keine interpretatorischen Details und keine Verzierungen

Folk: Flöten, Fagott, Perkussion, 120 bpm, kurze Artikulation, deutliche Schwerpunktbetonung, Verzierungen (Triller)

Maestoso: Hörner, Tuba, Pauken, durchschnittlich 70 bpm, breit aber abgesetzt artikuliert, mehr dynamische Arbeit

Harfensolo: die Harfe spielt freie Figurationen, durchschnittlich 60 bpm, mehr dynamische Arbeit.

End: alle beenden

Die Abfolge in den Klangbeispielen ist in Tabelle 6.6 aufgelistet. Aufgrund der Vielzahl der Instrumente kam zur Tongebung nicht *Vienna Instruments* zum Einsatz sondern der *VSampler* (Speedsoft 2006), ebenfalls ein Software-Sampler aber mit einer weniger Speicher-intensiven Standard-Samplebibliothek. Gleiches gilt auch für die beiden folgenden Musikstücke.



CD-Tracks
18 und 19

CD-Track	Abfolge der Interpretationen
18	Flat, Folk, Maestoso, Harfensolo, Maestoso, Folk, Harfensolo, End
19	Harfensolo, Folk, Harfensolo, Folk, Harfensolo, Folk, Maestoso, Folk, Maestoso, Folk, Maestoso, Folk, Harfensolo, End

Tabelle 6.6: Die Abfolge der Interpretationen in den Klangbeispielen 18 und 19, *Spielstück*.



CD-Tracks
20, 21, 22,
23

Building a Cathedral: Die Musik *Building a Cathedral* demonstriert eine orchestrale Instrumentation und eine cineastischere Tonsprache. Die Tempi reichen von 20 bis 110 bpm. Es wird sehr stark mit dem parallelen Arrangement gearbeitet. Im Vergleich mit den anderen vorgestellten Musiken ist die stilistische und ausdrucksmäßige Bandbreite hier am weitesten gefächert. Details zum Ablauf in den Klangbeispielen gibt Tabelle 6.7. Die Interpretationen sind hier programmatisch und heißen

Beginning: Harfe, Pauken, Celli, 20 bpm, *ppp-p*

Construction: Hörner, Pauken, Celli, Kontrabässe, 55 bpm, *pp-mf*

Serious: erste Violinen, Hörner, Pauken, Celli, Kontrabässe etwas lauter, 60 bpm, *pp-f*

Danger: Posaunen und Streicher ohne Kontrabass, 110 bpm, *mf-ff*

Finale 1: Hörner, Posaunen, Tuba, 100 bpm, *mf-ff*

Finale 2: alle Streicher, Trompeten, Hörner, Posaunen, Tuba, Pauken, 110 bpm, *mf-fff*.

End: alle beenden

Evaluator: Dies ist tatsächlich keine Demonstrations- oder Testmusik, sondern Teil einer musikpsychologischen Studie, bei der die Musik-Engine zum Einsatz kam. Die Probanden sollten Rubato, Schwerpunktbetonung und Artikulation der »Polonoise« aus der Ouvertüre »La Galliarde« (TWV 55:D13) einstellen. Die Latenzvorgabe für Musikwechsel war 500 Millisekunden. Das entsprach in etwa der Länge der Rubato Frames. Sind die Vorgaben deutlich unter der Länge der Rubato Frames, können bei großen Änderungen des Rubato-Parameters

CD-Track	Abfolge der Interpretationen
20	Beginning, Construction, Serious, Danger, Finale 1, Finale 2, End
21	Construction, Finale 1, Finale 2, Finale 1, Construction, End
22	Danger, Finale 2, Danger, Finale 2, End
23	Finale 2, Finale 1, Danger, Serious, Construction, Beginning, End

Tabelle 6.7: Die Abfolge der Interpretationen in den Klangbeispielen 20 bis 23, *Building a Cathedral*.

sprunghafte Timing-Veränderungen entstehen. Für so kurzfristige Interaktionen, in denen die Interpretation differenziert eingestellt werden soll, war der Lösungsansatz eigentlich nicht vorgesehen. Entsprechend ineffizient war die Realisierung (eine vorgefertigte Interpretation für jede Einstellungsmöglichkeit). Eine Schnittstelle, welche direkten interaktiven Zugriff auf einzelne Parameter einer laufenden Interpretation ermöglicht, wäre also eine sinnvolle Erweiterung. Ungeachtet dieser Punkte konnte die Studie aber erfolgreich durchgeführt werden.

Die Musikbeispiele zeigen bereits ein weites Spektrum an musikalischer Vielfalt und Variabilität. Das bringt das parallele Arrangement ohnehin schon mit sich. Es wird aber von den Möglichkeiten der wandelbaren Interpretation noch deutlich erweitert. Einzelne Stimmenkombinationen können interpretatorisch treffender umgesetzt werden und müssen sich nicht mit bspw. einem Durchschnittstempo begnügen, das für alle Stimmenkombinationen funktionieren muss; so war es noch in Kapitel 4. Schon ohne das parallele Arrangement kann durch die interpretatorischen Möglichkeiten eine große ausdrucksmäßige Spanne bedient werden.

Jeder interpretatorische Aspekt wird separat behandelt; das Tempo wird anders überleitet als die Dynamik oder die Artikulation usw. – ein modularer Ansatz. Die gezeigten Beispiele demonstrieren dies vorrangig mit traditionellen Musikstilen. Die Modularität macht es aber einfach, Erweiterungen vorzunehmen und auch andere Aspekte mit einzubeziehen, etwa aus der elektronischen Musik, wo u.a. mit Akustik- und Surround-Effekten, mit elektronischen Verzerrungen und mit Parametern der Klangsynthese gearbeitet wird. Soweit sich diese durch Sequenzen von Controller-Events darstellen und ansteuern lassen, erwachsen auch aus dem MIDI-Format keine Restriktionen.

Prinzipiell ist dieser Ansatz aber nicht auf das MIDI-Format beschränkt. Das Konzept lässt sich ebenso auch auf andere Formate übertragen, etwa *cSound score*. Jeder Agent kann in seinem eigenen Wiedergabemodus arbeiten. Diese Modi werden bislang nur dafür genutzt, um die spezialisierte Ansteuerung verschiedener Sampler zu automatisieren. Das kann aber noch viel weiter gehen und eben auch unterschiedliche Musikformate einbeziehen.

Die Autonomie der Agenten kommt in besonderem Maße den Überleitungen zugute. Die Agenten reagieren unterschiedlich, jeder entsprechend seiner jeweiligen Situation. So kann es vorkommen, dass ein Agent noch den Schlussston spielt und dann beendet, während ein anderer bereits vorher eine Auftaktfigur spielt. In solchen Fällen wirkt die Überleitung meist besonders gelungen. Ein solches »Hand in Hand« mit Überlappungen wird erst dadurch möglich, dass jede Stimme für sich behandelt wird, nicht alle Stimmen gleich – also durch den dezentralen Ansatz, der jeder Stimme/jedem Agenten Autonomie gibt.

Das setzt sich darin fort, dass jeder interpretatorische Aspekt für sich übergeleitet wird und auch nicht zwangsläufig exakt am Zielpunkt nach Überleitungsplan ankommen muss. Durch die nur lose Orientierung am formalen Zielpunkt wird die Struktur der Interpretation mit einbezogen. Wenn ein Crescendo in der Zielinterpretation erst einen halben Takt später beendet ist und dieser Punkt noch im Latenzintervall liegt, wird dies auch als Ziel der Dynamiküberleitung verwendet, anstatt eine zusätzliche Anweisung direkt auf dem Zielpunkt zu erzeugen. So entstehen oft selbst dann plausible Überleitungen, wenn keine Informationen zur Formgliederung des Musikstücks vorliegen, nur die Taktart.

Allerdings können auch unerwünschte Effekte auftreten. An starken Strukturgrenzen (d.h., von Abschnitten, Phrasen, Figuren) kann das Beenden von Agenten doch noch recht abrupt wirken. Hier sollte insbesondere noch mit der Länge und Lautstärke der Schlussstöne gearbeitet werden, denn oft sind sie zu laut und zu kurz. Auch eine zusätzliche Verlangsamung, asynchron zu den weiterspielenden Agenten, kann für die Schlusswirkung zuträglich sein.

Außerdem können Löcher entstehen: Zwei Interpretationen mit disjunkter Besetzung werden ineinander übergeleitet. Dabei kann es vorkommen, dass die Agenten der Ausgangsinterpretation enden, deutlich bevor die Agenten der Zielinterpretation einsetzen. Um dies zu verhindern, sollten die Agenten sich koordinieren und gegebenenfalls schwächere Zielpunkte für die Überleitung zugunsten eines durchgehenden musikalischen Geschehens in Kauf nehmen.

Im Demo *Building a Cathedral* werden auch die interpretatorischen Grenzen des Ansatzes allmählich wahrnehmbar. Extrem unterschiedliche Tempi und Dynamik lassen sich nur schwer über kurze Latenzen überleiten. In verschiedenen Situationen, insbesondere bei schwachen Zielpunkten, erscheint die Überleitung zu kurz bzw. Tempo- und Dynamikübergänge zu rasch. Solche Überlei-

tungen brauchen eine gewisse Mindestlänge. Es müsste ein Maß für die zu überbrückende Differenz eingeführt werden, um solche Situationen zu erkennen und in der Planung der Überleitung mit zu berücksichtigen.

Positiv zu erwähnen ist weiter, dass die stilistischen Unterschiede im parallelen Arrangement größer sein können als im Ansatz von Kapitel 4. Tatsächlich ist das musikalische Material einiger Interpretationen von *Building a Cathedral* nicht mit den anderen kombinierbar. Hier ist es wieder besser, wenn starke Zielpunkte und wenig Überlappung beim Überleiten entstehen. Der musikalischen Homogenität tut es gut, wenn eine zwischengelagerte Interpretation vorliegt, die eine Brücke zwischen zwei Extremen schlägt. Kann solch eine »Zwischeninterpretation« auch automatisch erzeugt werden? Das bereits erwähnte Maß für die zu überbrückende Differenz könnte hierbei dienlich sein.

Der Abschnitt 6.3 gibt außerdem einen Eindruck von der Vielseitigkeit des vorgestellten Ansatzes, der auch Grundlagen für weitergehende Entwicklungen schafft. So implementiert dieser Ansatz zwar keine Techniken, um zwischen gänzlich verschiedenen Musikstücken Überleitungen zu generieren, das Agenten-System stellt aber einen vielversprechenden Ausgangspunkt dafür dar. Das Prinzip einer Überleitung zwischen verschiedenen Musikstücken/Musikobjekten ließe sich wie folgt skizzieren. Die Agenten des Ausgangs-Musikobjektes variieren ihr musikalisches Material so, dass sie einen Anschluss an die Agenten des Ziel-Musikobjektes erreichen. Hier wird die Kommunikation zwischen den Agenten bedeutender. Nicht nur, dass sie melodisch aneinander anschließen müssen, auch die Harmonik, also das Zusammenspiel aller gerade spielenden Agenten muss koordiniert sein.

Musik als Medium zur direkten Interaktion

WER sich mit Musik in audiovisuellen Medien beschäftigt, wird traditionell zuerst und vorrangig auf ihre nicht-diegetische Erscheinungsform stoßen. Auch Kapitel 2 hatte den Fokus auf eben diese gelegt. Vergleichsweise wenig Beachtung erfährt die diegetische Musik, also die Musik, deren Klangquelle Teil des Szeneninventars/der Diegese ist (ein Radio, Straßenmusikanten, eine Musicbox, ein Autoradio). Da sie Teil der Szene ist und innerhalb dieser auch wahrgenommen wird, kann sie einen Einfluss auf die Handlung und die Handelnden ausüben, kann zum inhärenter Teil des szenischen Geschehens werden. In interaktiven Medien kann sie sogar zum Objekt werden, mit dem der Benutzer direkt interagiert.

Dieses Kapitel geht auf praktische und ästhetische Aspekte der diegetischen Musik ein und zeigt die neuen Möglichkeiten und Herausforderungen auf, die das interaktive Medium hierfür bereithält, speziell im Kontext von Video- und Computerspielen. Das führt zu konkreten Design-Richtlinien und zeigt, dass es hier nicht ausreicht, die Musik wie in einem CD-Player einfach nur abzuspielen. Das Problemfeld umfasst u.a. die Simulation von Raumakustik, Klangausbreitung, die Generierung ausdrucksvoller Darbietungen, die automatische Komposition und Variation von Musik.

Die Techniken dafür sind meist schon vorhanden. Einige wurden in den vorangegangenen Kapiteln vorgestellt. In diesem Kapitel wird gezeigt, wie sie zum Einsatz kommen können, um diegetische Musik angemessen zu inszenieren und Interaktion mit Musik zu ermöglichen. Dem liegen folgende Veröffentlichungen zu Grunde:

- (Stockmann, Berndt & Röber 2008),
- (Berndt 2011a).

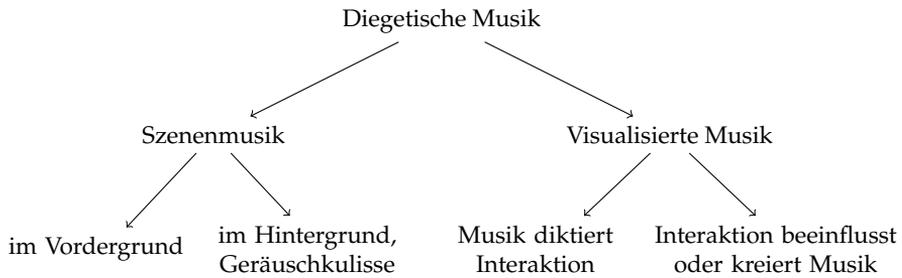


Abbildung 7.1: Ein systematischer Überblick über die verschiedenen Erscheinungsformen der diegetischen Musik.

In unterschiedlichen Spiele-Genres und Spielsituationen tritt diegetische Musik in ganz unterschiedlicher Weise in Erscheinung. Das hat zur Folge, dass auch die Art der interaktiven Einflussnahme auf die Musik sehr verschieden ausfällt. Grundlage dieses Kapitels soll die in Abbildung 7.1 vorgestellte Typologie der diegetischen Musik sein, die in den folgenden Abschnitten detaillierter ausgearbeitet wird. Es folgt zunächst ein kurzer Überblick.

Frühe Beispiele für diegetische Musik können bereits in klassischen Theater- und Opernwerken beobachtet werden, etwa im Finale von W. A. Mozarts *Don Giovanni* (KV 527, 1787 uraufgeführt), einer Ballszene, bei welcher die Musiker auf der Bühne spielen. Da sich das im Theater aber nachteilig auf die Dialog-Verständlichkeit auswirken kann, nimmt die Musik hier oft eine vordergründige Rolle ein. Solange Musik läuft, schweigt der Rest. Erst als der Tonfilm die Möglichkeiten bot, Vorder- und Hintergrund differenziert abzumischen, wurde Musik auch als Teil der Geräuschkulisse interessant, etwa bei der Vertonung einer Bar- oder Disco-Szene. Hintergrund-Elemente konnten nun leiser abgemischt werden, so dass die Konzentration der Zuschauer auf den gesprochenen Text gelenkt wird, vergleichbar mit dem wohlbekannten *Cocktailparty-Effekt* (Arons 1992).

Eine weitere Erscheinungsform von diegetischer Musik entstand im Zusammenhang mit Musik-basierten Computerspielen (auch Musikspiele oder Musik-Videospiele). Sie hat ihren Ursprung in der Ästhetik von Musikvideos: Musik, die auf dem Bildschirm visualisiert wird. Hier wird das virtuelle Szenarium sprichwörtlich durch die Musik erbaut. Musikalische Elemente und Eigenschaften definieren Objekte in der virtuellen Welt und ihre Positionierung oder sind Auslöser für Ereignisse im Spielablauf (z.B. Basedrum-Schläge, durch die Hindernisse auf einer Rennstrecke platziert werden, oder klangfarbliche Veränderungen, die Veränderungen im Farbschema auslösen). Die Visualisierungen haben üblicherweise einen stilisierenden Charakter und nichts mit photorealisti-

schen Darstellungen zu tun. Typische Beispiele für Musikspiele sind *Audiosurf: Ride Your Music* (Fitterer 2008), *Vib-Ribbon* (NanaOn-Sha 1999) und *Amplitude* (Harmonix 2003).

Die Musik muss nicht vollständig vorkomponiert sein. Spiele wie *Rez* (Sega 2001) zeigen, dass auch Spielerinteraktionen als Auslöser für musikalische Ereignisse dienen können. Durch das Spielen des Spiels wird seine Musik kreiert. Die Vermutung liegt nahe, dass es sich hier eher um eine besonders reaktive Musik aus dem »Off« handelt. Da das musikalische Geschehen aber direkt von der Mechanik der Spielwelt (der Diegese) erzeugt wird, muss auch dies als diegetische Musik angesehen werden. Das wird dadurch unterstützt, dass die Spielwelt als Erzeugungsmedium der Musik dient und entwickelt wurde. Die stilisierte visuelle Präsentation tut ihr Übriges, um die ästhetische Entfernung zwischen Musik und Bild zu verringern. Solche Musikspiele sind vom Konzept des virtuellen Musikinstruments, also eines »Musizierwerkzeugs«, nicht mehr weit entfernt, wie etwa das Spiel *Electroplankton* (Iwai 2005) und die rege Diskussion darüber, ob dies noch ein Spiel zu nennen ist, zeigen (Herber 2006).

7.1 Szenenmusik

Die anschaulichste Erscheinungsform von diegetischer Musik ist die Musik, die in der Szene gespielt wird. Hier ist sie meist autonom. Wenn etwa in der Szene ein Radio angeschaltet wird, erklingen in sich schlüssige für sich stehende Musikstücke, die meist nicht erst für den Film bzw. das Spiel entstanden, sondern schon vorher existierten. Das markiert auch den wichtigsten Unterschied zur nicht-diegetischen Musik, die für gewöhnlich nicht für sich allein stehen kann, ohne die visuelle oder die Handlungsebene. Schon Adorno & Eisler (1947a) machen diesen Unterschied deutlich. Deshalb wird Filmmusik vor der CD-Veröffentlichung oft neu arrangiert. Nicht selten fallen Stücke sogar ganz weg, die im Film zwar erklingen, aber sich nicht in die Dramaturgie der CD-Zusammenstellung einfügen. Diegetische Musik hingegen wird als autonome Musik in der Szene gespielt, wie etwa das Wanderlied »The Road Goes Ever On« in der Verfilmung *The Lord of the Rings: The Fellowship of the Ring* (Tolkien 1954, Shore 2001). In dem fiktiven Universum ist das ein präexistentes Volkslied.

7.1.1 Funktionen

Als Element der Soundkulisse dient diegetische Musik dazu, dem inszenierten Szenarium Authentizität zu verleihen. Was wären Bars, Discos, Cafés, Kasinos¹, Supermärkte und Konzerte ohne Musik? Dabei muss sie nicht neutral bleiben und kann weitere Aufgaben wahrnehmen. Sie kann Repräsentant für den Zustand der Umgebung sein. Wenn die Straßenmusikanten plötzlich aufhören zu spielen, ist das mehr als nur eine abrupte Einschnitt in der Soundkulisse. Es kann ein Signal sein; etwas ist passiert, das sie beim Spielen unterbrochen hat, etwas hat sich grundlegend geändert.

Umgekehrt kann es auch sein, dass dramatische Ereignisse geschehen, etwa dass der Protagonist angegriffen wird, aber die Musik im Hintergrund reagiert nicht. Statt dessen spielen fröhliche Melodien. Solch ein indifferentes Verhältnis zwischen Vordergrund und Hintergrund kann die Dramatik des Geschehens unterstreichen, wird manchmal sogar als philosophischer Kommentar interpretiert: Was immer dort passiert, für den Rest der Welt ist es bedeutungslos. »Das Leben geht weiter.«²

Zwar ist die Klangquelle Teil der virtuellen Szene, sie muss aber nicht unbedingt sichtbar sein. Der Klang eines Grammophons reicht aus, um zu signalisieren, dass sich eines in der Szene befindet, nur eben gerade nicht sichtbar ist. Auf diese Weise dient die diegetische Musik, ebenso wie diegetische Sound-Effekte im Allgemeinen, dazu, nicht sichtbare Elemente der Szene zu assoziieren. »Hinter der Tür steht der Kerl, der immer diese Melodie summt!« Die Art, wie er die Melodie summt, kann etwas über seinen inneren Zustand aussagen (z.B. fröhlich pfeifend, mit seufzendem Tonfall oder in sich gekehrt, während er etwas anderes tut). Musikalisch begabte Akteure mögen sogar die Tonart ändern (Dur/Moll), die Melodie variieren oder darüber improvisieren.

Je mehr die Musik aus dem Hintergrund hervortritt, desto wichtiger wird auch ihre Inszenierung. Hat der Sänger gut zur Musik agiert? Stimmt der Fingersatz des Pianisten mit der Musik überein? Die Musik kann zum Regulator von Bewegung und Schauspiel werden. Man denke an das tanzende Liebespaar. Berühmt ist auch das Finale aus Alfred Hitchcocks (1956) *The Man Who Knew Too Much*. Während einer Konzertaufführung von Arthur Benjamins Kantate *Storm Clouds* versucht der Attentäter, an der lautesten Stelle des Musikstücks synchron mit einem Beckenschlag zu schießen, um unerkannt zu bleiben.

¹siehe (Collins, Tessler, Harrigan, Dixon & Fugelsang 2010)

²Lissa (1965), Seite 166.

7.1.2 Design-Prinzipien

Diegetische Musik ist, im Gegensatz zu ihrem nicht-diegetischen Counterpart, Teil der Szene und unterliegt folglich auch den dort geltenden akustischen Verhältnissen. Eine hallige Kirche, ein Wohnzimmer, eine Waldszene, jeder Ort hat seine eigene Akustik und seine eigenen Resonanzen. Discomusik klingt außerhalb des Gebäudes anders als innerhalb; die Wände filtern Höhen und Mitten, nur die tiefen Frequenzen dringen durch. Das ändert sich drastisch, wenn man die Disco betritt. Diegetische Musik klingt nicht wie eine perfekt aufgenommene und abgemischte Studioproduktion. Ein Solosänger im Orchester ist auf CD-Aufnahmen immer gut zu hören, bei Live-Aufführungen kann er vom Orchester aber auch übertönt werden³.

Auch die Positionierung der Klangquellen im Stereo- bzw. Surroundpanorama hat nichts mit einer traditionellen Musikproduktion gemein. Diegetische Musik muss von dort her klingen, wo sie gespielt wird. Der Mensch kann Soundquellen in der realen Welt mit einer Genauigkeit von bis zu 2 Grad lokalisieren (Fastl & Zwicker 2007). Das ist bei Stereo- und Surroundsystemen zwar deutlich schlechter, aber selbst Stereo-Beschallung kann noch ungefähre Richtungsinformationen vermitteln. Wenn sich die Soundquelle bewegt oder der Spieler in der Lage ist, sich in der virtuellen Welt zu bewegen und zu drehen, wird die Lokalisierungsleistung wieder deutlich präziser (Röber 2008). Wenn sich also der visuelle Repräsentant einer Soundquelle relativ zum Hörer bewegt, darf dieser seinen Sound nicht »verlieren«. Das würde die Präsenz und Glaubwürdigkeit des virtuellen Objektes schwächen. Die Musik sollte ebenso wie jeder andere Sound-Effekt an dem Ort im Panorama positioniert werden, an dem sie gespielt wird.

Sofern es sich dabei nur um einen einzelnen Lautsprecher handelt, mag eine einzelne Soundquelle dafür genügen. Aber angenommen, es wäre ein Ensemble mit mehreren Musikern, ein ganzes Orchester, und der Spieler kann dicht heran kommen, vielleicht sogar zwischen den Musikern herumlaufen und sich jeden einzelnen aus der Nähe anhören. Nicht zu vergessen, dass jedes Instrument sein eigenes Abstrahlverhalten hat. Ein Trompeter klingt von hinten anders als von vorn, denn die Frequenzbänder breiten sich unterschiedlich aus – ein Fakt, den Tontechniker bei der Mikrofonierung berücksichtigen müssen (Meyer 2009).

Aber wie weit muss man als Entwickler hier gehen? Wie viel Realismus ist notwendig? Die Frage wird von dem allgemeinen Realismus der Inszenierung beantwortet. Cartoons sind etwa relativ tolerant gegenüber auditiven Inkonsistenzen. Selbst Hollywood-Film-Ästhetik zielt nicht auf eine realistische Wiedergabe

³Mit diesem Problem müssen u.a. die Architekten von Opernsälen umgehen, müssen den Raum so konzipieren, dass die Sänger auf der Bühne einen akustischen Vorteil gegenüber dem Orchester im Graben haben.

der Szene ab, sondern auf den Effekt, der mit einem bestimmten Sound-Design beim Zuschauer erzielt wird. Ekman (2009) beschreibt weitere Situationen, in denen die subjektive Wahrnehmung von der physikalischen Realität abweicht. Gründe dafür können sein: Aufmerksamkeit (*Cocktailparty-Effekt*), Stress, auditive Beeinträchtigungen (z.B. Schwerhörigkeit in bestimmten Frequenzbereichen), körpereigene Geräusche (die von außen nicht zu hören sind, von innen aber sehr deutlich), körpereigene Resonanzen, Halluzinationen (z.B. der sog. *subjektive Tinnitus*) usw.

Die diegetische Musik gehört auf die gleiche Ebene wie alle anderen Sound-Effekte in der virtuellen Umgebung. Sie fällt damit in den Zuständigkeitsbereich von Audio-Engines wie bspw. *OpenAL* (Loki Software & Creative Technology 2009), *DirectSound* (Microsoft Corporation 2009), *FMOD Ex* (Firelight Technology 2009) und *AM3D* (AM3D A/S 2009). Einen Ansatz zum Akustik-Rendering auf Basis von Graphik-Hardware beschreiben Röber, Kaminski & Masuch (2007), Röber (2008). Eine Audio-Engine speziell für mobile Geräte stellt Stockmann (2007) vor.

Es genügt aber nicht, die Musik nur mit der richtigen Akustik, dem richtigen Panorama und den richtigen Filter-Effekten abzuspielen. Um den Eindruck zu stärken, eine Live-Aufführung zu erleben, ist es oft nützlich, wenn die Darbietung zu einem gewissen Grad fehlerbehaftet ist – ganz im Sinne von »realer als die Realität«. Die Tonschwankungen eines alten Plattenspielers und das Knistern des Radios bei Empfangsstörungen sind typische Beispiele. Ganz zu schweigen von all den Dingen, die Musikern, auch Profis, bei Live-Aufführungen passieren, wie etwa Intonationsschwankungen, schlechtes (sprich asynchrones) Zusammenspiel und falsche Töne. Bei Studio-Produktionen kann ein Musikstück so oft wiederholt werden, bis eine perfekte Einspielung gelingt oder aus dem aufgenommenen Material zusammengeschnitten werden kann. Aber bei Live-Aufführungen passieren all diese Dinge und können auch nicht im Nachhinein korrigiert werden. Es macht den Live-Eindruck authentischer, wenn all das auch in der diegetischen Musik vorkommt.

7.1.3 Nichtlinearität und Interaktivität

Diese Authentizität geht aber schon dann wieder verloren, wenn der Spieler ein Musikstück mehr als einmal hört. Eine typische Situation in einem Computerspiel: Der Spieler betritt eine Szene zum wiederholten Mal. Die diegetische Musik fängt dabei jedesmal von vorn an, als hätten die Musiker in dem Moment aufgehört, als der Spieler die Szene verließ und nur darauf gewartet, dass er wieder zurückkommt. Solch ein Déjà-vu-Effekt schadet der Glaubwürdigkeit

der virtuellen Welt. Die Musik muss weiterspielen, selbst wenn sie nicht hörbar ist, und wenn der Spieler zurückkehrt, muss er etwas verpasst haben.

Eine ähnliche Situation tritt dann auf, wenn sich der Spieler sehr lange in einer Szene aufhält und das eine Musikstück immer und immer wiederkehrt. Es wird bald deutlich, dass die Jazzband nur ein Stück kann, der Radiosender nur ein Lied im Repertoire hat usw. In gewisser Weise ist damit das Ende der virtuellen Welt erreicht. Es gibt nicht mehr. Eine mögliche Lösung kann sein, dass man eine größere Auswahl von Stücken bereitstellt, aus der zufällig das nächste ausgewählt wird, quasi wie eine Musicbox. Trotzdem kann ein Stück mehrfach auftauchen. Hier ist es wichtig, dass die Darbietungen nicht identisch sind. Das Radio hat nicht immer an der gleichen Stelle im Lied eine Empfangsstörung. Die Musiker versuchen mit jedem Durchlauf, besser zu spielen, konzentrieren sich darauf, dass ihnen die alten Fehler nicht noch einmal unterlaufen, und machen neue. Die Interpretation kann auch absichtsvoll variieren: »Diesmal spielen wir es schneller!« Techniken dafür wurden in den Kapiteln 5 und 6 vorgestellt.

Aber die ausdrucksvolle Interpretation ist nicht die einzige Möglichkeit, Varianz in den musikalischen Ablauf zu bringen. Das sequentielle und parallele Arrangement erlauben es, die Bausteine der Musik immer wieder neu miteinander zu kombinieren, so dass jede Wiederholung wie ein Neuarrangement oder eine Variation wirkt. Ein Ansatz dafür wurde in Kapitel 4 vorgestellt. Durch generative Techniken kann der Raum der musikalischen Varianz sogar noch viel weiter gespannt werden. Man stelle sich eine virtuelle Jazzband vor, welche die ganze Zeit über improvisiert, kontinuierlich neues kreiert ohne sich je zu wiederholen. Ausgangspunkt dafür kann eine vorgegebene Melodie sein, wie etwa in dem evolutionären Ansatz *GenJam* (Miranda & Biles 2007). Die beiden Systeme *MeloNet* und *JazzNet* erzeugen Melodieumspielungen auf Basis eines zuvor trainierten neuronalen Netzes (Hörnel & Menzel 1999, Hörnel 2000). Stenzel (2005) variiert die Harmonisierung basierend auf dem sog. *Kadenzgraph*.

Aber Musik kann nicht nur variiert werden; sie kann auch frei von jeder Vorlage generiert werden. Ein neuerer Ansatz zur Melodiegenerierung wird z.B. von Klinger & Rudolph (2006) vorgestellt. Verfahren zur Harmonisierung von Melodien werden u.a. von Ebcioğlu (1992), Schottstaedt (1989), Verbiest et al. (2009) beschrieben. Andere Ansätze zur Musikkomposition finden sich bei Löthe (2003), Taube (2004), Pozzati (2009). Die Vielfalt der Ansätze zur algorithmischen Musikgenerierung ist groß. Einen Überblick geben Papadopoulos & Wiggins (1999), Pachet & Roy (2001).

Ein Aspekt, der die diegetische Musik in interaktiven Medien ganz besonders betrifft, blieb bislang jedoch noch unbenannt. Als Teil der Diegese kann die Musik auch von dieser beeinflusst werden, insbesondere vom Spieler. Welcher Spieler ist nicht versucht, die Musiker einmal anzuklicken und zu sehen, was pas-

siert? Im einfachsten Fall schaltet er damit ein Radio ein und aus oder wählt einen neuen Song in der Musicbox. Die Interaktion mit virtuellen Musikern kann jedoch komplizierter ausfallen. Hierbei können zwei Modi unterschieden werden: die *destruktive* und die *konstruktive* Interaktion.

Destruktive Interaktion: Diese Art der Interaktion »stört« den Musiker beim Spielen. Der Spieler mag mit ihm reden, ihn anrempeln, seine Aufmerksamkeit ablenken. Das kann dazu führen, dass er nicht mehr synchron mit dem Rest des Ensembles spielt, rhythmisch unsauber wird (Kapitel 5 stellte dazu Modelle vor) oder sogar falsche Töne spielt. Die destruktive Interaktion beeinflusst die Qualität der Darbietung. Ein einfacher Ansatz, falsche Töne zu kreieren, kann folgendermaßen aussehen. Es werden zufällig einige der zu spielenden Töne ausgewählt – je mehr der Musiker abgelenkt wird, desto mehr Töne – und um ein zufälliges Intervall in der Tonhöhe versetzt. Die Intervallspanne wird größer gewählt, je mehr der Musiker abgelenkt wird; z.B. der diatonische Nachbarton für kleine Fehler, größere Intervalle und tonartfremde Töne für große Fehler.

Konstruktive Interaktion: Die konstruktive Interaktion beeinflusst die Musik an sich. Man stelle sich einen Jazzmusiker vor, der angespornt vom Applaus des Publikums immer abenteuerlichere Improvisationen wagt. Oder ein Straßenmusiker, der depressiv vor sich hin spielt, bis der Spieler ihm eine Münze gibt und mit einer sich zum Heiteren wandelnden Musik belohnt wird. Ein berühmtes Beispiel für konstruktive Interaktion mit den Musizierenden kann in dem Computerspiel *Monkey Island 3: The Curse of Monkey Island* (LucasArts 1997) beobachtet werden. Die Piraten-Crew des Spielers singt das Lied »A Pirate I Was Meant To Be«; zwischen den Strophen wählt der Spieler im Multiple-Choice-Verfahren eine Antwort, auf die sich dann die nächste Strophe reimt. Aufgabe des Spielers ist es, die Antwort zu finden, auf die keiner einen Reim findet, so dass die Crew wieder zurück an die Arbeit geht. Die variable Abfolge von Strophen und Zwischenspielen ist hier durch sequentielles Arrangement umgesetzt.

So viel Aufwand, solch ein großes und komplexes Arsenal an Werkzeugen für eine Musik, die meist nur im Hintergrund spielen soll? Ist das alles nötig? Die Antwort ist »nein«. Dies soll ein Überblick über all die Werkzeuge sein, von denen das eine oder andere für die Inszenierung bestimmter Situationen nützlich sein kann. Zudem bilden sie die Basis dafür, dass Musik auch in interaktiven Medien in den Vordergrund treten kann, zum interaktiven Element wird und so auch für den Spieler aus nächster Nähe erlebt werden kann, wie es sonst nur Musikern vorbehalten ist.

7.2 Visualisierte Musik

Anstatt nur die Darbietung von Musik zu visualisieren, also spielende Musiker oder Soundquellen zu zeigen, kann auch Musik an sich visualisiert werden. Tatsächlich wird dabei nicht Musik als ganzes visualisiert, sondern eine Auswahl von musikalischen Merkmalen, wie etwa rhythmische Muster oder melodische Konturen. In Musikspielen, ebenso wie in Musikvideos, werden diese stilisierten Darstellungen oft mit »realistischen« Darstellungen collagiert.

Solche Collagen können in der *Guitar Hero*-Serie (Harmonix 2006–2010) beobachtet werden. Während im Hintergrund eine Konzertaufführung dargestellt wird, werden im Vordergrund die Gitarrenriffs angezeigt, die der Spieler ausführen soll. *PaRappa the Rapper* (NanaOn-Sha 1996) zeigt ebenfalls die Sänger und eine unauffällige Zeile von Symbolen in der oberen Bildschirmhälfte, welche die als nächstes zu drückenden Tasten anzeigt und das Timing, um mit der Musik mitzuhalten. In *Audiosurf* wird die komplette Szene aus der Musik abgeleitet: der Verlauf der Rennstrecke, die Platzierung von Hindernissen und Power-Ups, das Farbschema, Hintergrundobjekte, visuelle Effekte und sogar die Fahrgeschwindigkeit. Aus der Musik werden also nicht nur die visuellen Gegebenheiten abgeleitet, sondern auch spielerische Ereignisse und ihre Qualität. Einige Musikstücke erzeugen schwierigere Rennkurse als andere.

7.2.1 Musik als Diegese

In Musikvideos sind die Visualisierungen ästhetisch freier als in Musikspielen, wo sie schließlich auch dazu dienen, die Spielmechanik zu vermitteln. Bspw. werden gleiche Interaktionen mit gleich bleibender Symbolik angezeigt. Die Visualisierungen sind zwangsläufig konsistenter und werden in einer wohl geordneten Weise präsentiert, die sich meist aus der Zeit-Tonhöhe-Notation des traditionellen Notendrucks ableitet – Zeitachse horizontal, Tonhöhe (allgemeiner: Ereignisqualität) vertikal, nahende Ereignisse scrollen von rechts nach links. Die Orientierung dieses Schemas kann variieren. Shultz (2008) unterscheidet drei Modi:

Reading Mode: Der Reading Mode entspricht der traditionellen Notation, wie bereits beschrieben. Ein Beispiel ist das Spiel *Donkey Konga* (Namco 2003).

Falling Mode: Die Zeitachse ist vertikal ausgerichtet und die Ereignisqualität horizontal. Nahende Ereignisse »fallen herab« (*Dance Dance Revolution* von Konami 1998).

Driving Mode: Der Driving Mode entspricht dem Falling Mode, jedoch verläuft die Zeitachse in z -Richtung (in die Tiefe). Herannahende Ereignisse kommen von vorn (*Guitar Hero*).

Die Darstellungen müssen aus musikalischer Sicht nicht akkurat sein. Oft werden sie zugunsten einer besseren Spielbarkeit vereinfacht. *Guitar Hero* zeigt z.B. keine exakten Tonhöhen an, sondern nur eine ungefähre melodische Kontur. Selbst die wird auf den geringeren Ambitus herunterskaliert, den der Spielcontroller bedienen kann. Die Visualisierungen müssen sich aber nicht zwangsläufig aus der Musiknotationspraxis ableiten und »Noten« anzeigen. Es kann sich auch um Struktureigenschaften der Musik handeln. In *Amplitude* ist es etwa das Zusammenspiel mehrerer Tracks (Rhythmus, Vocals, Bass etc.), die als parallele »Fahrspuren« dargestellt sind. Klangfarbe wird oft in visuelle Farbe überführt (*Audiosurf*). Andere Visualisierungstechniken basieren auf der Wellendarstellung des Audiosignals oder seiner Fourier-Transformation (Spektraldarstellung). Sie finden oft in Visualisierungs-PlugIns für Medien-Player Anwendung. Der Vollständigkeit halber sollte noch erwähnt werden, dass dabei nicht nur statische Bilder entstehen sollten. Da die Musik/das Audiosignal sich über die Zeit verändert, müssen auch die Visualisierungen »in Bewegung bleiben«.

In Musikspielen, wie auch in Musikvideos, stellt die Musik die zentrale Schicht des Mediums dar. Sie unterliegt keinen außermusikalischen Abhängigkeiten etwa zur visuellen Schicht, ganz im Gegensatz zu den meisten Ausprägungen der nicht-diegetischen Musik (Kapitel 2) und zur Szenenmusik (Abschnitt 7.1). Umgekehrt, nun ist die visuelle Ebene die begleitende, die sich unterordnet. Sie stellt jedoch keine Soundquellen im traditionellen Sinne (wie bei der Szenenmusik) dar, so dass der Begriff der diegetischen Musik als Musik, deren Soundquelle Teil der Diegese ist, hier nicht angewendet werden kann. Trotzdem ist diese Musik diegetisch zu nennen, sogar mehr noch als die visuelle Ebene, die ja nur eine Übersetzung musikalischer Merkmale in visuelle Metaphern ist und unsere visuell geprägten Interaktionsmechanismen erleichtert. Die Bildebene illustriert, kommentiert, konkretisiert und kanalisiert Assoziationen, welche die Musik evoziert (Kungel 2004). Aber die Interaktion findet in der musikalischen Domäne statt. Hier wird die Diegese durch die Musik begründet. Diese *musikalische Diegese* ist der Raum der musikalischen Möglichkeiten, den der Spieler mehr oder weniger frei erforschen kann.

In dieser Welt ist die Musik keinen Restriktionen unterworfen. Die visuelle Ebene muss folgen. Die imaginäre Welt, die sich daraus ergibt, hat keine logischen oder rationalen Grenzen. Die Rennstrecken in *Audiosurf* verlaufen frei im schweirelosen Raum; selbst der Hintergrund hat nichts mit konventionellen Himmels- oder Weltraumdarstellungen gemein. Praktische Grenzen in der Musikdarbietung, wie sie für die Szenenmusik existieren (Empfangsstörungen des Radios,

falsche Töne usw.), gibt es nicht. Hier kann sich die Musik in höchster Qualität präsentieren.

7.2.2 Interaktivität in der musikalischen Domäne

Jedoch sind die Möglichkeiten, diese musikalische Welt interaktiv zu erforschen, noch sehr limitiert. Oft beschränkt es sich darauf, dass die Musik Tempo und Rhythmus von Geschicklichkeitsübungen vorgibt, ohne darauf zu reagieren, wenn der Spieler diese gut oder schlecht meistert. Das Prinzip entspricht dem der sog. *On-Rails Shooter*-Spiele, die eine fest vorgerenderte Videosequenz zeigen, während die Aufgabe des Spielers sich darauf reduziert, auftauchende Ziele abzuschießen. Ein einzelnes feststehendes Musikstück ist letztlich nichts anderes als eine fest vorgegebene Kamerafahrt durch eine viel größere Welt.

Die Musik muss nicht so feststehen, und Aufgabe des Spielers muss nicht nur sein, mit ihr mitzuhalten. »Music videogames would benefit from an increasing level of player involvement in the music.«⁴ Die Diegese, die Spielwelt, ist der Raum der musikalischen Möglichkeiten. Die letztlich zu hörende Musik spiegelt die Reaktionen der Spielwelt auf die Interaktionen des Spielers wieder. Das beginnt damit, dass nur die Töne (allgemeiner: musikalischen Ereignisse) gespielt werden, die der Spieler tatsächlich trifft bzw. auslöst, nicht diejenigen, welche er hätte treffen sollen. Das Spiel *Rez* (Sega 2001) präsentiert sich visuell zwar als *On-Rails Shooter*, aber auf musikalischer Ebene steht nur ein ostinates Grundschema fest (hauptsächlich Rhythmus-Patterns); alle übrige musikalische Aktivität geht vom Spieler aus. Bei jedem Durchspielen entsteht eine andere musikalische Ausgabe. Hier stellt Williams (2006) fest: »it is a pleasure not just to watch, but also to listen to someone who knows how to play *Rez* really well, and in this respect *Rez* comes far closer to realising the potential of a music videogame – that of musical creation, rather than repetition.«⁵

In *Rez* spannt der Strom von Zielen die musikalischen Möglichkeiten auf. Die spielerische Freiheit mag immer noch eingeschränkt sein, aber das gibt den Entwicklern Ansatzpunkte, um eine gewisse Kontrolle über die Musikdramaturgie zu behalten. Hier liegt auch die Grenze dessen, was mit vorkomponiertem und vorproduziertem Material möglich ist. Mehr Interaktivität setzt auch mehr musikalische Flexibilität voraus. Dabei können zwei Wege eingeschlagen werden:

- Interaktion durch musikalische Primitivereignisse,
- Interaktion mit High-Level-Strukturen und Design-Prinzipien.

⁴Williams (2006), Seite 7.

⁵Williams (2006), Seite 7/8.

Musikalische Primitivereignisse sind einzelne Töne, Klänge, Geräusche auch einzelne perkussive Schläge. Es können sogar Gruppen von solchen Primitivereignissen noch als Primitivereignis gelten, wenn sie noch keine musikalische Figur bilden (z.B. Cluster und Arpeggien). Selbst kurze motivische Figuren können noch Primitivereignisse sein, wenn sie in ihrer Form weitgehend statisch/unvariabel sind. Die Spielmechanik ermöglicht dem Spieler, diese Ereignisse auszulösen und ihre Eigenschaften (Tonhöhe, Lautstärke, Klangfarbe, die Dichte des Clusters usw.) festzulegen. Das führt schließlich zu virtuellen Musikinstrumenten. Dabei kann es sich um Nachbildungen real existierender Instrumente handeln (Klavier, Geige, Posaune), deren Handhabung mit Maus, Tastatur oder Gamepad sich aber sehr von den Originalen unterscheidet. Deshalb entstanden verschiedene Controller, die in Form und Handhabung den Originalen nachempfunden sind, etwa der Gitarren-Controller von *Guitar Hero*, die *Donkey Konga*-Bongos, der Turntable-Controller von *DJ Hero* (FreeStyleGames 2009) und die lange Liste von MIDI-Instrumenten. Roads (1996) gibt einen Überblick über solche professionellen musikalischen Eingabegeräte.

Aber es müssen nicht nur real existierende Instrumente nachgeahmt werden. Die technischen Möglichkeiten erlauben noch viel mehr und auch ganz andere Interaktionsmetaphern, die eine Grundlage für elektronische Musikinstrumente sein können. Das wird bspw. vom Gesten-basierten *Theremin* (Theremin 1924) demonstriert, von den mit Sensoren versehenen *Drum Pants* (Hansen & Jensenius 2006), dem berührungsempfindlichen *Arcontinuo* (Bertin, de Ioannes & Sylteros 2010) und dem Ultraschall-basierten Instrument von Ciglar (2010a,b), das durch eine Anordnung von 97 Ultraschallsendern den Klang mit der Hand taktil spürbar macht und es mittels eines Ultraschallempfängers ermöglicht, den Klang auch mit der Hand zu formen.

Die zentrale Aufgabe beim Design von virtuellen und elektronischen Musikinstrumenten besteht darin, geeignete Interaktionsmetaphern zu entwickeln, die es dem Anwender erlauben, gewisse Ton- und Klangeigenschaften zu steuern. Im spielerischen Umfeld kann dies das Anvisieren und Abschießen von Zielen sein (*Rez*) oder das Anstoßen von Objekten, die dann wieder mit anderen Objekten kollidieren (*Electroplankton*). In der Veröffentlichung »A Musical Instrument based on Interactive Sonification Techniques« (Stockmann et al. 2008) stellen wir ein Musikinstrument vor, das auf den Techniken zur Sonifikation und interaktiven Exploration von Volumendaten von Stockmann (2008) basiert. Die *Tone Wall* (siehe Abbildung 7.2(a)) ist eine virtuelle Wand, in die ein Interaktor eindringen kann – Handgesten, die durch einen Sensor getrackt werden. Abhängig von der Position innerhalb der Wand werden unterschiedliche Tonparameter gesteuert. Entlang der Horizontalen (x -Achse) wird die Klangfarbe gesteuert (von dumpf/obertonarm bis schrill/obertonreich). Die Vertikale (y -Achse) bestimmt die Tonhöhe. Entlang der z -Achse wird die Tonlautstärke abgetragen (je

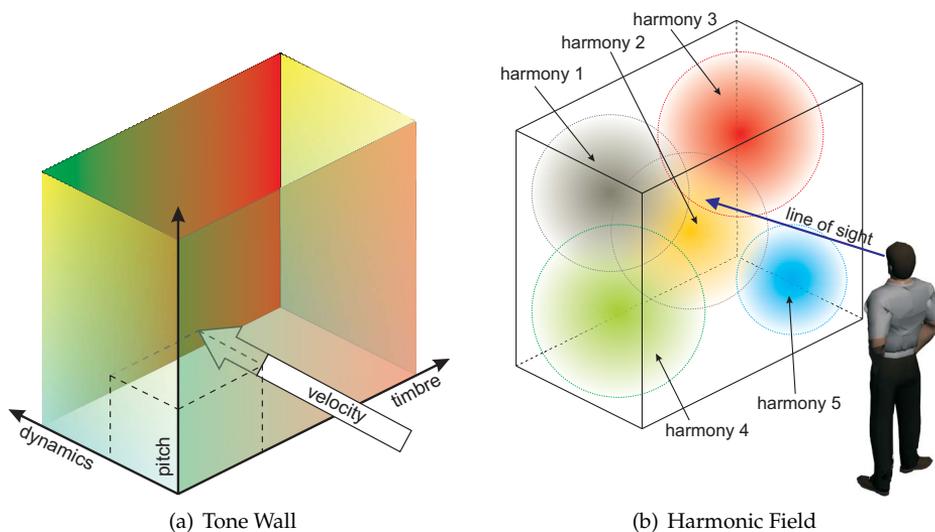


Abbildung 7.2: Ein virtuelles Musikinstrument, das durch Hand- und Kopfgeesten gespielt wird.

weiter in der Wand, desto lauter erklingt der Ton). Das Einschwing-Verhalten des Tones wird durch die Geschwindigkeit gesteuert, mit welcher der Interaktor in die Wand eindringt. Eine denkbare Erweiterung wäre, dass die Wand mit Partikeln gefüllt wird, die durch den Interaktor zum Klingen angeregt werden und mit anderen Partikeln interagieren können; dabei kann jedes Partikel seine eigenen Klangeigenschaften haben. Im *Harmonic Field* (siehe Abbildung 7.2(b)) können Wolken definiert werden, die je einer Harmonie entsprechen. Durch Head-Tracking wird ermittelt, wohin sich der Benutzer wendet, d.h., welche Wolken er »ansieht«. Aus ihnen wird die Harmonik eines begleitenden Ostinatos abgeleitet. Die *Tone Wall* ist das »Lead-Instrument«; mit dem *Harmonic Field* wird die Begleitung generiert. Denkbar wäre, dass auch das *Harmonic Field* durch einen Hand-Interaktor bedient werden kann.

Der Spieler muss aber nicht allein musizieren. Eine autonome Begleitung kann im Hintergrund mitgehen. Dafür werden oft repetitive Musiken verwendet (Ostinato, Vamp, Riff), die jedoch auf Dauer langweilig werden können. Variationstechniken können Abhilfe schaffen. Durch paralleles Arrangement können unterschiedliche Klangfarben, Begleitstimmen und -rhythmen miteinander kombiniert werden. Kapitel 6 hat die Techniken dazu eingeführt. Generative Techniken können frei über einem harmonischen und rhythmischen Grundschemata improvisieren; der vorige Abschnitt 7.1 hat hierfür bereits einige Ansätze vor-

gestellt. Außerdem muss sich nicht immer ein und das selbe Grundschema wiederholen. Es kann sich auch ändern, etwa nach dem Prinzip »spiele 6 mal *A*, dann 2 mal *B* und 8 mal *C*«. Solch eine vorgefertigte Begleitung verwendet z.B. *GenJam* (Miranda & Biles 2007). Ist dem Spieler dieser Ablauf vorher nicht bekannt, ist es gut, wenn er auf Änderungen vorbereitet wird, etwa durch sog. *Fill Ins* (auch *Breaks*, *Pickups*) vom Schlagzeug, die auf die Änderung hinspielt. Ferner ist es möglich, einzelne Akkorde durch verwandte Akkorde mit ähnlichem Tonmaterial zu ersetzen, so dass es für den Spieler leichter ist, trotz Varianz passende Töne zu finden und auf Änderungen zu reagieren.

Der Spieler kann vor diesem Hintergrund frei musizieren. Aber sollte er dabei wirklich alle Freiheiten haben? Sollte er sogar unpassend und falsch spielen können? Um gerade unerfahrene Spieler nicht von vornherein schon zu entmutigen, kann ein vereinfachter Schwierigkeitsgrad angeboten werden, bei dem nur die Töne spielbar sind, die zur Musik passen. Statt eines »falschen« Tones erklingt automatisch der nächstgelegene passende Ton. Durch eine kurze zeitliche Verzögerung kann der Ton in die metrische und rhythmische Grundstruktur eingepasst/quantisiert werden. Die Lautstärke kann automatisch gesteuert werden. Umgekehrt kann sich auch die Begleitung nach dem Musiker richten und sich etwa in Tempo und Dynamik an den Spieler anpassen (Dannenberg 1989). In höheren Schwierigkeitsstufen werden solche Hilfssysteme nach und nach abgeschaltet. Erfahrene Spieler können frei in dem größeren Raum von Möglichkeiten experimentieren und sich ausprobieren.

Die Interaktion mit High-Level-Strukturen und Design-Prinzipien ist weniger direkt. Die Musik spielt autonom und reagiert auf das Verhalten des Benutzers. Während die zuvor beschriebenen Musikinstrumente eher als »Musizierwerkzeuge« wahrgenommen werden, kann hier viel stärker der Eindruck einer musikalischen Diegese, einer virtuellen Welt voller Entitäten, die ein eigenes Verhalten haben, auf den Spieler reagieren und mit ihm interagieren, entstehen. Der Spieler beeinflusst das Arrangement des musikalischen Materials oder die Design-Prinzipien, nach denen es generiert wird. In *Amplitude* (im Standard-Gameplay-Modus) wird das Arrangement von mehreren parallelen Tracks beeinflusst. Ein Track repräsentiert eine Komponente des Musikstücks, wie Bass, Vocals, Perkussion etc. Über einen Geschicklichkeitstest kann jeder Track einzeln für eine gewisse Zeit aktiviert werden. Der Test leitet sich aus den melodischen und rhythmischen Eigenschaften des Materials ab, das er repräsentiert. Ziel des Spieles ist es, alle Tracks zu aktivieren. Ist das erreicht, geht es nahtlos in den nächsten Abschnitt des Musikstücks weiter.

In *Amplitude* ist die Musik vorkomponiert und deshalb wenig variabel. Letztlich führt jeder Lauf zur gleichen Zielmusik. In anderen Ansätzen wird die Musik in Echtzeit beim Spielen generiert. Benutzerinteraktion steuert den Generationsprozess, was schließlich zu neuer Musik führt. Für diese Konstellation von au-

tonomer Musikgenerierung und Interaktion prägte Chapel (2003) den Begriff *Active Musical Instrument*. »The system actively proposes musical material in real-time, while the user's actions [...] influence this ongoing musical output rather than have the task to initiate each sound.«⁶ Da ein solches aktives Musikinstrument um jeden Kompositionsalgorithmus herum konstruiert werden kann, gibt es quasi keine stilistischen Grenzen. Die Musik muss nicht einmal repetitiv sein, wie es in *Amplitude* noch der Fall ist.

Das erste dieser aktiven Musikinstrumente stellte Chadabe (1985) vor. Während die Musik automatisch generiert wird, steuert der Benutzer ausdrucksmäßige Parameter, wie Artikulation, Tempo und Klangfarbe. Chapel (2003) benutzt Fraktale zur Musikgenerierung, die vom Benutzer bearbeitet werden können. Eldridge (2002) benutzt selbstaussgleichende Systeme; Störung (aus dem Gleichgewicht bringen) des Systems verursacht musikalische Aktivität. Das Musikspiel *Electroplankton* für den Nintendo DS bietet verschiedene Spielmodi (sog. Plankton-Typen) an, welche die musikalische Domäne jeweils unterschiedlich aufbauen. Darunter befindet sich ein Melodie-Progressions-Graph (Plankton-Typ Luminaria): Jeder Knoten des Graphen steht für einen Ton. Die Kanten repräsentieren die Übergänge zum nächsten Ton und können vom Spieler gesetzt werden. Bis zu vier Partikel können den Graphen zeitgleich durchlaufen und erzeugen einen vierstimmigen Satz. Der Plankton-Typ Tracy ist ein melodischer Interpret für graphische Kurven, die vom Spieler frei gezogen und modifiziert werden können.

Brown, Wooller & Kate (2007) stellen mit dem *Morph Table* einen Tabletop-basierten Ansatz vor. Auf dem Tabletop werden Würfel (allgemeiner: Tangibles; das müssen nicht nur Würfel sein) platziert. Jeder Würfel repräsentiert einen Track im Musikstück, dessen Wiedergabe in dem Moment aktiviert wird, in welchem der Würfel auf dem Tabletop platziert wird. Für jeden Würfel sind zwei Patterns vordefiniert. Sie stehen für den linken und rechten Rand des Tabletops. Abhängig von der relativen Position des Würfels zwischen diesen beiden Extremen werden sie mit Hilfe der *Music Morphing*-Techniken von Wooller & Brown (2005) neu kombiniert. Durch die vertikale Position des Würfels werden weitere Effekte gesteuert. Einen der ersten Tabletop-Ansätze präsentieren Jordà, Kaltenbrunner, Geiger & Bencina (2005), Kaltenbrunner, Jordà, Geiger & Alonso (2006), Jordà, Geiger, Alonso & Kaltenbrunner (2007) mit dem *ReacTable*. Hier werden Synthesizer- und Sequencer-Komponenten durch Tangibles repräsentiert, die miteinander verbunden und bezüglich verschiedener Parameter (Tonhöhe, Lautstärke, Klangfarbe, Effekte etc.) gesteuert werden können.

Eine der großen Stärken von Tabletop-Interfaces ist, dass es mehreren Anwendern erlaubt, zugleich an der Gestaltung der Musik mitzuwirken. Das Musi-

⁶(Chapel (2003), Seite 50.

zieren war schon immer eine kollaborative Angelegenheit, die Interaktion zwischen den Musikern und gegenseitige Inspiration fordert und fördert. Musikspiele können dabei die Rolle von Instrumenten oder von eigenständigen Musikern übernehmen und den Benutzern eine Plattform zum Experimentieren und Umsetzen von eigenen Ideen bieten. Nichtmusiker werden dabei auf spielerische Art und Weise an Prinzipien der musikalischen Gestaltung (z.B. serielle Komposition), der Klangsynthese und des Arrangements herangeführt. Dieses große künstlerische und pädagogische Potential wird gerade durch die Interaktivität eröffnet, die den Benutzer nicht zum passiven Zuhörer sondern zum aktiven Teilnehmer werden lässt. Die große Popularität von Musikspielen sollte dazu ermutigen, dieses Potential weiter auszuschöpfen. Wie das aussehen kann, hat dieser Abschnitt gezeigt und damit für die in den vorangegangenen Kapiteln vorgestellten Techniken ein lohnendes Anwendungsfeld auch über die aktuellen Anwendungen hinaus aufgezeigt.

Schlussbemerkungen

INTERAKTION mit Musik muss nicht beschränkt sein auf den eigentlichen Akt des Musizierens, der nur Musikern vorbehalten ist, oder auf spezielle Musikbearbeitungsprogramme. Ebenso ist die Interaktion auf visueller Ebene nicht nur innerhalb von Bildbearbeitungsprogrammen möglich und man muss auch kein Zeichner oder Graphiker sein, um mit graphischen Benutzeroberflächen umzugehen. Im Zeitalter der interaktiven Medien muss die interaktive Musik eine ebensolche Selbstverständlichkeit werden, wie es die interaktive Graphik heute schon ist. Dafür muss sie die Fesseln der starren Aufnahme überwinden. Interaktive Medien sind nicht der Ort für schlichte Reproduktion. Wer Musik für interaktive Medien schafft, muss ihr die Vergänglichkeit und Wandelbarkeit wieder zurückgeben, die sie im Zeitalter ihrer technischen Konservierung und Reproduzierbarkeit eingebüßt hat, und muss die Musik für die vielfältigen Formen und Möglichkeiten der Nichtlinearität öffnen. In den vorangegangenen Kapiteln wurden die Techniken hierfür systematisch aufgearbeitet und einige von ihnen auch entwickelt.

Als eines der schillerndsten und vielseitigsten Anwendungsszenarien gelten Video- und Computerspiele. Kapitel 2 hat die Vor- und Entwicklungsgeschichte der Spielemusik nachvollzogen. In der Gegenüberstellung mit der Entwicklungsgeschichte der Filmmusik ließen sich Parallelen erkennen, sowohl was ihre technische Entwicklung angeht als auch ihre funktionelle Anwendung. In beiden Fällen entwickelte sich die Musik von der schlicht unterhaltenden Begleitscheinung hin zu einem narrativen Werkzeug, das seine ganz eigenen Inhalte vermittelt und Aufgaben erfüllt. In beiden Fällen emanzipierte sich die Musik als eine eigenständige mediale Ebene, die zu den anderen Ebenen auf vielfältige Weise in Beziehung treten kann und ihr multimediales Zusammenspiel wertvoll bereichert. In interaktiven Medien kommt die Ebene der Interaktion hinzu, die eine Reihe neuer Möglichkeiten eröffnet. Hier tritt die Musik in Beziehung zum Spieler und seinem Handeln. Beide, sowohl Film- als auch Spielemusik, krankten zu Beginn an der technisch unzulänglichen Koordination mit den anderen

Ebenen. Während Filmmusik über diese frühen Probleme längst hinausgewachsen ist, hat die Musik in interaktiven Medien an vielen Stellen noch damit zu kämpfen.

Kapitel 3 hat eine Systematik vorgestellt, nach der sechs grundsätzliche Herangehensweisen an eine wandlungsfähige Musik unterschieden werden können. Das beginnt mit dem sequentiellen und parallelen Arrangement statischer Musikfragmente. Hier ist das musikalische Material noch vorgefertigt und im Einzelnen auch unveränderlich. Veränderlich ist lediglich seine Montage, also die Kombination zeitgleich und zeitlich nacheinander erklingender Fragmente. Bei der ausdrucksvollen Interpretation ist die Musik ebenfalls vorkomponiert, aber die Art, wie sie gespielt wird, also die Mittel, mit welchen der Interpret sie ausdrucksmäßig ganz unterschiedlich einzufärben vermag, sind veränderlich. Mit der Reharmonisierung und der komponierten Überleitung wurden zwei Konzepte vorgestellt, welche die immer noch vorkomponierte Musik auf kompositorischer Ebene editieren. Die letztendliche Abkehr von allem Vorgefertigten repräsentiert die generative Musik.

Jede dieser sechs Herangehensweisen hat ihre eigenen Vorteile, Nachteile und Herausforderungen, denen sie jede für sich unterliegen. Sie schließen sich aber nicht gegenseitig aus und lassen sich miteinander kombinieren. In Kapitel 4 wurde eine Lösung erarbeitet, die das sequentielle und das parallele Arrangement miteinander kombiniert. Aus der Synergie profitieren beide, denn die Vorteile des einen gleichen die Nachteile des anderen aus. Wo das sequentielle Arrangement relativ schwerfällig reagiert, hilft das reaktivere parallele Arrangement aus. Wo das parallele Arrangement musikalisch nur relativ lokal bleibt, kann das sequentielle Arrangement größere musikalische Distanzen überwinden. Kapitel 4 verbindet beide und entwirft einen vollautomatischen mehrdimensionalen Arrangementmechanismus für die Vertonung dreidimensionaler narrativer Umgebungen. Hier wird noch mit vorgefertigtem musikalischem Material gearbeitet, das in sich weitgehend unveränderlich bleibt. Für eine größere Variabilität sind Techniken notwendig, die detailliertere Eingriffe ermöglichen, etwa Timing und Dynamik einer Audioaufnahme zielgerichtet verzerren. Veränderungen auf kompositorischer Ebene, z.B. die Versetzung einzelner Töne in polyphonen Audiosignalen, sind heute allerdings noch sehr problematisch und aufwendig. Hier liegen die praktischen Grenzen von Audioformaten. Auf Basis symbolischer Musikformate lassen sich solche Eingriffe weit einfacher umzusetzen. Die vorgestellte MIDI-Portierung ebnet hierfür den Weg.

Eine weitere Kombination zweier Lösungskonzepte stellte Kapitel 6 mit der Verbindung von parallelem Arrangement und ausdrucksvoller Interpretation vor. Das parallele Arrangement war bislang auf eher mechanische Überblendtechniken angewiesen und interpretatorisch starr. Die ausdrucksvolle Interpretation war darauf beschränkt, ein weitgehend unveränderliches musikalisches Mate-

rial darzubieten. In ihrer Kombination gibt es diese Nachteile nicht mehr und ein noch größerer Raum an musikalischen Wandlungsmöglichkeiten wird aufgespannt. Mit einem Agenten-basierten Lösungsansatz wird das flexible Spiel eines Orchesters simuliert, das in Echtzeit auf die Anweisungen eines »virtuellen Dirigenten« reagiert. Die Resultate sind in den meisten Fällen überzeugend, haben aber auch die Grenzen des Ansatzes aufgezeigt. Die liegen dort, wo sehr große musikalische und interpretatorische Unterschiede überbrückt werden müssen und die Überleitung in sehr kurzer Zeit durchgeführt wird bzw. werden muss. Es fehlt ein Maß, das die zu überbrückende Differenz angibt. Auf Grundlage dessen wäre es denkbar, den notwendigen Aufwand, der in eine hinreichend ausführliche Überleitung investiert werden muss, zu bestimmen und in die Planung und Umsetzung der Überleitung mit einfließen zu lassen.

Solch ein *reaktives Orchester*, wie in Kapitel 6, ist freilich schon weit jenseits dessen, was mit vorgefertigten Musikeinspielungen möglich ist. Es werden formale Beschreibungen für die charakteristischen Merkmale des ausdrucksvollen Spiels benötigt, so dass sie maschinell adäquat nachgebildet und flexibel bearbeitet werden können. Diese Grundlagenarbeit leistete Kapitel 5. Timing, Dynamik und Artikulation gelten als die drei Säulen der ausdrucksvollen Interpretation. Für sie wurde eine umfangreiche Phänomenologie ausgearbeitet und durch mathematische Modelle umgesetzt. Diese kamen nicht nur in Verbindung mit dem *reaktiven Orchester* zum Einsatz, sondern konnten sich auch als Werkzeug in einigen musikwissenschaftlichen und -psychologischen Studien bewähren. Es wurde gezeigt, dass die Modelle wesentliche Merkmale menschlicher Interpretationen plausibel nachbilden. Es sind nichtsdestotrotz Approximationen. Beim Menschen lassen sich mehr Details beobachten, die von anderen Einflussfaktoren herrühren, wie etwa der Motorik beim Spiel eines Instruments, der Kommunikation im Ensemblespiel. Bis heute sind nicht alle Ursachen bekannt. Die vorgestellten Modelle bilden aber bereits jetzt einen umfangreichen Werkzeugkasten, mit dem Interpretationen analysiert und synthetisiert werden können.

Eine Reihe weiterer Anwendungen stellte Kapitel 7 vor. Es schlägt den Bogen zurück zu den Spieleapplikationen und thematisiert das Feld der diegetischen Musik. Diese bringt im interaktiven Kontext einige besondere Möglichkeiten und Herausforderungen mit. Es wurde eine Systematisierung der Erscheinungsformen von diegetischer Musik in interaktiven Medien vorgestellt, Design-Prinzipien erarbeitet und gezeigt, wie mit den verschiedensten Techniken das große Potential einer interaktiven Musik realisiert werden kann.

Interaktive Musik ist keine Unmöglichkeit, auch für Nichtmusiker, und musikalische Interaktivität erschöpft sich nicht mit den Funktionalitäten eines CD-Players (Play, Skip, Pause). Kapitel 7 hat gezeigt, wie vielgestaltig sie sein kann. Eine Reihe technischer Lösungen dafür wurde entwickelt und in dieser Arbeit

vorgestellt. Die Systematik in Kapitel 3 gibt einen Eindruck davon, dass noch ein großes Spektrum weiterer Ansätze offen ist. Sie alle dienen dazu, eine lebendige wandelbare Musik zu schaffen, welche der Lebendigkeit und Wandelbarkeit des interaktiven Mediums gewachsen ist. Hier muss Musik wieder eine vergängliche Erscheinung sein, woraus ihre Lebendigkeit erwächst. Uns bleibt dann die Zeit ihrer Aufführung, um aufzupassen, hinzuhören, zu erleben und *mitzumachen*.

Narrative Funktionen der Musik: Ergänzung

A.1 Nach Zofia Lissa

Zofia Lissa (1965) formulierte die folgenden achtzehn Funktionen, die Musik im Film erfüllen kann oder aber in Verbindung mit der Musik stehen.

1. Musikalische Illustration Illustrative Klangstrukturen entstehen dadurch, dass gewisse allgemeine Gestaltqualitäten von auditiven Erscheinungen auf das musikalische Material übertragen werden. Es können auch visuelle Bewegungserscheinungen als Designate illustrativer Motive vorkommen, besonders deutlich in Zeichentrickfilmen ausgeprägt – daher etablierte sich der Begriff *Micky Mousing*.

2. Musikalische Unterstreichung von Bewegungen Die musikalische Stilisierung visueller Bewegungsabläufe führte Lissa bereits im Zusammenhang mit der musikalischen Illustration (1.) ein und bespricht sie in dieser Kategorie differenzierter. Der Charakter der visuellen Bewegung wird mit Mitteln der Dynamik, Rhythmik, Agogik, Melodik, Klangfarbe usw. musikalisch nachgeahmt. Hier ist der Synchronismus, wie Eisenstein (1940) ihn beschreibt, das maßgebende Prinzip. Lissa bemerkt weiterhin, dass, obwohl die Verbindung von Bild und Musik hier am stärksten und die Musik am wenigsten autonom ist, sie aber »in dieser Funktion [...] am lockersten der Dramaturgie des Filmwerks verbunden«¹ ist! Aus der heutigen Praxis lässt sich zu den bisherigen Schilderungen noch die Akzentuierung ergänzen: Gilt es, Bewegungen, Filmschnitte und Kamerafahrten besonders zu betonen, werden sie mit musikalischen Akzenten (z.B. betonte

¹Lissa (1965), Seite 124.

Taktzeiten, melodische Ziele, Crescendoziele) synchronisiert. Es ist gewissermaßen eine Unterstreichung, ein musikalisches Ausrufezeichen.

3. Musikalische Stilisierung realer Geräusche Viele visuell dargestellte Erscheinungen haben mit ihnen verbundene spezifische auditive Eigenheiten: Türen knarren, Bienen summen, Wasser plätschert, Vögel zwitschern usw. Diese können in ihrer realistischen Gestalt wiedergegeben oder musikalisch stilisiert werden. Dabei werden ihre strukturellen Eigenheiten, die Gestaltqualität, im Sinne der musikalischen Illustration auf klangliche Strukturen übertragen. Da das Bild das visuelle Designat ergänzt und dadurch die subjektive Interpretation des Hörers bereits eindeutig fokussiert, kann die Stilisierung sehr großzügig ausfallen. Sie kann sich weiter vom Prototyp entfernen und dabei besondere Funktionen erfüllen, die dem nackten Geräusch vorenthalten bleiben. Sie kann zur Grundlage eines längeren Ablaufes werden, sich musikalisch entwickeln und das Geräusch emotional aufladen. Auch die Verschmelzung mit dem Geräuschvorbild tritt häufig auf: Aus dem rhythmischen Rattern eines Eisenbahnzuges kristallisiert sich die Musik heraus, die zur Grundlage für eine ganze Szene wird.² Oder die Musik geht wieder in das Geräusch über, um unmerklich zu verklingen.

4. Musik als Repräsentation des dargestellten Raums Die Bühne des Theaters, der Rand der Leinwand, der Rahmen des Monitors, das alles sind Bildbegrenzungen. Was innerhalb zu sehen ist, ist nur ein Ausschnitt eines größeren Ganzen (ein Zimmer, ein Haus, eine Stadt, usw.). Die unerbittliche Grenze des Bilderrahmens lässt nichts jenseitiges zu. Die Welt endet am Rand des Bildes. Schon in Theater und Oper hatte Musik die Aufgabe, über diese Grenze hinaus zu weisen. Sie macht den Bilderrahmen unscharf, weicht ihn auf und zeigt, dass die Welt dort noch weitergeht. Die zuvor vorgestellten Typen von Funktionen (1. bis 3.) können hierzu eingesetzt werden, ohne dass ihr jeweiliges Designat im Bild gezeigt wird. Es ist eben gerade nicht sichtbar, aber zu hören. Hier wird mit der Assoziationsfähigkeit, Hörgewohnheit und Hörkultur des Publikums gearbeitet. Darüber hinaus kann Musik gewisse qualitative Eigenschaften des Raumes vermitteln. Durch bestimmte Satztechniken kann ein Gefühl von Weite oder Enge evoziert werden. Das kann objektiv auch das Bild vermitteln. Aber der subjektive Eindruck, wie eine Person den Raum empfindet (frei oder erdrückend), lässt sich nur durch die Hinzunahme musikalischer Mittel erzeugen. Das Bild kann diesem Eindruck sogar entgegenstehen, ohne dass sich für den Zuhörer qualitativ etwas ändert. Lissa resümiert: Bei dieser Form des Kontra-

²Die musikalische Stilisierung des Eisenbahngeräusches findet bspw. im Film *The Peacemaker* (deutscher Titel *Projekt: Peacemaker*), vertont von Hans Zimmer (1997), Anwendung.

punkts werden »visuelle Wahrnehmungs-Vorstellungen unter Vermittlung auditiver Wahrnehmungs-Vorstellungen um visuelle Phantasie-Vorstellungen ergänzt«³ und so manches rein visuelle filmisches Mittel überflüssig.

5. Musik als Repräsentation der dargestellten Zeit Musik kann mit Hilfe bestimmter stilistischer Mittel eine damit verbundene Zeitassoziation hervorrufen. Mit den musikalischen Mitteln des 15./16. Jahrhunderts kann beim musikkulturell konditionierten Publikum der Verweis auf Renaissance und Barock abgerufen werden. Deutlich synthetische Klänge und ungewohntere Klangstrukturen können einen hohen Grad an Technisierung und Fortschritt suggerieren und als »Zukunftsmusik« verstanden werden. So ist es bspw. möglich, der Inszenierung eine Aura von Authentizität zu geben, oder im Kontrast zu den Bildern auf eine andere Zeit(-epoche) zu verweisen.

Mit Musik lässt sich die Zeit aber nicht nur anzeigen, was Lissa beschreibt. Spätere Autoren zum Thema Filmmusik und damit verwandten Themen, wie Kungel (2004), Schneider (1990) und Wingstedt (2004), kennen die Musik auch als Mittel, mit welchem das Zeitempfinden des Publikums beeinflusst werden kann. Mit vorrangig rhythmischen und metrischen Mitteln kann die Zeit relativiert werden. Eine dichte *metrische Prägnanz*⁴ wirkt drängelnd, beschleunigend. Liegen die Rhythmus-gebenden Impulse weit auseinander, wird Ruhe ausgestrahlt; die Zeit wird gedehnt, der Augenblick versinnlicht. Tatsächlich ist das eigentliche Musiktempo weniger bedeutsam für diesen Effekt als die Dichte der Rhythmus-gebenden Impulse. Auch das Verhältnis zu Bildschnitten und Bewegungen im Bild hat Einfluss auf die wahrgenommene Zeitgeschwindigkeit, da die innermusikalischen Proportionen unterbewusst in Verhältnis zu Schnittrhythmus und Bewegungen gesetzt werden. Ein besonderer Effekt vermag das Verrinnen der Zeit sogar kurz anzuhalten: Eine stetige Steigerung von Tempo, Rhythmik, Instrumentation und Lautstärke gefolgt von einem plötzlichen Innehalten auf einem einzigen Ton oder einer Pause erzeugt ein Gefühl von Zeitlosigkeit und Losgelöstheit, ein »Verharren im ›Jetzt‹«⁵. Im Allgemeinen bleibt dieser Effekt den besonders dramatischen Situationen vorbehalten.

6. Deformation des Klangmaterials Mit akustischen Effekten, wie Hall und Echo, werden besondere atmosphärische Eindrücke erzielt, die der Absicht der Komposition zuträglich sein können. Viel Hall kann bspw. das Gefühl von Weite, Offenheit oder Leere intensivieren. Die Deformation des Klangmaterials beinhaltet auch die Verfremdung von Klängen, wodurch der Klangcharakter eines

³Lissa (1965), Seite 144.

⁴Die metrische Prägnanz bezeichnet den vorherrschenden Abstand Rhythmus-gebender Impulse.

⁵Kungel (2004), Seite 178.

Musikstücks sehr unterschiedliche Facetten annehmen kann. Die Verfremdung ist ein beliebtes Mittel für parodistische Zwecke und, um eine Szene surreal erscheinen zu lassen. In *Raumpatrouille*⁶ dient sie der Charakterisierung einer fernen Zukunft.

7. Musik als Kommentar Die Möglichkeiten der Autoren, zu dem Gezeigten Stellung zu nehmen, sind im Film sehr eingeschränkt im Gegensatz z.B. zum literarischen Werk. Sie erschöpfen sich oft schon in der Wahl inszenatorischer Stilmittel, d.h. bestimmter ästhetischer Herangehensweisen. Durch die Kraft ihrer Subtilität ist hingegen die Musik in besonderem Maße als Medium für den Kommentar geeignet. »Äußerst häufig versteckt sich hinter einem emotionalen Kommentar, der die Szene interpretiert und über die im Bild gezeigten Elemente hinausgeht [...] ein moralisches Urteil, eine philosophische Verallgemeinerung, ein kritisches Verhältnis, ein ironischer Unterton.«⁷ Die Musik ist das Sprachrohr der Autoren.

8. Musik in ihrer natürlichen Rolle Dies ist Musik, deren Quelle Teil der fiktiven Welt ist, die also auch von den dortigen Charakteren gehört wird. Hierfür gibt es eine ganze Reihe von Bezeichnungen: Musik im Bild, Szenenmusik, Source Music, diegetische Musik, Musik im On. Sie kann zutiefst inniger Ausdruck des musizierenden Charakters und Bestandteil der Handlung sein (z.B. in Musikfilmen). Oft ist sie jedoch lediglich ein Requisit, nicht mehr als ein Hintergrundgeräusch (Nachtlokal-, Kaffeehaus-, Radiomusik), das der Stimmigkeit der Inszenierung dient.

9. Musik als Ausdrucksmittel psychischer Erlebnisse Die Gestalt psychischer Ausdrucksbewegungen wird auf die klanglichen Strukturen der Musik übertragen. Dazu gehören u.a. der Typ der Bewegung, der Motorik, die Art der Rhythmik (leicht oder schwerfällig), der Agogik (langsam oder rasch), die Tonalität (Dur, Moll, atonal, geräuschbasiert), der Charakter der Melodik (fließend, kantabel oder gebrochen), die Klangfarbe der Instrumente, der Charakter der musikalischen Gattungen, die neben der emotionalen auch eine assoziative Ladung transportieren können (Kirchenchoral, Militärmarsch, Tanzmusik) usw. Die Musik vermittelt allerdings keine konkreten psychischen Erlebnisse, sondern deren allgemeinen Typ. Das Bild konkretisiert den Ausdruck, wie auch Casella & Paiva (2001) beschreiben, und wird zugleich von der Musik verallgemeinert. »Darauf eben beruht u.a. die dialektische Einheit von Musik und Bild im

⁶Der vollständige Titel lautet *Raumpatrouille – Die phantastischen Abenteuer des Raumschiffs Orion*, vertont von Peter Thomas (1966).

⁷Lissa (1965), Seite 158.

Film. [...] Außerdem vertieft die Musik den emotionalen Ausdruck des Bildes, da sie [...] den Zuschauer emotional affiziert, während das Bild nur vorgestellte Gefühle, die dem Filmhelden zugeschrieben werden und ebenso imaginativ sind wie er, hervorrufen kann.«⁸ Darüber hinaus kann die Musik Zeichen psychischer Erscheinungen sein, die mit visuellen Mitteln nicht oder nur begrenzt darstellbar sind und im Sinne der Ergänzung und des Kontrapunkts auch nicht dargestellt werden müssen. Hier tritt die Musik auf als:

1. Zeichen von Wahrnehmungen,
2. Mittel der Repräsentation von Erinnerungen,
3. Widerspiegelung von Phantasievorstellungen,
4. Mittel zur Aufdeckung von Trauminhalten,
5. Mittel zur Aufdeckung von Halluzinationen,
6. Mittel zum Ausdruck von Gefühlen,
7. Zeichen von Willensakten.

»Im Film kann im Prinzip alles, was visuell erfassbar ist, gezeigt werden; im Falle imaginativer psychischer Inhalte können aber nur einige ihrer *Korrelate* gezeigt werden. Ihre innere Dynamik indes kann direkt nur die Musik aufdecken. [...] Ihre dramaturgische, integrierende Rolle ist sehr groß. Sie ist häufig die Grundlage der subjektiven Montage, d.h. jener intellektuellen Operationen des Zuschauers, die für das volle Verständnis der künstlerischen Aussage des Films unerlässlich sind.«⁹

10. Musik als Grundlage der Einfühlung Die Vermittlung emotionaler Inhalte, wie in der vorherigen Kategorie beschrieben, hilft dem Zuschauer, sich in einen gezeigten Charakter hinein zu versetzen. Die Einfühlung kann sich also auf Bestandteile der Szene beziehen, aber auch auf die Szene als Ganzes oder sogar längere Sequenzen. Sie kann sie emotional zusammenfassen, den emotionalen Kommentar bilden und dem Zuschauer helfen, den Ablauf des Ganzen einführend mitzuerleben, also nachzuvollziehen. Dazu zählt auch die Immersion in das Szenarium, das sich hinein versetzt Fühlen in die gezeigte Umgebung, Landschaft, Situation. Hier wirkt die Musik auch der Sprunghaftigkeit der Einstellungen und Schnitte und der Episodenhaftigkeit der Szenen integrierend entgegen. Sie spannt einen großen Bogen, fasst zu größeren zusammengehörenden Komplexen zusammen und gibt der Welt die Kontinuität zurück, die ihr der Bildschnitt nimmt.

⁸Lissa (1965), Seite 175.

⁹Lissa (1965), Seite 176.

11. Musik als Symbol Musik kann Gegenstände, Charaktere, Zustände, Situationen, Milieus usw. symbolisieren. Ihr Erklingen ruft beim Zuschauer das symbolisierte Designat wieder in Erinnerung. Wichtige Grundlage hierfür ist der von Schneider (1990) beschriebene »Semantisierungsprozess«¹⁰. Durch die Kombination mit dem Bild wird beim ersten Erklingen der Musik der entsprechende Assoziationsraum aufgespannt. Später wird das Bild nicht mehr benötigt. Solche Symboliken müssen nicht immer extra eingeführt werden. Viele symbolhafte Musiken haben ihre assoziative Ladung bereits aus dem kulturellen Usus und können beim Publikum als bekannt vorausgesetzt werden. Hierzu zählen beispielsweise Nationalhymnen. Musikalische Stilmittel implizieren häufig geographische, kulturelle und historische Zuordnungen (irische Musik, das österreichische Lied, Pop-Musik der 80er Jahre usw.). Die Melodielehre von Diether de la Motte (1993) schlüsselt einige solcher typischen Eigenschaften auf. Vorsicht ist allerdings geboten, wenn der Film auch in anderen Kulturkreisen zur Aufführung kommen soll. Das kulturelle Allgemeingut kann dort andere Ausprägungen vorweisen und andere Assoziationen wecken.

12. Musik als Mittel zur Antizipation des Handlungsinhalts Häufig eilt die Musik inhaltlichen Entwicklungen voraus und schafft auf diese Weise eine innere organische Verbindung zwischen den sonst episodenhaften Szenen. Dabei kann Spannung und Erwartung erzeugt werden. Überdies bereitet sie den Zuschauer auf bestimmte Inhalte der folgenden Szenen vor. In dieser Funktion geht die Musik auch weit über das bloße Vorwegnehmen von Handlungsinhalten hinaus. Um etwa einen Überraschungseffekt zu provozieren, kann die Musik im Vorfeld eine gegenläufige Entwicklung andeuten.

13. Musik als formal einender Faktor In der Kunst bilden Form und Inhalt eine dialektische Einheit. Die Form unterstützt die Vermittlung des Inhaltes. Bspw. ergibt sich das Versmaß und Reimschema in Gedichten nicht zufällig, sondern wird bewusst gestaltet. In manchen Fällen provoziert der Inhalt auch eine bestimmte Form. Musik ist als Form bildendes Mittel schon seit jeher bekannt. Schon in den frühesten Volksliedern kann man musikalische Formen (Liedformen wie AAB oder ABA) finden, die den Text strukturieren und innere Querbezüge herstellen (de la Motte 1993, Altmann 2001). Im Film ist die Musik als Form bildendes Mittel unerlässlich. Das beginnt mit den Einleitungs- und Schlussmusiken, die eine große Klammer über das gesamte Werk spannen. Die Musik kann die Kapitel wie eine Art Zwischenaktmusik voneinander absetzen, überleiten, auf sie einstimmen. Dadurch, dass Musik meist dramaturgisch begründet zum Einsatz kommt, wichtige Stellen markiert und Ausrufezeichen setzt, spiegelt sie

¹⁰Schneider (1990), Seite 84.

das dramaturgische Skelett des Filmes wider. Durch stilistische und motivische Techniken werden inhaltliche Querverbindungen ermöglicht. Dem Zuschauer können so bewusst oder unbewusst Zusammenhänge vermittelt werden, die allein mit visuellen und textuellen Mitteln nicht oder nicht überzeugend darstellbar sind. Die Musik wird zu einem integrierenden Faktor, der die einzelnen Teile des Films zu einem größeren Ganzen verbindet.

14. Mehrfunktionalität und Mehrschichtigkeit der Filmmusik Die bisherigen Erläuterungen lassen bereits erkennen, dass die Musik nicht auf ausschließlich eine der beschriebenen Funktionen beschränkt bleiben muss. Sie erfüllt oft mehrere zugleich. Hinzu kommen die anderen auditiven Schichten (Geräusch und Rede), die auch wieder verschiedene Funktionen erfüllen können. Je nach Intention und Situation kann eine Schicht dominierender sein als andere. Innerhalb dieses komplexen Gewebes kann die Musik wieder parallel oder kontrapunktisch wirken. Wo in der Literatur, im Theater und in frühen Filmen die zeitliche Sequentialität vorherrschte (die Inhalte werden nacheinander präsentiert), ermöglicht diese Mehrschichtigkeit nun eine kondensierte Form, in der verschiedene Schichten simultan agieren und sogar zueinander in Bezug gesetzt werden können. Das kann zur Dynamisierung des dramaturgischen Ablaufs genutzt werden und bereichert die Filmtechnik um vormals ungeahnte künstlerische Mittel.

15. Funktionen der Geräuscheffekte und

16. Funktion der Rede im Film Geräusch und Rede sind von der Musik unabhängige auditive Schichten, die sehr direkte eindeutige Informationen transportieren. Oft wird von einem konkurrierenden Verhältnis dieser Schichten untereinander gesprochen. Geräusch und Rede seien wichtiger, weshalb die Musik grundsätzlich leiser zu sein habe. Tatsächlich sind sie eher miteinander verflochten denn konkurrenz. Selbst die Sprache kann in den Hintergrund treten und dem Geräusch oder der Musik die Bühne freigeben. Dadurch kann die Aufmerksamkeit des Zuschauers auf sonst nur beiläufige subtile Einzelheiten (z.B. Mimik und Gestik) gelenkt werden, die hierbei etwa symbolhafte Qualität erhalten können.

17. Funktion der Stille im Film Lissa beschreibt die Stille als Pause, welcher auditive Erscheinungen vorangegangen sind oder folgen. Die Stille kann sich auf nur eine auditive Schicht erstrecken (z.B. die Abwesenheit jeglicher Hintergrundgeräusche) aber auch auf alle. Sie kann interpunktiven Charakter haben,

d.h., das Vorher und Nachher voneinander trennen. Sie kann mit der Intention gesetzt sein, das Weiterdenken des Zuschauers zu provozieren, die Stille eigenständig auszufüllen. Sie kann dramaturgisches Mittel sein, um Spannung auf das zu erwartende aufzubauen. Im Abschnitt über die Funktion der Musik als Repräsentation der dargestellten Zeit (5.) wurde ihre Bedeutung für den Effekt des Zeitanhaltens bereits beschrieben. Ebenfalls über Lissas Schilderungen hinausgehend darf nicht unerwähnt bleiben, dass die Musik nicht ständig spielen sollte, um ihr dramaturgisches Potential nicht zu verbrauchen. In sparsamer ausgewogener Dosierung ist der Musikeinsatz wirklich etwas besonderes, das der ganzen Szene eine besondere spezifische Betonung verleiht und ihre Bedeutung für das Gesamtwerk hervorhebt. Diese Wirkung geht verloren, wenn ständig Musik spielt.

18. Nichtfunktionelle Zuordnung der audio-visuellen Schichten im Tonfilm

In dieser Kategorie vereint Lissa diejenigen auditiven Erscheinungen, die sich durch keine der bisherigen Funktionen qualifizieren lassen. Dabei handelt es sich hauptsächlich um Erscheinungen, die aus ästhetischen Gründen Einzug gefunden haben, um die innermusikalische Konsistenz sicher zu stellen, ein allgemeines Gestaltungsprinzip kohärent fortzuführen oder fließend überzuleiten.

A.2 Nach Johnny Wingstedt

Die narrativen Funktionen von Musik im Film sind nach Wingstedt (2004) wie folgt klassifiziert.

Emotionale Klasse Die emotionalen bzw. emotionalisierenden Funktionen der Musik sind am häufigsten anzutreffen, meist in Kombination mit Funktionen anderer Klassen. Zu den emotionalen/emotionalisierenden Funktionen zählen:

- Aufdecken und Beschreiben der Gefühle von Charakteren,
- Darstellen von Beziehungen zwischen Charakteren,
- Erzeugen von (insbesondere emotionaler) Glaubwürdigkeit,
- Täuschen des Publikums (durch falsche Glaubwürdigkeit),
- Vermitteln einer allgemeinen Stimmung und
- Provozieren von Vorahnungen.

Musiken, die Funktionen dieser Klasse bedienen, helfen dem Betrachter, sich in Personen, Umgebungen und Situationen einzufühlen.

Informative Klasse Die Funktionen, die den drei Unterkategorien dieser Klasse angehören, eint die Absicht, Informationen zu vermitteln, nicht Emotionen. Wie Schneider (1990) bereits feststellte und auch Marshall & Cohen (1988) und Cohen (1998) in ihren eigenen Untersuchungen belegen, wird die Bedeutung von Musik mit dieser Funktion sehr konkret und eindeutig wahrgenommen. Entsprechend mächtig ist diese Musik als narratives Werkzeug.

1. Vermitteln von analogen Inhalten: Die Funktionen dieser Kategorie zielen darauf ab, inhaltliche Zusammenhänge, Analogien und visuell nicht darstellbare Inhalte zu vermitteln. Dazu gehört:
 - Verdeutlichen der Verwandtschaft von Situationen,
 - Vermitteln unausgesprochener Gedanken und
 - Bekräftigen der subjektiven Interpretation einer Situation.
2. Evozieren von assoziativen Inhalten: Von den assoziativen Ladungen, die viele Musiken bereits aus dem kulturellen Usus mitbringen, war bereits die Rede. Hier sind Funktionen vertreten, wie:
 - Evozieren von Zeitepochen,
 - Evozieren eines kulturellen Umfeldes und
 - Anzeigen eines sozialen Status' oder Milieus.
3. Wiedererkennung: Musikalische Inhalte können mit außermusikalischen Phänomenen assoziiert werden (z.B. mit Personen, Gegenstände, Situationen). Dieser Wiedererkennung muss zunächst der Semantisierungsprozess vorausgehen. Ein häufiges Mittel hierfür ist die Verwendung von Leitmotivik im Zusammenhang mit dem visuell gezeigten Designat, das später auch ohne die visuelle Repräsentation wieder assoziiert werden kann.

Beschreibende Klasse Anders als die informative Klasse, die eher passiv Assoziationen und Analogien vermittelt, beschreiben die Funktionen dieser Klasse hauptsächlich physikalische Phänomene in der Szene, nehmen aktiv am Geschehen teil.

1. Beschreibung des Settings: Die Funktionen dieser Kategorie entsprechen denen von Lissas Kategorie 4 – Musik als Repräsentation des dargestellten Raums. Sie weisen über den Bildrand hinaus und vermitteln:

- abstrakt die physikalische Atmosphäre (z.B. Tag, Jahreszeit) und
 - konkret das physikalische Setting (z.B. Ozean, Wald).
2. Beschreibung physikalischer Aktivität: Ebenfalls bereits bei Lissa bekannt ist die musikalische Bewegungsillustration (Kategorie 2). Wingstedt unterscheidet zusätzlich zwischen:
- direkter Bewegungsillustration (Micky Mousing; die Bewegung ist visuell sichtbar) und
 - konzeptioneller Bewegungsillustration (die Musik impliziert Bewegung, die nicht sichtbar sein muss; z.B. hinter einer verschlossenen Tür nähern sich Schritte).
 - Wenn Bewegungen sehr schnell sind, können dem Auge Details entgehen. Durch Musik können diese Details herausgearbeitet und hörbar gemacht werden. Das Ohr ergänzt das langsamere Auge (Chion 1994).

Führende, leitende Klasse Musik dieser Klasse lenkt das Auge, die Gedanken, Meinung und Erwartung des Zuschauers. Sie ist ein Werkzeug zum Hervorheben und Verstärken visuell schwacher Details. Solche Musik kann auch den Zugang zu einem tieferen Verständnis einer Szene oder des Gesamtwerkes erleichtern.

1. Indikative Kategorie: Hier agiert Musik als eine Art Scheinwerfer, der die Aufmerksamkeit des Betrachters auf Elemente in der Szene, Bewegungen oder Teile eines Dialogs lenkt und Betonungen bzw. Schwerpunkte setzt. In Kombination mit emotionalen und beschreibenden Funktionen können darüber hinaus qualitative Eigenschaften (Schönheit, Gefahr, Stärke usw.) vermittelt werden. Zu diesen Funktionen zählt:
- die Richtung der Aufmerksamkeit lenken und
 - die Fokussierung auf Details.
2. Maskierende Kategorie: Neben dem Überdecken ungewollter Geräusche (Filmprojektor beim Stummfilm) zählt zu diesen Funktionen:
- das Verstärken oder Überdecken (Ausmaskieren) schwacher Elemente (z.B. schlechtes Schauspiel) und
 - das Aufblähen schwacher Inhalte (Werbe-Jingles).

Zeitliche Klasse Musik wird oft auch bezeichnet als Zeitkunst, als gestaltete Zeit. Sie kann ohne die zeitliche Dimension nicht existieren. Nahe liegender Weise ist sie gerade für zeitbezogene Funktionen prädestiniert.

1. Musik wirkt dem Bildschnitt und der daraus resultierenden Episodenhaftigkeit entgegen und sorgt für Kontinuität:
 - im engeren Sinne (Szenenübergänge),
 - längerfristig (szenenübergreifend, längere Sequenzen) und
 - allumfassend (über das Gesamtwerk).
2. Struktur und Form definieren: Musik hat auf die zeitliche Abfolge der Szenen eine strukturgebende Wirkung (durch motivische Querbezüge, musikalischen Gestus etc.). Dadurch werden Funktionen möglich, wie:
 - Gestaltung der Form der Erzählung,
 - Beeinflussung der Wahrnehmung von Zeit und Geschwindigkeit,
 - Beeinflussung der Wahrnehmung von Form, Länge und Tempo des dramaturgischen Flusses,
 - Aufbauen von Erwartungen und Vorahnungen und
 - das formale Abrunden einer Szene und des Films als Ganzes.

Rethorische Klasse Zusätzlich zu ihren semiotischen Funktionen, kann die Musik der Szene auch eine weitere eigenständige und häufig verallgemeinernde Inhaltsschicht hinzufügen. Diese wirkt im Sinne eines Kommentars, einer Auseinandersetzung mit dem gezeigten und als Mittel zur Beeinflussung der Interpretation des Zuschauers. Die große suggestive Kraft des Films liegt in dieser Musik begründet, die das subtile Vermitteln solcher Inhalte überhaupt erst ermöglicht. Zu den rethorischen Funktionen zählt:

- das Vortreiben und Kommentieren der Erzählung,
- das (Ver-)Urteilen – ein philosophisches und politisches Statement geben,
- sich (und somit auch den entmündigten Zuschauer, Schneider 1990) für eine Seite entscheiden,
- das Verfremden (etwas aus einem anderen Blickwinkel betrachten) und
- das Beeinflussen, wie die Bedeutung eines bestimmten Wortes oder Satzes wahrgenommen wird.

Wie die meisten Autoren zum Thema Filmmusik führt auch Wingstedt (2004) weiterhin die Unterscheidung zwischen diegetischer und nichtdiegetischer Musik an. Auch die bei Lissa aufgeführte 14. Kategorie (Mehrfunktionalität und Mehrschichtigkeit der Filmmusik) bespricht er. Darüber hinaus verweist er noch auf die von Cohen (1998) genannte ästhetische/ästhetisierende Funktion der Musik, die ohnehin offensichtlich und allgegenwärtig ist, schließlich stellt sie den allgemeinen Rahmen, in welchem musikalische Vorgänge konzipiert werden (z.B. stilistische Grundentscheidungen). Genau genommen ist dies keine abstrakt zu betrachtende Funktion, da sie sich aus der Gesamtanlage und somit aus der Planung der übrigen Funktionen ergibt.

iMuse im Überblick

Einer der Höhepunkte der automatischen Arrangementansätze ist die *Interactive Music Streaming Engine* (kurz: *iMuse*) von Land & McConnell (1994). Hier gibt der Komponist zu jedem Musikstück zusätzliche Angaben hinzu, welche der Engine Anhaltspunkte für ein musikalisch sinnvolles Arrangieren in Echtzeit liefern. *iMuse* unterstützt sowohl verschiedene Audioformate (einschl. CD-Audio) als auch MIDI.

Audiodaten können sequentiell aneinander gereiht werden. Sie können mehrspurig synchron arrangiert und abgemischt werden. Die Wiedergabe kann zwischen den Audiodaten hin und her springen und dabei beliebige Positionen innerhalb der Musikstücke anspringen, nicht nur den Beginn einer Datei. Diesbezüglich ist *iMuse* also bereits mächtiger als viele seiner jüngeren Konkurrenten. Trotzdem bleiben die Eingriffsmöglichkeiten in Audiodaten natürlich nur begrenzt. Seine wahren Stärken spielt *iMuse* im MIDI-Bereich aus.

Auch hier setzt der Komponist bestimmte Marken im Stück, die über die Schnittstelle der Engine aktiviert werden können. Kommt die Wiedergabe an einer aktivierten Marke an, wird nahtlos zu einer anderen Position im Stück oder in einem gänzlich anderen Stück gesprungen. Die Marken können auch als Auslöser für weitere Ereignisse verwendet werden: Ein-, Ausschalten von MIDI-Kanälen, Instrument-, Panorama-, Stimmungsveränderungen und Transposition einzelner MIDI-Kanäle, Lautstärkeveränderung einzelner oder aller MIDI-Kanäle und Tempoänderungen.

Die Möglichkeiten, die *iMuse* zum Arrangieren und Adaptieren bietet, sind außerordentlich vielseitig. Das sequentielle Arrangieren der Musik ermöglicht die Einbeziehung von vorkomponierten Überleitungen. Das geht sogar so weit, dass mehrere verschiedene Überleitungen für den gleichen Wechsel zwischen zwei Musikstücken zum Einsatz kommen können. Dies sei an einem Beispiel illustriert:

- Das laufende Musikstück *A* hat in den ersten acht Takten die Tonart C-Dur zu Grunde liegen. Die nächsten acht Takte stehen in As-Dur. Der Schlussteil steht in c-Moll.
- Das Musikstück *B*, zu dem gewechselt werden soll, beginnt in B-Dur und wechselt nach vier Takten zu c-Moll.
- Tritt das den Musikwechsel auslösende Ereignis im C-Dur-Teil von *A* auf, wird eine Überleitung dazwischen gesetzt, die über F-Dur moduliert.
- Tritt das den Musikwechsel auslösende Ereignis hingegen im As-Dur-Teil von *A* auf, wird eine andere Überleitung verwendet, die über Es-Dur moduliert.
- Aus dem c-Moll-Teil von *A* heraus kommt keine Überleitung zum Einsatz. Statt dessen wird direkt in den c-Moll-Teil von Musikstück *B* gesprungen, der B-Dur-Teil übersprungen.

Dieses verhältnismäßig schlichte Beispiel demonstriert die Mächtigkeit und die Vielseitigkeit, die allein dieser eine Aspekt von *iMuse* bietet. Die Möglichkeit, mehrere auf konkrete musikalische Situationen spezialisierte Überleitungen zu verwenden, erlaubt einen in sich äußerst stimmigen musikalischen Ablauf, der keinerlei Inkonsistenzen erkennen lässt. Der Aufwand, welcher in die Komposition der Überleitungen investiert werden kann, kennt jedoch ebenfalls nach oben hin keine Grenzen. Es wäre bspw. möglich, eine spezialisierte Überleitung für den Wechsel nach jedem einzelnen Takt zu schreiben. Je komplexer das virtuelle Szenarium wird, desto weiter steigt der Aufwand: Die möglichen Zielmusikstücke, in die gewechselt werden kann, können von größerer Zahl sein. Auch die Rückrichtung – der Wechsel zurück in die Ausgangsmusik – muss normalerweise möglich sein. Der Komponist muss selbst ein geeignetes Maß finden und dies in der Anlage seiner Musik berücksichtigen, so dass möglichst wenig Überleitungen an möglichst vielen Stellen funktionieren.

Hilfreich können auch die zahlreichen weiteren (oben aufgezählten) Funktionen sein. Dazu seien einige Beispiele genannt:

- Hinzunehmen und Ausschalten einzelner Stimmen (z.B. verschiedene Variationen einer Melodie) oder der Wechsel der Instrumentation (z.B. Klavier statt Streicher) können den Charakter ein und der selben Komposition unterschiedlich färben, ohne dass ein Wechsel zu einer gänzlich anderen Musik von Nöten ist. Es fehlen hierfür aber Überleitungstechniken; die MIDI-Kanäle werden einfach stumm geschaltet.
- Durch Transposition der Zielmusik kann eine aufwendige Modulation umgangen werden.

- Auch die Abspielgeschwindigkeit kann geändert werden. Auch hierfür fehlen Überleitungstechniken.

iMuse ist mit Abstand der am weitesten entwickelte Lösungsansatz, der in dem nichtwissenschaftlichen Umfeld der Computer- und Videospielevertonung entstanden ist. Vom Komponisten wird allerdings ein gewisses Maß an technischer Einsicht besonders im Umgang mit MIDI-Daten und Einarbeitung in die vielfältigen Möglichkeiten der Engine verlangt. Auch lizenzrechtliche Belange schränken die größere Verbreitung und Nutzung der Engine ein. Bezeichnender Weise kommen bisher die meisten und ausgeklügeltsten Spielemusiken von den Autoren Land & McConnell selbst und ihrem zeitgleich bei *LucasArts* beschäftigten Kollegen Clint Bajakian.

Stilistischer Pluralismus

Der Großteil der kompositorischen Werke bis heute bewegt sich innerhalb eines jeweiligen Zeit-, Personal- oder werkspezifischen Stiles, verlässt diesen nicht und bleibt damit immer in sich stimmig und ästhetisch konsistent. Einen deutlich weiteren Raum des musikalischen Stilbegriffes erforschte u.a. der Komponist Bernd Alois Zimmermann (*1918, †1970) mit seinem kompositorischen Prinzip des »zeitlichen und stilistischen Pluralismus«¹.

Kann ein musikalisches Werk mehrere Stile beherbergen, ohne an Zusammenhalt zu verlieren? Zimmermanns pluralistische Auffassung kennzeichnet die zeitlich parallele »Mehrgleisigkeit verschiedener Metren, Rhythmen und Tonalitäten bis hin zu collageartigen Schichtungen musikalischer Zitate«². Der fremde Stil wird in das Gefüge des im Werk dominierenden Stiles eingewoben. Während einige Stimmen eine in sich schlüssige stilistische Einheit bilden, ergänzen andere Stimmen den Satz um Mittel eines oder mehrerer anderer Stile. Jede Stil-schicht ist in sich stimmig, weist aber Bezugspunkte zu anderen Schichten auf, was integrierend wirkt. Beispielhaft dafür steht Zimmermanns Oper *Die Soldaten*. In ihr dominiert die serielle Musik und die Darmstädter Avantgarde. Darin eingewoben finden sich Jazz-Elemente (in Akt 2 Szene 1 und Akt 4 Szene 3) und sogar ein barocker Choralatz zu Heinrich Isaacs Innsbruck-Lied (in Akt 2 Szene 2), das in dem narrativen Zusammenhang der Oper wohl eher den geistlichen Text »O Welt, ich muss Dich lassen« annotieren soll.

Als ein weiteres Beispiel für den stilistischen Pluralismus steht auch die Komposition *INtroVENTION*³ von Michael Denhoff, in welcher er u.a. eine Invention von J. S. Bach mit seinen eigenen Stilmitteln verbindet. Als Vorreiter dieser Strömung gilt Charles Ives (*1874, †1954). Sein Werk *The Unanswered Question*

¹Denhoff (1988), Seite 27.

²Denhoff (1988), Seite 36.

³Die Partitur von *INtroVENTION* veröffentlicht Denhoff auf seiner Webseite: <http://www.denhoff.de> (Stand: März 2011).

wirkt »wie eine Vorausnahme dessen, was Zimmermanns pluraistische Auffassung kennzeichnet«⁴ – gleichermaßen seine Komposition *Central Park in the Dark*. Heute ist das Nebeneinander und Übereinander mehrerer Stile zu einem wichtigen Ausdrucksmittel besonders in der Filmmusik geworden, wo gezielt mit der assoziativen Ladung bestimmter Stile und Stilmittel gearbeitet wird.

Als kompositorisches Prinzip des stilistischen Pluralismus stellt Peter Dickinson (1989) die Stilmodulation vor. Den Begriff Stilmodulation definiert er wie folgt: »That is, the use of different musical styles within a single work in a way which is as calculated as any other element of control.«⁵ Er unterscheidet drei Typen von Stilmodulationen:

Einfache Stilmodulation: Stilwechsel passieren direkt. Die dabei entstehende potpourristische Wirkung ist dem musikalischen Zusammenhalt jedoch oft abträglich.

Progressive Stilmodulation: Stilwechsel passieren in mehreren diskreten Stufen. Die Ausdrucksmittel des einen Stiles werden schrittweise in die des anderen umgeformt. Die Art der Rhythmik kann nach und nach in eine andere transformiert werden. Die harmonische Sprache kann schrittweise überführt werden; usw. Bei ähnlichen oder verwandten Stilen fällt dies naturgemäß leichter und wirkt organischer als bei sehr unterschiedlichen oder gar gegensätzlichen Stilen. Der Prozess des Übergangs muss hier zeitlich umfangreicher disponiert und die Stufen möglichst klein gehalten werden.

Stilvermischung: Dickinson beschreibt diese Art der Stilmodulation als ein im zeitlichen Verlauf abwechselndes Nebeneinander verschiedener Stile auf engstem Raum (z.B. kurze Cluster-Einwürfe in einer barocken Arie). Auch Zimmermanns Form der Polystilistik, in der mehrere Stile zeitgleich erklingen, zählt zur Stilvermischung. Besondere Aufmerksamkeit muss hier der Bindung zwischen den stilistisch eigenständigen Parteien zukommen. Sie kann sich in der großstrukturellen Anlage widerspiegeln, im harmonischen, melodischen/motivischen oder ausdrucksmäßigen Bezug zueinander, in ihrem inhaltlich konnotativen oder kommunizierenden Verhältnis zueinander usw. Ives Vorwort zu *The Unanswered Question* beschreibt Beispiele hierfür: Die einzelnen Stile stehen für an der Szene beteiligte Charaktere, ihr Duktus umschreibt ihre Handlungsweise – die fragende Trompete, die schweigenden Druiden (Streicher) und die anderen Menschen (Flöten), welche erfolglos einer Antwort nachjagen und letztendlich doch bloß die Frage nachäffen.

⁴Denhoff (1988), Seite 36.

⁵Dickinson (1989), Seite 208.

Performance-Übungen

Wie verlaufen Ritardando, Accelerando, Crescendo, Decrescendo? Wie verändern Artikulationsanweisungen die Art, wie eine Note gespielt wird? In einer Studie wurden professionelle Musiker gebeten, eine Reihe von Tempo-, Dynamik- und Artikulationsübungen auszuführen. Sie wurden dabei aufgenommen und die Aufnahmen später ausgemessen (siehe Kapitel 5 Abschnitt 5.1). Die Übungen sind als Ergänzung zu echten Musikaufführungen konzipiert und verzichten bewusst auf jeglichen musikalischen Kontext (wie z.B. Melodistrukturen), der eine bestimmte Merkmalsausprägung forcieren könnte. Die Musiker hatten Tonrepetitionen zu spielen. So konnten die Merkmale sowohl bei richtiger Musik als auch hier unter »neutralisierten« Bedingungen beobachtet werden. Nachfolgend ist das Notenmaterial zu den Übungen dargestellt. Von den Dynamik- und Tempoübungen hatte jeder Musiker jeweils nur eine zu spielen.

Vivace.



Vivace.



Abbildung D.1: Artikulationsübung

Dynamiktest

$\text{♩} = 120$

The musical score consists of six staves of music in a single system. The first staff starts with a tempo marking of quarter note = 120 and a dynamic marking of *f*. The second staff has a dynamic marking of *p*. The third staff has a dynamic marking of *f*. The fourth staff has a dynamic marking of *p*. The fifth staff has a dynamic marking of *f*. The sixth staff has a dynamic marking of *p*. The music is written in a single treble clef with a common time signature. The notes are quarter notes, and there are slurs and hairpins indicating dynamic changes.

Abbildung D.2: Dynamikübung 1

Dynamiktest

$\text{♩} = 120$

The musical score consists of six staves of music in a single system. The first staff starts with a tempo marking of quarter note = 120 and a dynamic marking of *p*. The second staff has a dynamic marking of *f*. The third staff has a dynamic marking of *p*. The fourth staff has a dynamic marking of *f*. The fifth staff has a dynamic marking of *p*. The sixth staff has a dynamic marking of *f*. The music is written in a single treble clef with a common time signature. The notes are quarter notes, and there are slurs and hairpins indicating dynamic changes.

Abbildung D.3: Dynamikübung 2

Tempotest

Musical score for Tempotest 1, consisting of five staves of music in 4/4 time. The first staff starts with 'allegro.' and changes to 'rit.' at measure 2, then 'adagio.' at measure 3. The second staff starts with 'acc.' at measure 1, changes to 'allegro.' at measure 2, 'rit.' at measure 3, and 'adagio.' at measure 4. The third staff starts with 'acc.' at measure 1, changes to 'allegro.' at measure 2, and 'rit.' at measure 3. The fourth staff starts with 'adagio.' at measure 1 and changes to 'acc.' at measure 2. The fifth staff starts with 'allegro.' at measure 1 and changes to 'rit.' at measure 2.

Abbildung D.4: Tempoübung 1

Tempotest

Musical score for Tempotest 2, consisting of five staves of music in 4/4 time. The first staff starts with 'adagio.' at measure 1, changes to 'acc.' at measure 2, and 'allegro.' at measure 3. The second staff starts with 'rit.' at measure 1, changes to 'adagio.' at measure 2, 'acc.' at measure 3, and 'allegro.' at measure 4. The third staff starts with 'rit.' at measure 1, changes to 'adagio.' at measure 2, and 'acc.' at measure 3. The fourth staff starts with 'adagio.' at measure 1 and changes to 'acc.' at measure 2. The fifth staff starts with 'allegro.' at measure 1 and changes to 'rit.' at measure 2.

Abbildung D.5: Tempoübung 2

Tempostudie

Können menschliche Hörer Unterschiede im Verlauf kontinuierlicher Tempoübergänge erkennen? In einer Tempostudie wurde das in Kapitel 5 erarbeitete Modell *Explizite Tempokurven* genutzt, mit dem die Verlaufscharakteristiken von Tempokurven systematisch beeinflusst werden können. Die Teilnehmer an der Studie bekamen 6 Paare von Tempoübergängen vorgespielt, die sich nur in der Einstellung des Parameters i_m unterschieden (siehe Tabelle E.1). Zur besseren Orientierung konnten sie auch den Notentext einsehen und sich gegebenenfalls Eintragungen machen (siehe Notenbild auf der folgenden Seite). Eine vollständige Auswertung der Studie wird in Kapitel 5 Abschnitt 5.3.1 vorgenommen.

Typ	i_m -Einstellungen			korrekte Antworten		
	1. Beispiel	2. Beispiel	Differenz	#	%	Signifikanz
Acccel.	0.7	0.3	0.4	16	88.9	0.001
Rit.	0.5	0.4	0.1	5	27.8	0.096
Acccel.	0.6	0.8	0.2	9	50.0	1.0
Rit.	0.4	0.4	0	8	44.4	0.815
Acccel.	0.4	0.3	0.1	9	50.0	1.0
Rit.	0.3	0.6	0.3	14	77.8	0.031

Tabelle E.1: Die Resultate der Tempostudie.

$\text{♩} = 40$ accel.
 $\text{♩} = 120$ rit.

3

5

7

9

11

13

15 $\text{♩} = 120$
 $\text{♩} = 40$

17 *decresc. al niente*

19

Abbildung E.1: Das Notenmaterial zur Tempostudie.

Literaturverzeichnis

- Aav, S. (2005), *Adaptive Music System for DirectSound*, Master's thesis, University of Linköping, Department of Science and Technology, Norrköping, Sweden.
- AdLib Inc. (1987), 'AdLib Soundcard'.
- Adorno, T. W. (1963), *Der getreue Korrepetitor: Lehrschriften zur musikalischen Praxis*, Vol. 15 of *Gesammelte Schriften*, S. Fischer, Frankfurt (Main), Germany.
- Adorno, T. W. (1978), *Musikalische Schriften I–III*, Vol. 16 of *Gesammelte Schriften*, Suhrkamp, Frankfurt (Main), Germany, chapter *Fragmente über Musik und Sprache*, pp. 251–255.
- Adorno, T. W. & Eisler, H. (1947a), *Composing for the Films*, Oxford University Press, New York, USA.
- Adorno, T. W. & Eisler, H. (1947b), *Komposition für den Film*, reissue april 2006 edn, Suhrkamp.
- Al Kasimi, A., Nichols, E. & Raphael, C. (2007), *A Simple Algorithm for Automatic Generation of Polyphonic Piano Fingerings*, in '10th Int. Society for Music Information Retrieval Conf. (ISMIR)', Austrian Computer Society (OCG), Vienna University of Technology, Vienna, Austria, pp. 131–132.
- Altmann, G. (2001), *Musikalische Formenlehre—Ein Handbuch mit Beispielen und Analysen. Für Musiklehrer, Musikstudierende und musikinteressierte Laien*, 8th revised edn, Schott, Mainz, Germany.
- Alty, J. L., Rigas, D. & Vickers, P. (2005), *Music and speech in auditory interfaces: When is one mode more appropriate than another?*, in 'Proceedings of ICAD 05—Eleventh Meeting of the International Conference on Auditory Display', Limerick, Ireland, pp. 1–7.
- AM3D A/S (2009), 'AM3D', am3d.com.

- Amanita Design (2009), 'Machinarium', Daedalic Entertainment. music by Tomáš Dvořák.
- Ames, C. (1989), 'The Markov Process as a Compositional Model: A Survey and Tutorial', *Leonardo Music Journal* 22(2), 175–187.
- Arons, B. (1992), 'A Review of The Cocktail Party Effect', *Journal of the American Voice I/O Society* 12, 35–50.
- Bach, C. P. E. (1753), *Versuch über die wahre Art das Clavier zu spielen*, faksimile reprint (1994), w. horn, ed. edn, Bärenreiter.
- Bartlett, D. (1996), Physiological Responses to Music and Sound Stimuli, in D. A. Hodges, ed., 'Handbook of Music Psychology', National Association for Music Therapy, IMR Press, San Antonio, Texas, pp. 343–385.
- Bartz, A. (2010), *Vertonung virtueller Stadtmodelle*, Bachelor thesis, Otto-von-Guericke University of Magdeburg, Department of Simulation and Graphics, Magdeburg, Germany.
- Baur, T., Böehler, F., Dieckmann, U., Frieß, E., Günther, H., Meier, S. & Nonnenmann, H.-U., eds (1999), *89 Oberstimmen zu Sätzen des Posaunen-Choralbuchs zum Evangelischen Gesangbuch*, Strube Verlag, Munich, Germany.
- Baur, T., Eckert, F. & Nonnenmann, H.-U. (2007), *Zwo, Drei, Vier: Bläuserschule für Posaunenchöre. Eine musikalische Grundausbildung*, 8th edn, buch & musik, ejw-Service GmbH, Stuttgart-Vaihingen, Germany.
- Bello, J., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. & Sandler, M. (2005), 'A Tutorial on Onset Detection in Music Signals', *IEEE Transactions on Speech and Audio Processing* 13(5), 1035–1047.
- Bello, J., Duxbury, C., Davies, M. & Sandler, M. (2004), 'On the Use of Phase and Energy for Musical Onset Detection in the Complex Domain', *IEEE Signal Processing Letters* 11(6), 553–556.
- Bemani (1999a), 'DrumMania', Konami Digital Entertainment.
- Bemani (1999b), 'Guitar Freaks', Konami Digital Entertainment.
- Bemani (1999c), 'HipHopMania', Konami Digital Entertainment.
- Berlioz, H. & Strauss, R. (1844/1904), *Grande Traité d'instrumentation et d'orchestration moderne (Treatise on Instrumentation)*, Kalmus Edition, New York, USA. english translation by T. Front (1948).

- Berndt, A. (2006), Repräsentation musikalischer Metainformationen, Diploma thesis, Otto-von-Guericke University, Department of Simulation and Graphics, Magdeburg, Germany.
- Berndt, A. (2008), *Liturgie für Bläser*, 2nd edn, Musikverlag Bruno Uetz, Halberstadt, Germany. transl.: Liturgy for Brass.
- Berndt, A. (2009), Musical Nonlinearity in Interactive Narrative Environments, in G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 355–358.
- Berndt, A. (2010), Decentralizing Music, Its Performance, and Processing, in M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 381–388.
- Berndt, A. (2011a), Diegetic Music: New Interactive Experiences, in M. Grimshaw, ed., 'Game Sound Technology and Player Interaction: Concepts and Developments', IGI Global, Hershey, PA, pp. 60–76.
- Berndt, A. (2011b), Musical Tempo Curves, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, University of Huddersfield, Huddersfield, UK.
- Berndt, A. (2012), Nun sich der Tag geendet hat, in F. Plewka, M. Schmeiß, T. Sterzig & J. Häußler, eds, 'Mitteldeutsches Bläserheft'. accepted and in press.
- Berndt, A. & Hartmann, K. (2006), 'Kompositions- und Arrangiertechniken für Musik in interaktiven Umgebungen', poster and exhibition at Game Days 2006.
- Berndt, A. & Hartmann, K. (2007a), Audio-Interactive Counterpoint, in 'Sound, Music and the Moving Image', Institute of Musical Research, University of London, England. invited talk.
- Berndt, A. & Hartmann, K. (2007b), Strategies for Narrative and Adaptive Game Scoring, in 'Audio Mostly 2007: 2nd Conf. on Interaction with Sound', Fraunhofer IDMT, Interactive Institute/Sonic Studio Piteå, Ilmenau, Germany, pp. 141–147.
- Berndt, A. & Hartmann, K. (2008), The Functions of Music in Interactive Media, in U. Spierling & N. Szilas, eds, 'Interactive Digital Storytelling (ICIDS) 2008', Springer, Erfurt, Germany, pp. 126–131. LNCS 5334.

- Berndt, A., Hartmann, K., Röber, N. & Masuch, M. (2006), Composition and Arrangement Techniques for Music in Interactive Immersive Environments, *in* 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute/Sonic Studio, Piteå, Sweden, pp. 53–59.
- Berndt, A. & Hähnel, T. (2009a), Der Computer als Interpret, *in* 'Gespräche um Telemann', Zentrum für Telemannpflege und -Forschung, Magdeburg, Germany. invited talk.
- Berndt, A. & Hähnel, T. (2009b), Expressive Musical Timing, *in* 'Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 9–16.
- Berndt, A. & Hähnel, T. (2010), Modelling Musical Dynamics, *in* 'Audio Mostly 2010: 5th Conf. on Interaction with Sound—Sound and Design', Interactive Institute/Sonic Studio Piteå, ACM, Piteå, Sweden, pp. 134–141.
- Berndt, A. & Knolle, N. (2001), 'Das „Musikalisches Würfelspiel“', an implementation of Mozart's musical dice game. also embedded within the online-role-playing-game Legend of the Green Dragon by Eric Stevens et al., www.atrahor.de, 2006.
- Berndt, A. & Theisel, H. (2008), Adaptive Musical Expression from Automatic Realtime Orchestration and Performance, *in* U. Spierling & N. Szilas, eds, 'Interactive Digital Storytelling (ICIDS) 2008', Springer, Erfurt, Germany, pp. 132–143. LNCS 5334.
- Bernhard, C. (1657), *Von der Singekunst oder -manier*.
- Bernstein, A. D. & Cooper, E. D. (1976), 'The Piecewise-Linear Technique of Electronic Music Synthesis', *Journal of the Audio Engineering Society* 24(6), 446–454.
- Bertin, C., de Ioannes, G. & Sylleros, A. (2010), The Arcontinuo: A Performed-Centered Electronic Musical Instrument, *in* M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 80–87.
- Bortz, J. (2005), *Statistik für Human- und Sozialwissenschaftler*, 6th edn, Springer, Berlin, Heidelberg, Germany.
- Boulangier, R. (2000), *The CSound Book: Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*, MIT Press.
- Bradter, C. (2007), *Ein systematischer Ansatz zur Evaluation von Lautheitsmodellen*, Uni-Edition, Berlin, Germany.

- Bresin, R. (2001), Articulation Rules for Automatic Music Performance, in R. Dannenberg, ed., 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, La Habana, Cuba, pp. 294–297.
- Brown, A. & Kerr, T. (2009), Adaptive Music Techniques, in 'Improvise: The Australasian Computer Music Conf.', Australasian Computer Music Association, Queensland University of Technology, Brisbane, Australia, pp. 26–31.
- Brown, A. R., Wooller, R. W. & Kate, T. (2007), The Morphing Table: A collaborative interface for musical interaction, in A. Riddel & A. Thorogood, eds, 'Proc. of the Australasian Computer Music Conf.', Australian National University Canberra, Canberra, Australia, pp. 34–39.
- Brown, A. & Sorensen, A. (2009), 'Interacting with Generative Music through Live Coding', *Contemporary Music Review* 28(1), 17–29. issue on Generative Music.
- Brown, A., Wooller, R. & Miranda, E. (2009), Interactive Evolutionary Morphing as a Music Composition Strategy, in E. R. Miranda, ed., 'Music As It Could Be: New Musical Worlds from Artificial Life', A&R Editions, Madison, Wisconsin, USA.
- Brown, S., Martinez, M. J. & Parsons, L. M. (2004), 'Passive music listening spontaneously engages limbic and paralimbic systems', *NeuroReport* 15(13), 2033–2037.
- Bräuninger, I. & Blumer, E. (2004), Tanz- und Bewegungstherapie, in W. Rössler, ed., 'Psychiatrische Rehabilitation', Springer, Berlin, Germany, chapter 29, pp. 380–387.
- Bullerjahn, C. (2001), *Grundlagen der Wirkung von Filmmusik*, Vol. 43 of *Forum Musikpädagogik*, Wissner-Verlag, Augsburg, Germany.
- Bushnell, N. & Alcorn, A. (1972), 'Pong', Atari.
- Casella, P. (2002), Music, Agents and Emotions, Licentiate thesis, Engenharia Informática e de Computadores, Instituto Superior Técnico, Universidade Técnica, Lisboa, Portugal.
- Casella, P. & Paiva, A. (2001), MAgentA: An Architecture for Real Time Automatic Composition of Background Music, in A. de Antonio, R. Aylett & D. Ballin, eds, '3rd Int. Workshop on Intelligent Virtual Agents', Vol. 2190/2001 of *Lecture Notes in Computer Science*, Springer, Madrid, Spain, pp. 224–232.
- Chadabe, J. (1985), 'Interactive Music Composition and Performance System', United States Patent Nr. 4,526,078. filed Sept. 1982.

- Chapel, R. H. (2003), *Realtime Algorithmic Music Systems From Fractals and Chaotic Functions: Towards an Active Musical Instrument*, PhD thesis, University Pompeu Fabra, Department of Technology, Barcelona, Spain.
- Chew, E. (2000), *Towards a Mathematical Model of Tonality*, PhD thesis, MIT, Stanford University, Cambridge, USA.
- Chion, M. (1994), *Audio-Vision: Sound on Screen*, Columbia University Press, New York, USA.
- Chiricota, Y. & Gilbert, J.-M. (2007), IMTool: An Open Framework for Interactive Music Composition, in 'Proc. of the 2007 Conf. on Future Play', ACM, Toronto, Canada, pp. 181–188.
- Ciglar, M. (2010a), An ultrasound based instruments generating audible and tactile sound, in 'Proc. of New Interfaces for Musical Expression (NIME) 2010', University of Technology Sydney, Sydney, Australia, pp. 19–22.
- Ciglar, M. (2010b), Tactile Feedback based on Acoustic Pressure Waves, in M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 143–146.
- Clarke, E. F. (1999), Rhythm and Timing in Music, in D. Deutsch, ed., 'The Psychology of Music', 2nd edn, Academic Press, Elsevier, San Diego, CA, USA, chapter 13, pp. 473–500.
- Clarke, E. F. & Windsor, W. L. (2000), 'Real and Simulated Expression: A Listening Study', *Music Perception* 17(3), 277–313.
- Cohen, A. J. (1998), The Functions of Music in Multimedia: A Cognitive Approach, in 'Proc. of the 5th Int. Conf. on Music Perception and Cognition', Seoul, Korea.
- Collins, K. (2006a), 'Flat Twos and the Musical Aesthetic at the Atari VCS', *Popular Musicology Online* **issue 1 musicological critique**.
- Collins, K. (2006b), 'Loops and Bloops: Music On the Commodore 64', *Media Culture* 8.
- Collins, K. (2007), 'In the Loop: Creativity and Constraint in 8-bit Video Game Audio', *twentieth-century music* 4(2), 209–227.
- Collins, K. (2008), *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*, MIT Press, Cambridge, Massachusetts, USA, London, UK.

- Collins, K. (2009), 'An Introduction to Procedural Music in Video Games', *Contemporary Music Review* 28(1), 5–15. issue on Generative Music.
- Collins, K., Tessler, H., Harrigan, K., Dixon, M. J. & Fugelsang, J. (2010), Sound in Electronic Gambling Machines: A Review of the Literature and its Relevance to Game Audio, in M. Grimshaw, ed., 'Game Sound Technology and Player Interaction: Concepts and Developments', IGI Global, Bolton, UK.
- Dannenberg, R. B. (1989), Real-Time Scheduling and Computer Accompaniment, in M. V. Mathews & J. R. Pierce, eds, 'Current Directions in Computer Music Research', number 2 in 'System Development Foundation Benchmark Series', MIT Press, Cambridge, Massachusetts, chapter 18, pp. 225–261.
- Dannenberg, R. B. (2006), The Interpretation of MIDI Velocity, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Tulane University, New Orleans, USA, pp. 193–196.
- Dannenberg, R. B., Pellerin, H. & Derenyi, I. (1998), A Study of Trumpet Envelopes, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, San Francisco, USA, pp. 57–61.
- de Haas, W. B., Rohrmeier, M., Veltkamp, R. C. & Wiering, F. (2009), Modeling Harmonic Similarity Using a Generative Grammar of Tonal Harmony, in '10th Int. Society for Music Information Retrieval Conf. (ISMIR)', Kobe, Japan, pp. 549–554.
- de la Motte, D. (1993), *Melodie*, Deutscher Taschenbuch Verlag.
- de Poli, G. (2004), 'Methodologies for Expressiveness Modelling of and for Music Performance', *Journal of New Music Research* 33(3), 189–202.
- de Quincey, A. (1998), Herman, Master's thesis, Division of Informatics, University of Edinburgh, Edinburgh, UK.
- Demoucron, M. (2008), On the control of virtual violins—Physical modelling and control of bowed string instruments, PhD thesis, Universite Pierre et Marie Curie (Paris 6) and the Royal Institute of Technology (KTH), Stockholm, Sweden.
- Denhoff, M. (1988), 'Stille und Umkehr—Betrachtungen zum Phänomen Zeit, ausgehend vom letzten Orchesterwerk von Bernd Alois Zimmermann', *MusikTexte* 24.
- Dickinson, P. (1989), 'Style-modulation: an approach to stylistic pluralism', *The Musical Times* 130(1754), 208–211.

- Dixon, S. (2000), Extraction of Musical Performance Parameters from Audio Data, in 'Proc. IEEE Pacific-Rim Conf. on Multimedia', IEEE Circuits and Systems Society, IEEE Signal Processing Society, and The Signal & Multimedia Processing Lab, Sydney, Australia, pp. 42–45.
- Dixon, S. (2006), Onset Detection Revisited, in 'Proc. of the 9th Conf. on Digital Audio Effects (DAFx'06)', CIRMMT, Schulich School of Music/McGill University, Montreal, Canada, pp. 133–137.
- Duxbury, C., Bello, J., Davies, M. & Sandler, M. (2003), Complex Domain Onset Detection for Musical Signals, in 'Proc. of the 6th Conf. on Digital Audio Effects (DAFx-03)', London, UK, pp. 1–4.
- Ebcioğlu, K. (1992), An Expert System for Harmonizing Chorales in the Style of J. S. Bach, in M. Balaban, K. Ebcioğlu & O. Laske, eds, 'Understanding Music with AI: Perspectives on Music Cognition', AAAI / MIT Press, Cambridge, UK, chapter 12, pp. 294–334.
- Eisenstein, S. M. (1940), Wertikalny Montasch, in 'Iskusstwo Kino', Moskau, Russia, chapter 9, 12.
- Eisenstein, S. M., Pudowkin, W. I. & Alexandrow, G. W. (1928), Manifest zum Tonfilm, in F.-Z. Albersmeier, ed., 'Texte zur Theorie des Films', 3rd edn, Reclam (1998).
- Ekman, I. (2009), Modelling the Emotional Listener: Making Psychological Processes Audible, in 'Audio Mostly 2009: 4th Conf. on Interaction with Sound', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 33–40.
- Eldridge, A. C. (2002), Adaptive Systems Music: Musical Structures from Algorithmic Process, in C. Soddu, ed., 'Proc. of the 6th Generative Art Conf.', Milan, Italy.
- Fabiani, M. (2009), A Method for the Modification of Acoustic Instrument Tone Dynamics, in A. Sarti, ed., 'Proc. of the 12th Int. Conf. on Digital Audio Effects (DAFx-09)', Politecnico di Milano, Como, Italy, pp. 359–364.
- Fabiani, M. (2010), Rule-Based Expressive Analysis and Interactive Resynthesis of Music Audio Recordings, in 'Interpretationsforschung 2010. Methoden computergestützter Analyse im Kreuzverhör der Disziplinen', Staatliches Institut für Musikforschung Preußischer Kulturbesitz (SIM PK), TU Berlin, Berlin, Germany.
- Farin, G. (2002), *Curves and Surfaces for CAGD: A Practical Guide*, 5th edn, Academic Press, San Diego, CA.

- Farnell, A. (2007), Synthetic game audio with Puredata, in 'Audio Mostly 2007: 2nd Conf. on Interaction with Sound', Fraunhofer IDMT, Interactive Institute/Sonic Studio Piteå, Ilmenau, Germany, pp. 76–83.
- Fastl, H. & Zwicker, E. (2007), *Psychoacoustics: Facts and Models*, Vol. 22 of *Information Sciences*, 3rd edn, Springer, Berlin, Heidelberg, Germany.
- Firelight Technology (2009), 'FMOD'. version 4.28, fmod.org.
- Fitterer, D. (2008), 'Audiosurf: Ride Your Music', Valve.
- Fober, D., Letz, S. & Orlarey, Y. (1996), Recent Developments of MidiShare, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Hong Kong University of Science and Technology, Hong Kong, China, pp. 40–42.
- Fober, D., Letz, S. & Orlarey, Y. (1999), MidiShare joins Open Source Softwares, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Beijing, China, pp. 311–313.
- Foley, J. D., van Dam, A., Feiner, S. K. & Hughes, J. F. (1997), *Computer Graphics: Principles and Practice*, The Systems Programming Series, revised 2nd edn, Addison-Wesley.
- FreeStyleGames (2009), 'DJ Hero', Activision.
- Friberg, A. (1995), A Quantitative Rule System for Musical Expression, PhD thesis, Royal Institute of Technology, Sweden.
- Friberg, A. (2008), Digital Audio Emotions—An Overview of Computer Analysis and Synthesis of Emotional Expression in Music, in 'Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)', Espoo, Finland, pp. 1–6.
- Friberg, A. & Bresin, R. (2000), 'Emotional Coloring of Computer Controlled Music Performances', *Computer Music Journal* 24(4), 44–63.
- Friberg, A., Bresin, R. & Sundberg, J. (2006), 'Overview of the KTH Rule System for Musical Performance', *Advances in Cognitive Psychology, Special Issue on Music Performance* 2(2–3), 145–161.
- Friberg, A. & Sundberg, J. (1999), 'Does Music Performance Allude to Locomotion? A Model of Final Ritardandi Derived from Measurements of Stopping Runners', *The Journal of the Acoustical Society of America* 105(3), 1469–1484.
- Friberg, J. & Gärdenfors, D. (2004), Audio Games: New Perspectives on Game Audio, in 'Proceedings from ACE conference'.

- Fricke, J. P. & Louven, C. (2008), Psychoakustische Grundlagen des Musikhörens, in H. Bruhn, R. Kopiez & A. C. Lehmann, eds, 'Musikpsychologie. Das neue Handbuch', Rowohl Taschenbuch Verlag, Reinbek bei Hamburg, Germany, chapter 5.2, pp. 413–436.
- Fyk, J. (1995), *Melodic Intonation, Psychoacoustics, and the Violin*, Organon, Zielona Góra, Poland.
- Gaar, R. (1997), *Festliche Oberstimmen zum Posaunen-Choralbuch*, Strube Verlag, Munich, Germany.
- Gabrieli, G. (1597), *Sacrae Symphoniae*, Vol. 1, Venice, Italy.
- Gabrielsson, A. (1985), 'Interplay Between Analysis and Synthesis in Studies of Music Performance and Music Experience', *Music Perception* 3(1), 59–86.
- Gabrielsson, A. (1999), Music Performance, in D. Deutsch, ed., 'The Psychology of Music', 2nd edn, Academic Press, Elsevier, San Diego, CA, USA, chapter 14, pp. 501–602.
- Garcia, J. M. (2006), From Heartland Values to Killing Prostitutes: An Overview of Sound in the Video Game Grand Theft Auto Liberty City Stories, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute, Sonic Studio, Piteå, Sweden, pp. 22–25.
- Gartland-Jones, A. (2003), MusicBlox: A Real-Time Algorithm Composition System Incorporating a Distributed Interactive Genetic Algorithm, in G. Raidl, ed., 'Proc. of EvoWorkshops/EuroGP2003, 6th European Conf. in Genetic Programming', Springer, Berlin, Germany, pp. 490–501.
- Gelineck, S. & Serafin, S. (2009), From Idea to Realization—Understanding the Compositional Processes of Electronic Musicians, in 'Audio Mostly 2009: 4th Conf. on Interaction with Sound—Sound and Emotion', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 111–115.
- Goebel, W. (2001), 'Melody lead in piano performance: Expressive device or artifact?', *Journal of the Acoustical Society of America* 110(1), 563–572.
- Goebel, W., Flossmann, S. & Widmer, G. (2010), 'Investigations of Between-Hand Synchronization in Magaloff's Chopin', *Computer Music Journal* 34(3), 35–44.
- Goertz, L. (2004), Wie interaktiv sind Medien?, in C. Bieber & C. Leggewie, eds, 'Interaktivität: Ein transdisziplinärer Schlüsselbegriff', Campus Verlag, Frankfurt/Main, Germany, pp. 97–117.

- Grachten, M., Arcos, J.-L. & de Mántaras, R. (2006), TempoExpress: An Expressivity-Preserving Musical Tempo Transformation System, in 'AAAI'06: Proc. of the 21st National Conf. on Artificial Intelligence', AAAI Press, Boston, Massachusetts, USA, pp. 1609–1612.
- GRAME (2006), *MidiShare Developer Documentation*, Computer Music Research Lab., France. version 1.91.
- Güttler, L., Pommer, M. & Neues Bachisches Collegium Musicum Leipzig (2000), Georg Philipp Telemann: Concerto for Trumpet in D major, in 'Trumpet Concertos Of The Baroque: Telemann, Baldassare, Torelli, Albinoni', Capriccio, Naxos. Originally recorded in 1985.
- Hall, C. V. & O'Donnell, J. T. (2009), Using Regular Expressions to Express Bowing Patterns for String Players, in G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 37–40.
- Hall, C. V., O'Donnell, J. T. & Findlay, B. (2009), Bowing Models for String Players, in G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 323–326.
- Hansen, S. H. & Jensenius, A. R. (2006), The Drum Pants, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute/Sonic Studio, Piteå, Sweden, pp. 60–63.
- Harmonix (2003), 'Amplitude', Sony.
- Harmonix (2006–2010), 'Guitar Hero Series', Harmonix, Neversoft, Vicarious Visions, Budcat Creations, RedOctane, Activision.
- Hartmann, K., Büchner, D., Berndt, A., Nürnberger, A. & Lange, C. (2007), Interactive Data Mining & Machine Learning Techniques for Musicology, in 'Conf. on Interdisciplinary Musicology 2007 (CIM07)', Tallinn, Estonia.
- Hefling, S. E. (1993), *Rhythmic Alteration in Seventeenth- and Eighteenth-Century Music. Notes Inégales and Overdotting*, Schirmer Books, New York, NY.
- Heinicke, S. (2008), Konzeption und Implementierung von Musik-Engine-Funktionalitäten auf der Basis des Midi-Standards, Internship report, Otto-von-Guericke University, Magdeburg, Germany.
- Herber, N. (2006), The Composition-Instrument: musical emergence and interaction, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute/Sonic Studio, Piteå, Sweden, pp. 53–59.

- Hermann, T., Baier, G., Stephani, U. & Ritter, H. (2006), Vocal sonification of pathologic EEG features, in 'Proceedings of the 12th Int. Conf. on Auditory Display', London, UK, pp. 158–163.
- Hähnel, T. (2008), *Auf der Suche nach der inneren Uhr: Musikalische, psychologische und neurologische Konzepte des Timings*, Vol. 14 of *Osnabrücker Beiträge zur systematischen Musikwissenschaft*, B. Enders, Electronic Publishing Osnabrück (epOs-Music), Osnabrück, Germany.
- Hähnel, T. (2010), From Mozart to MIDI: A Rule System for Expressive Articulation, in 'Proc. of New Interfaces for Musical Expression (NIME) 2010', University of Technology Sydney, Sydney, Australia, pp. 72–75.
- Hähnel, T. & Berndt, A. (2010a), Expressive Articulation for Synthetic Music Performances, in 'Proc. of New Interfaces for Musical Expression (NIME) 2010', University of Technology Sydney, Sydney, Australia, pp. 277–282.
- Hähnel, T. & Berndt, A. (2010b), Synthetic and Pseudo-Synthetic Music Performances: An Evaluation, in 'Proc. of the 3rd Int. Conf. of Students of Systematic Musicology (SysMus10)', University of Cambridge, Cambridge, UK.
- Hähnel, T. & Berndt, A. (2011a), Eighth-Notes Performances: Kinds of Inégalité, in 'Audio Mostly 2011: 6th Conf. on Interaction with Sound—Sound and Context', University of Coimbra, Interactive Institute/Sonic Studio Piteã, ACM, Coimbra, Portugal.
- Hähnel, T. & Berndt, A. (2011b), Studying Interdependencies in Music Performance: An Interactive Tool, in 'Proc. of New Interfaces for Musical Expression (NIME) 2011', University of Oslo, Norwegian Academy of Music, Oslo, Norway, pp. 48–51.
- Hiebert, G. (2006), *OpenAL 1.1 Specification and Reference*, Loki Software and Creative Technology.
- Hiller, L. A. & Isaacson, L. M. (1959), *Experimental Music: Composing with an Electronic Computer*, McGraw Hill, New York.
- Hitchcock, A. (1956), 'The Man Who Knew Too Much', Paramount Pictures.
- Hoeberechts, M., Demopoulos, R. J. & Katchabaw, M. (2007), A Flexible Music Composition Engine, in 'Audio Mostly 2007: 2nd Conf. on Interaction with Sound', Fraunhofer IDMT, Interactive Institute/Sonic Studio Piteã, Ilmenau, Germany, pp. 52–57.

- Hoeberechts, M. & Shantz, J. (2009), Realtime Emotional Adaptation in Automated Composition, in 'Audio Mostly 2009: 4th Conf. on Interaction with Sound', Glasgow Caledonian University, Interactive Institute/Sonic Studio Piteå, Glasgow, Scotland, pp. 1–8.
- Holzweissig, E. & Holzweissig, C. (1996), *Klavierschule für den Elementarunterricht*, revised edn, Edition C.F. Peters, Frankfurt/Main, Germany.
- Honing, H. (2003), 'The Final Retard: On Music, Motion, and Kinematic Models', *Computer Music Journal* 27(3), 66–72.
- Hörnel, D. (2000), *Lernen musikalischer Strukturen und Stile mit neuronalen Netzen*, Shaker Verlag, Aachen, Germany.
- Hörnel, D. & Menzel, W. (1999), 'Learning Musical Structure and Style with Neural Networks', *Computer Music Journal* 22(4), 44–62.
- Huckle, T. & Schneider, S. (2002), *Numerik für Informatiker*, Springer, Berlin, Heidelberg, Germany.
- Hudson, R. (2004), *Stolen Time. The History of Tempo Rubato*, reprinted edn, Clarendon Press/Oxford University Press, Oxford, UK.
- Illert, M. (1994), Die Motorik des Musizierens, in H. Bruhn, R. Oerter & H. Rösing, eds, 'Musikpsychologie: Ein Handbuch', Rowohlt, Reinbek, Germany, pp. 639–349.
- Iwai, T. (2005), 'Electroplankton', Indies Zero, Nintendo.
- Jehan, T. (2010), 'Swinger', Music Hack Day, San Francisco, USA.
- Jerkert, J. (2003), Measurements and models of musical articulation, Master's thesis, KTH Royal Institute of Technology, Stockholm, Sweden.
- Jordà, S., Geiger, G., Alonso, M. & Kaltenbrunner, M. (2007), The reacTable: Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces, in 'Proc. of Tangible and Embedded Interaction', Baton Rouge, Louisiana, USA.
- Jordà, S., Kaltenbrunner, M., Geiger, G. & Bencina, R. (2005), The reacTable, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Barcelona, Spain.
- Jørgensen, K. (2006), On the Functional Aspects of Computer Game Audio, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute, Sonic Studio, Piteå, Sweden, pp. 48–52.

- Jørgensen, K. (2008), *Left in the dark: playing computer games with the sound turned off*, in K. Collins, ed., 'From Pac-Man to Pop Music: Interactive Audio in Games and New Media', Ashgate, Hapshire, England, chapter 11, pp. 163–176.
- Jørgensen, K. (2010), *Time for New Terminology? Diegetic and Non-Diegetic Sounds in Computer Games Revisited*, in M. Grimshaw, ed., 'Game Sound Technology and Player Interaction: Concepts and Developments', IGI Global, Bolton, UK.
- Kaltenbrunner, M., Jordà, S., Geiger, G. & Alonso, M. (2006), *The reacTable: A Collaborative Musical Instrument*, in 'Proc. of the Workshop on "Tangible Interaction in Collaborative Environments"(TICE), at the 15th Int. IEEE Workshops on Enabling Technologies', Manchester, U.K.
- Karlin, F. & Wright, R. (2004), *On the Track; A Guide to Contemporary Film Scoring*, revised 2nd edn, Taylor & Francis, New York, London.
- Keller, H. (1973), *Phrasing and Articulation: A Contribution to a Rhetoric of Music*, english edn, W. W. Norton & Company, Inc., New York, USA. L. Gardine (translator).
- Kühhirt, U. & Rittermann, M. (2007), *Interaktive audiovisuelle Medien*, Fachbuchverlag Leipzig/Carl Hanser Verlag, Munich, Germany.
- Kim, S. & André, E. (2004a), *A Generate and Sense Approach to Automated Music Composition*, in 'Proc. of the 9th Int. Conf. on Intelligent User Interfaces', ACM Press, Funchal, Madeira, Portugal, pp. 268–270.
- Kim, S. & André, E. (2004b), *Composing Affective Music with a Generate and Sense Approach*, in 'Proc. of Flairs 2004 - Special Track on AI and Music', AAAI Press.
- Kirnberger, J. P. (1767), 'Der allezeit fertige Polonaisen und Menuetten Komponist', published by G.L. Winter, Berlin, Germany. trans.: The Ever Ready Composer of Polonaises and Minuets.
- Klinger, R. & Rudolph, G. (2006), *Evolutionary Composition of Music with Learned Melody Evaluation*, in N. Mastorakis & A. Cecchi, eds, 'Proc. of the 5th WSEAS Int. Conf. on Computational Intelligence, Man-Machine Systems and Cybernetics', World Scientific and Engineering Academy and Society (WSEAS), Venice, Italy, pp. 234–239.
- Konami (1998), 'Dance Dance Revolution', Disney, Keen, Nintendo.

- Kopiez, R. & Auhagen, W. (1998), *Controlling Creative Processes in Music*, Vol. 12 of *Schriften zur Musikpsychologie und Musikästhetik*, Peter Lang Publishing Group, Germany.
- Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N. & Neuhoff, J. (1997), Status of the field and research agenda, in 'Sonification Report', National Science Foundation.
- Krumhansl, C. & Shepard, R. (1979), 'Quantification of the Hierarchy of Tonal Functions Within a Diatonic Context', *Journal of Experimental Psychology: Human Perception and Performance* 5(4), 579–594.
- Kungel, R. (2004), *Filmmusik für Filmemacher—Die richtige Musik zum besseren Film*, Mediabook-Verlag, Reil, Germany.
- Kurzweil, P., Frenzel, B. & Gebhard, F. (2008), *Physik Formelsammlung für Ingenieure und Naturwissenschaftler*, Studium Technik, Vieweg, Wiesbaden, Germany.
- Kuttruff, H. (2000), *Room Acoustics*, 4 edn, Spon Press, Taylor & Francis, London, UK.
- Land, M. Z. & McConnell, P. N. (1994), 'Method and apparatus for dynamically composing music and sound effects using a computer entertainment system', United States Patent Nr. 5,315,057. filed Nov. 1991.
- Lang, F. (1931), 'M—Eine Stadt sucht einen Mörder', Vereinigte Star-Film, Paramount Pictures. Lang's first sound film.
- Langer, P. (2007), 'Planung von Spiele-Audio', */GameStar/dev* 1, 45–47.
- Leech-Wilkinson, D. (2010), Making Music with Alfred Cortot: Ontology, Data, Analysis, in 'Interpretationsforschung 2010. Methoden computergestützter Analyse im Kreuzverhör der Disziplinen', Staatliches Institut für Musikforschung Preußischer Kulturbesitz (SIM PK), TU Berlin, Berlin, Germany.
- Legend Entertainment (2003), 'Unreal II: The Awakening', Atari. music by Jeremy Soule.
- Lehrdahl, F. & Jackendorf, R. (1983), *A Generative Theory of Music*, MIT Press, Cambridge.
- Lenaerts, D. (2006), *Interactive Music—Or: When the Conductor doesn't know when Psycho Killers will drive the Knife into the young Damsel's Back*, Master's thesis, Katholieke Hogeschool, Communication and Multimedia Design, Limburg, Belgium.

- Lerdahl, F. (1988), 'Tonal Pitch Space', *Music Perception* 5(3), 315–349.
- Lerdahl, F. (1996), 'Calculating Tonal Tension', *Music Perception* 13(3), 319–363.
- Lionhead Studios (2001), 'Black & White', Electronic Arts.
- Lionhead Studios (2005), 'Fable', Microsoft.
- Lissa, Z. (1965), *Ästhetik der Filmmusik*, Henschel, Leipzig, Germany.
- Livingstone, S. R. (2008), Changing Musical Emotion through Score and Performance with a Compositional Rule System, PhD thesis, The University of Queensland, Brisbane, Australia.
- Loki Software & Creative Technology (2009), 'OpenAL', openal.org. version 1.1.
- Löthe, M. (2003), Ein wissensbasiertes Verfahren zur Komposition von frühklassischen Menuetten, PhD thesis, University of Stuttgart, Intelligent System Group, Stuttgart, Germany.
- LucasArts (1997), 'Monkey Island 3: The Curse of Monkey Island'. Music by M. Land.
- Lucasfilm Games (1990), 'Loom'. music by Tchaikovsky (Swan Lake) adapted by George A. Sanger.
- Maempel, H.-J. (2010), Musikaufnahmen als Datenquelle der Interpretationsanalyse, in 'Interpretationsforschung 2010. Methoden computergestützter Analyse im Kreuzverhör der Disziplinen', Staatliches Institut für Musikforschung Preußischer Kulturbesitz (SIM PK), TU Berlin, Berlin, Germany.
- Maestre, E. (2009), Statistical Modeling of Violin Bowing Parameter Contours, in G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 243–250.
- Maestre, E., Bonada, J. & Mayor, O. (2006), Modeling musical articulation gestures in singing voice performances, in 'Proc. of the 121th Audio Engineering Society Convention', Audio Engineering Society (AES), San Francisco, Ca, USA.
- Maestre, E. & Gómez, E. (2005), Automatic Characterization of Dynamics and Articulation of Expressive Monophonic Recordings, in 'Proc. of the 118th Audio Engineering Society Convention', Audio Engineering Society (AES), Barcelona, Spain.

- Maestre Gómez, E. (2006), Coding Instrumental Gestures. Towards a quantitative description of instrumental gestures in excitation-continuous musical instruments, Diploma thesis, Universitat Pompeu Fabra, Barcelona, Spain.
- Manz, J. & Winter, J., eds (1976), *Baukastensätze zu Weisen des Evangelischen Kirchengesangbuches*, Evangelische Verlagsanstalt, Berlin, Germany.
- Marks, A. (2008), *The Complete Guide to Game Audio: For Composers, Musicians, Sound Designers, Game Developers*, 2nd edn, Focal Press, USA.
- Marshall, S. K. & Cohen, A. J. (1988), Effects of Musical Soundtracks on Attitudes toward Animated Geometric Figures, in 'Music Perception', pp. 95–112. Vol. 6, No. 1.
- Mathews, M. V. & Rosler, L. (1968), 'Graphical Language for the Scores of Computer-Generated Sounds', *Perspectives of New Music* 6(2), 92–118.
- Mazzola, G., Göller, S. & Müller, S. (2002), *The Topos of Music: Geometric Logic of Concepts, Theory, and Performance*, Birkhäuser Verlag, Zurich, Switzerland.
- McLean, A. & Wiggins, G. (2010), Petrol: Reactive Pattern Language For Improvised Music, in M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 331–334.
- Meyer, J. (2009), *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instrument Makers*, Modern Acoustics and Signal Processing, 5th edn, Springer, New York, USA. translator: U. Hansen.
- Michels, U. (2001), *dtv-Atlas Musik*, Vol. 1, 20th edn, Deutscher Taschenbuch Verlag.
- Microsoft Corporation (2009), 'DirectX', microsoft.com. version 11.0.
- MIDI Manufacturers Association (2001), *The Complete MIDI 1.0 Detailed Specification*. version 96.1 2nd edition.
- Miedreich, F. (2000), *Zeitliche Streuung von Handlungen—Empirischer Test des Wing-Kristofferson Modells*, Shaker Verlag, Aachen, Germany.
- Miller, B. O., Scarborough, D. L. & Jones, J. A. (1992), On the Perception of Meter, in M. Balaban, K. Ebcioğlu & O. Laske, eds, 'Understanding Music with AI: Perspectives on Music Cognition', AAAI / MIT Press, Cambridge, UK, chapter 18, pp. 428–447.

- Miranda, E. R. & Biles, J. A., eds (2007), *Evolutionary Computer Music*, 1st edn, Springer, USA.
- Moffat, D. C. & Kiegler, K. (2006), Investigating the effects of music on emotions in games, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute, Sonic Studio, Piteå, Sweden, pp. 37–41.
- Monolith Productions (2000), 'The Operative: No One Lives Forever', Fox Interactive, Electronic Arts.
- Monolith Productions (2002), 'No One Lives Forever 2: A Spy in H.A.R.M.'s Way', Sierra Entertainment.
- Mozart, L. (1789), *Gründliche Violinschule*, R. Jung, ed., Faksimile-reprint (1968), 3rd edn, Bärenreiter, Augsburg, Germany.
- Mozart, W. A. (1787), 'Musikalisches Würfelspiel: Anleitung so viel Walzer oder Schleifer mit zwei Würfeln zu componieren ohne musikalisch zu seyn noch von der Composition etwas zu verstehen', Köchel Catalog of Mozart's Work KV1 Appendix 294d or KV6 516f.
- Namco (2003), 'Donkey Konga', Nintendo.
- NanaOn-Sha (1996), 'PaRappa the Rapper', Sony.
- NanaOn-Sha (1999), 'Vib Ribbon', Sony.
- Nintendo EAD (1985), 'Suber Mario Bros.', Nintendo. music by Koji Kondo.
- Nintendo EAD (1998), 'The Legend of Zelda—Ocarina of Time', Nintendo. music by Koji Kondo.
- Nishikado, T. (1978), 'Space Invaders', Taito, Midway Games.
- Origin Systems (1990), 'Wing Commander'.
- Origin Systems (1991), 'Wing Commander II—Vengeance of the Kilrathi'. music by Dana K. Glover.
- Orlarey, Y. (1990), An Efficient Scheduling Algorithm for Real-Time Musical Systems, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, University of Glasgow, Glasgow, Scotland, pp. 194–198.
- Pachet, F. & Roy, P. (2001), 'Musical Harmonization with Constraints: A Survey', *Constraints Journal* 6(1), 7–19.

- Page, J. & Kelly, M. (2007), *Next Gen: More Than Extra Channels?*, in 'Game Developers Conference 2007', CMP Game Group, San Francisco, CA.
- Papadopoulos, G. & Wiggins, G. (1999), *AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects*, in 'AISB Symposium on Musical Creativity', Edingurgh, Scotland.
- Parncutt, R. (1998), *Recording Piano Fingering in Live Performance*, in B. Enders & N. Knolle, eds, 'KlangArt-Kongress 1995', Universitätsverlag Rasch, Osnabrück, Germany, pp. 263–268.
- Parncutt, R. (2003), *Accents and expression in piano performance*, in K. W. Niemöller, ed., 'Perspektiven und Methoden einer Systematischen Musikwissenschaft', Peter Lang, Frankfurt/Main, Germany, pp. 163–185. Festschrift Fricke.
- Parncutt, R., Sloboda, J. A. & Clarke, E. F. (1999), 'Interdependence of Right and Left Hands in Sight-read, Written, and Rehearsed Fingerings of Parallel Melodic Piano Music', *Australian Journal of Psychology* 51(3), 204–210.
- Pauli, H. (1976), *Filmmusik: Ein historisch-kritischer Abriß*, in H. C. Schmidt, ed., 'Musik in Massenmedien', Schott, Mainz, Germany. revised in 1977 and 1981.
- Perl, H. (1998), *Rhythmische Phrasierung in der Musik des 18. Jahrhunderts. Ein Beitrag zur Aufführungspraxis*, Taschenbücher zur Musikwissenschaft vol. 90, 2nd revised edn, Florian Noetzel, Verlag der Heinrichshofen-Bücher, Wilhelmshaven, Germany.
- Piranha Bytes (2001), 'Gothic', Shoebox (Egmont Interactive, dtp Entertainment). music by Kai Rosenkranz.
- Piranha Bytes (2002), 'Gothic II', JoWooD Entertainment. music by Kai Rosenkranz.
- Piranha Bytes (2006), 'Gothic 3', JoWooD Entertainment. music by Kai Rosenkranz.
- Pozzati, G. (2009), *Infinite Suite: Computers and Musical Form*, in G. Scavone, V. Verfaillie & A. da Silva, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Schulich School of Music/McGill University, Montreal, Canada, pp. 319–322.
- Puckette, M. (2010), *Crafting Synthesis Patches onto Live Musical Instruments*, in M. Schedel & D. Weymouth, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Stony Brook University, New York, USA, pp. 250–253.

- Quantz, J. J. (1752), *Versuch einer Anweisung, die Flöte traversière zu spielen*, Bärenreiter. Reprint (1997), H. Augsburg.
- Quarteroni, A., Sacco, R. & Saleri, F. (2007), *Numerical Mathematics*, Vol. 37 of *Texts in Applied Mathematics*, 2nd edn, Springer, Berlin, Heidelberg, Germany.
- Rains, L. & Logg, E. (1979), 'Asteroids', Atari.
- Ramirez, R., Hazan, A. & Maestre, E. (2005), Intra-Note Features Prediction Model for Jazz Saxophone Performance, in 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Universitat Pompeu Fabra, Barcelona, Spain.
- Rasch, R. A. (1979), 'Synchronisation in Performed Ensemble Music', *Acustica* 43, 121–131.
- Ratai, J. (2005), 'Der Jazzwürfel—Ein harmonisches Würfelspiel', Fraunhofer Institut Medienkommunikation. <http://netzspannung.org/learning/meimus/jazz/>.
- Rathgeber, V. (1733), *Missa Civilis, Opus 12, Nr. 8*, Johann Jakob Lotter Verlag, Augsburg, Germany.
- Röber, N. (2008), Interaction with Sound: Explorations beyond the Frontiers of 3D virtual auditory Environments, PhD thesis, Otto-von-Guericke University, Magdeburg, Germany.
- Röber, N., Deutschmann, E. & Masuch, M. (2006), Authoring of 3D Virtual Auditory Environments, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute/Sonic Studio, Piteå, Sweden, pp. 15–21.
- Röber, N., Kaminski, U. & Masuch, M. (2007), Ray Acoustics using Computer Graphics Technology, in 'Proc. of the 10th Int. Conf. on Digital Audio Effects (DAFx-07)', LaBRI University Bordeaux, Bordeaux, France, pp. 117–124.
- Röber, N. & Masuch, M. (2005a), Leaving the Screen: New Perspectives in Audio-Only Gaming, in '11th Int. Conf. on Auditory Display (ICAD)'
- Röber, N. & Masuch, M. (2005b), Playing Audio-Only Games: A Compendium of Interacting with Virtual, Auditory Worlds, in '2nd Int. Digital Games Research Association Conf. (DIGRA)'
- Repp, B. H. (1996), 'Patterns of note onset asynchronies in expressive piano performance', *Journal of the Acoustical Society of America* 100(6), 3917–3932.

- Riemann, H. (1884), *Musikalische Dynamik und Agogik: Lehrbuch der musikalischen Phrasierung auf Grund einer Revision der Lehre von der musikalischen Metrik und Rhythmik*, D. Rahter, Hamburg, Germany.
- Roads, C. (1996), *The Computer Music Tutorial*, MIT Press.
- Robertson, J., de Quincey, A., Stapleford, T. & Wiggins, G. (1998), Real-time music generation for a virtual environment, in 'Proc. of the ECAI 98 Workshop on AI/Alife and Entertainment', Brighton, UK.
- Rohrmeier, M. (2007), A generative grammar approach to diatonic harmonic structure, in '4th Sound and Music Computing Conf. (SMC'07)', University of Athens, Ionian University, Lefkada, Greece, pp. 97–100.
- Rubisch, J., Husinsky, M., Doppler, J., Raffaseder, H., Horsak, B., Ambichl, B. & Figl, A. (2010), A mobile music concept as support for achieving target heart rate in preventive and recreational endurance training, in 'Audio Mostly 2010: 5th Conf. on Interaction with Sound—Sound and Design', Interactive Institute/Sonic Studio Piteå, ACM, Piteå, Sweden, pp. 142–145.
- Russell, J. (1980), 'A Circumplex Model of Affect', *Journal of Personality and Social Psychology* 39(6), 1161–1178.
- Sanderson, P., Wee, A., Seah, E. & Lacherez, P. (2006), Auditory alarms, medical standards, and urgency, in 'Proceedings of the 12th Int. Conf. on Auditory Display', London, UK, pp. 24–27.
- Sanger, G. A. (2003), *The Fat Man on Game Audio: Tasty Morsels of Sonic Goodness*, New Riders.
- Saunders, C., Hardoon, D., Shawe-Taylor, J. & Widmer, G. (2008), 'Using String Kernels to Identify Famous Performers from their Playing Style', *Intelligent Data Analysis* 12(4).
- Schneider, N. J. (1989), *Handbuch Filmmusik II: Musik im dokumentarischen Film*, 2nd edn, Uvk Verlag, Munich, Germany.
- Schneider, N. J. (1990), *Handbuch Filmmusik I: Musikdramaturgie im neuen Deutschen Film*, 2nd edn, Verlag Ölschläger, Munich, Germany.
- Schottstaedt, W. (1989), Automatic Counterpoint, in M. Mathews & J. Pierce, eds, 'Current Directions in Computer Music Research', number 2 in 'System Development Foundation Benchmark Series', MIT Press, Cambridge, Massachusetts, chapter 16, pp. 199–214.

- Schütz, H. (1664), *Historia, der Freuden- und Gnadenreichen Geburth Gottes und Marien Sohn, Jesu Christi, Unsers Einigen Mittlers, Erlösers und Seeligmachers*, Bärenreiter (1955), Kassel, Germany. SWV 435.
- Schütz, M. (2008), *Handbuch Populärmusik*, Strube Verlag, Munich, Germany.
- Sega (2001), 'Rez', Sega.
- Serra, X. (2007), State of the Art and Future Directions in Musical Sound Synthesis, in 'IEEE 9th Workshop on Multimedia Signal Processing', Chania, Crete, pp. 9–12.
- Sevsay, E. (2005), *Handbuch der Instrumentationspraxis*, 1st edn, Bärenreiter, Kassel, Germany.
- Shaffer, L. H. (1981), 'Performances of Chopin, Bach and Bartok: Studies in Motor Programming', *Cognitive Psychology* **13**, 326–376.
- Shore, H. (2001), 'The Lord of the Rings: The Fellowship of the Ring', film score.
- Shore, H. (2002), 'The Lord of the Rings: The Two Towers', film score.
- Shore, H. (2003), 'The Lord of the Rings: The Return of the King', film score.
- Shultz, P. (2008), Music Theory in Music Games, in K. Collins, ed., 'From Pac-Man to Pop Music: Interactive Audio in Games and New Media', Popular and Folk Music Series, Ashgate, Hampshire, UK, chapter 12, pp. 177–188.
- Sloboda, J. A. & Juslin, P. N. (2001), *Music and Emotion: Theory and Research*, Oxford University Press, New York, NY.
- Speedsoft (2006), 'VSampler 3.5.2 Professional', <http://www.vsampl3.com> [last visited: Jan. 2010]. software sampler including a standard sample library.
- Stapleford, T. (1998), The Harmony, Melody, and Form of HERMAN, a Real-Time Music Generation System, Master's thesis, Division of Informatics, University of Edinburgh, Edinburgh, UK.
- Stenzel, M. (2005), Automatische Arrangieretechniken für affektive Sound-Engines von Computerspielen, Diploma thesis, Otto-von-Guericke University, Department of Simulation and Graphics, Magdeburg, Germany.
- Stockhausen, K. (1956), *Klavierstück XI*, Universal Edition.
- Stockmann, L. (2007), Designing an Audio API for Mobile Platforms, Internship report, Otto-von-Guericke University, Magdeburg, Germany.

- Stockmann, L. (2008), *Interaktive Sonifikation von 2D/3D-Daten*, Diploma thesis, Otto-von-Guericke University, Faculty of Computer Science, Department of Simulation and Graphics, Magdeburg, Germany.
- Stockmann, L., Berndt, A. & Röber, N. (2008), A Musical Instrument based on Interactive Sonification Techniques, in 'Audio Mostly 2008: 3rd Conf. on Interaction with Sound', Interactive Institute/Sonic Studio Piteå, Piteå, Sweden, pp. 72–79.
- Sundberg, J., Friberg, A. & Bresin, R. (2003), 'Attempts to Reproduce a Pianist's Expressive timing with Director Musices Performance Rules', *Journal of New Music Research* 32(3), 317–325.
- Sundberg, J. & Verrillo, V. (1980), 'On the Anatomy of the Retard: A Study of Timing in Music', *The Journal of the Acoustical Society of America* 68(3), 772–779.
- Taube, H. K. (2004), *Notes from the Metalevel: Introduction to Algorithmic Music Composition*, Taylor & Francis, London, UK. Studies on New Music Research.
- Theremin, L. S. (1924), 'Method of and Apparatus for the Generation of Sounds', United States Patent Nr. 73,529.
- Thiel, W. (1981), *Filmmusik in Geschichte und Gegenwart*, Henschel Verlag, Berlin, Germany.
- Thomas, P. (1966), 'Raumpatrouille—Die phantastischen Abenteuer des Raumschiffs Orion', Westdeutscher Rundfunk (WDR), Office de Radiodiffusion Télévision Française (ORTF), Bavaria Film. film score.
- Tobler, H. (2004), CRML—Implementierung eines adaptiven Audiosystems, Master's thesis, Fachhochschule Hagenberg, Medientechnik und -design, Hagenberg, Austria.
- Todd, N. P. (1989), 'A computational model of rubato', *Contemporary Music Review* 3, 69–88.
- Todd, N. P. M. (1992), 'The dynamics of dynamics: A model of musical expression', *The Journal of the Acoustical Society of America* 91(6), 3540–3550.
- Todd, N. P. M. (1995), 'The Kinematics of Musical Expression', *The Journal of the Acoustical Society of America* 97(3), 1940–1949.
- Tolkien, J. R. R. (1954, 1955), *The Lord of the Rings Trilogy*, 1st edn, George Allen & Unwin, London, UK.
- Toru, I. (1980), 'Pac-Man', Namco, Midway.

- Tuohy, D. R. & Potter, W. D. (2006), An Evolved Neural Network/HC Hybrid for Tablature Creation in GA-based Guitar Arranging, in T. H. Park, G. Essl & I. Fujinaga, eds, 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, Tulane University, New Orleans, Louisiana, USA, pp. 576–579.
- Verbiest, N., Cornelis, C. & Saeys, Y. (2009), Valued Constraint Satisfaction Problems Applied to Functional Harmony, in 'IFSA World Congress EUSFLAT Conference', International Fuzzy Systems Association, European Society for Fuzzy Logic and Technology, Lisbon, Portugal, pp. 925–930.
- Vienna Symphonic Library GmbH (2010a), 'Vienna Instruments', <http://vsl.co.at/> [last visited: March 2010]. Software sampler.
- Vienna Symphonic Library GmbH (2010b), 'Vienna Symphonic Library', <http://vsl.co.at/> [last visited: March 2010]. Sample library of orchestral instruments.
- Wallis, I., Ingalls, T. & Campana, E. (2008), Computer-Generating Emotional Music: The Design of an Affective Music Algorithm, in 'Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)', Espoo, Finland, pp. 7–12.
- Walter, K. (2002), *Grenzen spielerischen Erzählens. Spiel- und Erzählstrukturen in graphischen Adventure Games*, Vol. 2 of *Medienwissenschaften*, universi Verlag, Siegen, Germany.
- Widmer, G. & Goebel, W. (2004), 'Computational Models of Expressive Music Performance: The State of the Art', *Journal of New Music Research* 33(3), 203–216.
- Williams, J. (1977), 'Star Wars Episode IV', Lucasfilm, 20th Century Fox. film score.
- Williams, L. (2006), Music videogames: the inception, progression and future of the music videogame, in 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute, Sonic Studio, Piteå, Sweden, pp. 5–8.
- Windsor, W. L. & Clarke, E. F. (1997), 'Expressive Timing and Dynamics in Real and Artificial Musical Performance: Using an Algorithm as an Analytical Tool', *Music Perception* 15(2), 127–152.
- Wing, A. M. & Kristofferson, A. B. (1973), 'Response delays and the timing of discrete motor responses', *Perception & Psychophysics* 14(1), 5–12.
- Wingstedt, J. (2004), Narrative functions of film music in a relational perspective, in 'ISME Proc.—Sound Worlds to Discover', Int. Society for Music Education, Spain.

- Wingstedt, J. (2006), REMUPP—a tool for investigating musical narrative functions, *in* 'Audio Mostly 2006: A Conf. on Sound in Games', Interactive Institute, Sonic Studio, Piteå, Sweden, pp. 42–47.
- Wingstedt, J. (2008), Making Music Mean: On Functions of, and Knowledge about, Narrative Music in Multimedia, PhD thesis, Luleå University of Technology, Department of Music and Media, Luleå, Sweden.
- Wingstedt, J., Liliedahl, M., Lindberg, S. & Berg, J. (2005), REMUPP: An Interactive Tool for Investigating Musical Properties and Relations, *in* 'Proc. of New Interfaces for Musical Expression (NIME) 2005', Vancouver, Canada, pp. 232–235.
- Wooller, R. W. & Brown, A. R. (2005), Investigating morphing algorithms for generative music, *in* 'Third Iteration: Third International Conference on Generative Systems in the Electronic Arts', Melbourne, Australia.
- Yoon, M.-J. & Lee, I.-K. (2006), Musical Tension Curves and its Application, *in* 'Proc. of the Int. Computer Music Conf. (ICMC)', International Computer Music Association, New Orleans, USA, pp. 482–486.
- Young, D. (2002), The Hyperbow Controller: Real-Time Dynamics Measurement of Violin Performance, *in* E. Brazil, ed., 'Proc. of the 2002 Conf. on New Instruments for Musical Expression (NIME-02)', Media Lab Europe, Dublin, Ireland.
- Zimmer, H. (1997), 'The Peacemaker', DreamWorks Pictures. film score.