

Sound-Technologien

Tibor Schütz

23.7.2003

Quelle:

<http://www.deinmeister.de/>

Inhaltsverzeichnis

1	Sound-Technologien	1
1.1	Der menschliche Hörsinn	1
1.1.1	Begriff: Dezibel	1
1.2	Anwendungsgebiete in Spielen	2
1.2.1	Soundtracks	2
1.2.2	Effekte	2
1.3	Softwaretechnische Aspekte	3
2	Notenbasierte Realisierung	4
2.0.1	Begriff: Stimme (Voice)	5
2.1	Der IBM-PC	5
2.2	Der C64: SID	6
2.3	AdLib: OPL2	7
3	Digitalisierte Wiedergabe: Sampling	8
3.1	SoundBlaster und Verwandte	8
3.2	Samplemixing	9
3.2.1	Mixing und Clipping	10
3.2.2	Stimmen- und Samplemanagement	10
3.2.3	Audiokompression	11
3.3	Abstraktion	11
3.4	MIDI und Module	12
4	Hardwarebeschleunigung - Hardwaremixing	13
5	CD-ROM + CD-DA	14
6	Raumklang	15
6.1	Verschiedene Varianten	15
6.1.1	Kunstkopfstereophonie, QSound, Virtual Surround	15
6.1.2	Dolby Surround	15
6.1.3	Quadrophonie, 5.1 / 6.1 / 7.1 etc.	15
6.2	DirectSound3D und Co.	16
6.3	Akustischer Nachbau von Räumen	17

6.3.1 Vereinfachte Raumakustik	17
7 Interaktive Musik	18
7.1 Sequentielle Titelwiedergabe	18
7.2 Parallele Titelwiedergabe	18
7.3 Module	19
7.4 DirectMusic	20
8 Adressen	21

Kapitel 1

Sound-Technologien

Der Hörsinn liefert nach dem Sehsinn den zweitwichtigsten Teil der von Menschen wahrgenommenen Informationen. Dementsprechend ist eine passende akustische Untermalung einer der wichtigsten Immersionsfaktoren in Computerspielen und VirtualReality-Umgebungen

1.1 Der menschliche Hörsinn

Um die Erfordernisse an eine audiotecnische Implementierung abschätzen zu können sind einige Grundkenntnisse bezüglich der menschlichen Schallwahrnehmung bedeutend.

1.1.1 Begriff: Dezibel

Die Amplitude von elektrischen und akustischen Signalen wird üblicherweise in Dezibel(dB) angegeben. Dabei ist zu beachten daß es sich um ein logarithmisches Maß handelt: Eine Erhöhung um 6dB entspricht einer Verdoppelung der Amplitude.

Der Bezugspunkt(0dB) für absolute Angaben ist in der Tontechnik die maximal verzerrungsfrei verarbeitbare Amplitude, für Lautstärkemessungen die Hörschwelle.

Der Dynamikbereich des menschlichen Gehörs umfaßt zwischen der Hörschwelle und dem ohne Hörschäden ertragbaren Lautstärkepegel rund 85-90dB. Dies entspricht einer 15-fachen Verdoppelung des leisesten hörbaren Geräusches.

Der Bereich der wahrnehmbaren Frequenzen beginnt bei 20Hz und endet je nach Alter zwischen unter 10kHz und knapp über 20kHz. Zusammen mit der nur schwach ausgeprägten Richtcharakteristik der Ohrmuscheln ergibt sich daraus eine im Vergleich zu Hunden oder Fledermäusen eine relativ schwach ausgeprägte Ortung der Schallquellen.

Demgegenüber ist das Ohr ziemlich empfindlich gegenüber zeitlichen Schwankungen als auch Tonhöhenänderungen. Dabei wird allerdings nicht der eigentliche Amplitudenverlauf eines Schallsignals erfaßt, sondern der Amplitudenverlauf der verschiedenen hörbaren Frequenzen wird wahrgenommen. Menschen die über das sogenannte absolute Gehör verfügen

können sogar nicht nur die Differenz benachbarter Frequenzen wahrnehmen sondern auch den absoluten Wert einer einzigen Frequenz.

Desweiteren werden große Schwankungen des Schallpegels, wie sie etwa bei lauter Musik oder Explosionen auftreten können werden nicht nur mit dem Ohr, sondern auch vom Tastsinn wahrgenommen.

1.2 Anwendungsgebiete in Spielen

Primär kann man zwei Arten von Sound in Spielen unterscheiden: Soundtracks und Effekte

1.2.1 Soundtracks

Die Hintergrundmusik in Spielen dient denselben Zwecken wie auch in Filmen. Sie verdeutlichen einerseits die einer Umgebung inwohnende Atmosphäre, andererseits lassen sich mit Musik auch Stimmungen ideal ausdrücken. Paradebeispiele für den ersten Fall sind die Karibikmusik aus *Monkey Island* als auch die Musik im Tempel bei *Unreal*, der zweite Fall findet sich am deutlichsten in *BladeRunner* am Ende des ersten Tages.

Auf der technischen Seite findet man eine nur lose Verknüpfung des Soundtracks mit dem momentanen Spielgeschehen, während die Qualität der Hintergrundmusik möglichst gut sein sollte - es kann durchaus vorkommen daß sie an manchen Stellen im Spiel den größten Teil der Aufmerksamkeit des Spielers an sich zieht.

1.2.2 Effekte

Effekte sind kurze, direkt ans Spielgeschehen eingebundene Audiosequenzen. Diese können ebenfalls zur Erzeugung einer gegebenen Atmosphäre verwendet werden, wie es bei Wellenrauschen, summenden Maschinen oder zirpenden Grillen der Fall ist. Viel wichtiger ist jedoch ihre Bedeutung zur Vermittlung konkreter Informationen. Bei der Sprachwiedergabe ist dies offensichtlich, aber auch Status und Positionsinformationen können akustisch übermittelt werden. Die Handhabung von Benutzeroberflächen aller Art im Spiel wie Menüs, die Bewegungssteuerung, Codeschlösser oder Abzugshebel wird zudem durch akustisches Feedback stark verbessert.

Dementsprechend sind beim Einsatz von Soundeffekten eine möglichst präzise Kontrolle in Abhängigkeit vom Spielgeschehen gewünscht, während die klangliche Qualität meist zweitrangig ist.

Die Trennung zwischen Hintergrundmusik und Effekten verwischt in einigen Fällen, etwa wenn lange Monologe ohne direkten Bezug zur Grafik wiedergegeben werden. Auch gibt es Fälle in denen Musik in Abhängigkeit vom Spielgeschehen abgespielt wurde, man denke

nur an die Musikbox in *LeisureSuitLarry1* oder die unzähligen Auftritte von Musikgruppen in verschiedenen Adventures.

1.3 Softwaretechnische Aspekte

Es fällt auf, daß die unterschiedlichsten Spielegenres weitgehend dieselben Anforderungen an die verwendete Musik- und Effektwiedergabe haben, während Grafik und Simulation meist hochspezifische Lösungen benötigen. Desweiteren ist die akustische Wiedergabe weitgehend unabhängig vom Rest des Spielecodes, so daß sich dieser leicht separat entwickeln läßt.

Die führte schon früh zu der Entwicklung standardisierten Bibliotheken zur Audiowiedergabe, die nicht selten von darauf spezialisierten Fremdfirmen übernommen werden konnte. Eine weitere Konsequenz dessen war, daß die Musik und die Effekte selbst meist auch von Dritten erstellt werden, so daß sich die Spieleentwickler ganz auf ihr Kerngebiet konzentrieren können.

Kapitel 2

Notenbasierte Realisierung

#8	Channel 1	Channel 2	Channel 3	Channel 4	Channel 5	Channel 6
0	G#4 11 -- S88	---	S88	G-4 10,45 S88	---	S87
1	---	---	G-4 10,45	---	F-5 01 -- S88	G-6 14 -- S88
2	---	---	---	---	---	---
3	G#4 12,30	G#4 11,32	G-4 10,21	G-4 10,22	---	---
4	G#4 12	---	---	G-4 10,22	---	---
5	---	---	---	---	---	---
6	G#4 13	G#4 12,15	---	G-4 10,10	---	---
7	---	G#4 12,32	---	---	---	---
8	G#4 11	---	---	---	---	D-6 15
9	---	G#4 13,32	---	---	---	---
10	---	---	---	---	---	---
11	G#4 12,30	G#4 11,32	D-6 10,14	---	---	---
12	---	---	D-6 10,21	---	---	---
13	G#3 12,40	---	---	---	---	---
14	G#4 13	G#4 12,15	G-5 10,45	D-6 10,07	---	---
15	G#3 13 -- 003	---	---	D-6 10,10	---	G-6 14,30
16	G#4 11	G#3 12,20	G-4 10,45	---	---	G-6 14
17	---	G#4 13,32	---	G-5 10,22	---	---
18	---	G#3 13,32 003	---	---	---	---
19	G#4 12,30	G#4 11,32	G-4 10,21	G-4 10,22	---	---
20	G#4 12	---	---	---	---	---
21	---	---	---	---	---	---
22	G#4 13	G#4 12,15	---	G-4 10,10	---	G-6 14,30
23	---	G#4 12,32	---	---	---	G-6 14
24	G#4 11	---	---	---	---	D-6 15
25	---	G#4 13,32	---	---	---	---

Ein naheliegender Ansatz für Musik auf dem Computer ist die Speicherung der zugrundeliegenden Noten, die bei der Wiedergabe mittels geeigneter Hardware wieder in Töne umgewandelt werden. Neben dem geringen Platzbedarf (je nach Aufwand einige dutzend bis einige hundert Bytes/Sekunde) besitzt diese Methode darüber hinaus den Vorteil, in

Verbindung mit geeigneter Hardware wenig Rechenzeit auf der CPU zu verbrauchen.

Die aus der herkömmlichen Musik bekannten Noten können auf zwei Arten gespeichert werden: Entweder als sequentielle Folge von Noten und deren Abspielzeit (Timestamp), wie es etwa beim MIDI-Format der Fall ist, oder als ein- bzw. zweidimensionale Tabelle, wie es die verschiedenen Tracker-Formate (Modules) handhaben. In dieser Tabelle ist die Zeit schon in einem festen Vielfachen des Grundtaktes quantisiert, die Tabelleneinträge entsprechen den dazugehörigen Noten. Bei einer zweidimensionalen Darstellung können die verschiedenen Spalten schon im Vorraus fest bestimmten Kanälen oder Instrumenten zugeordnet werden.

2.0.1 Begriff: Stimme (Voice)

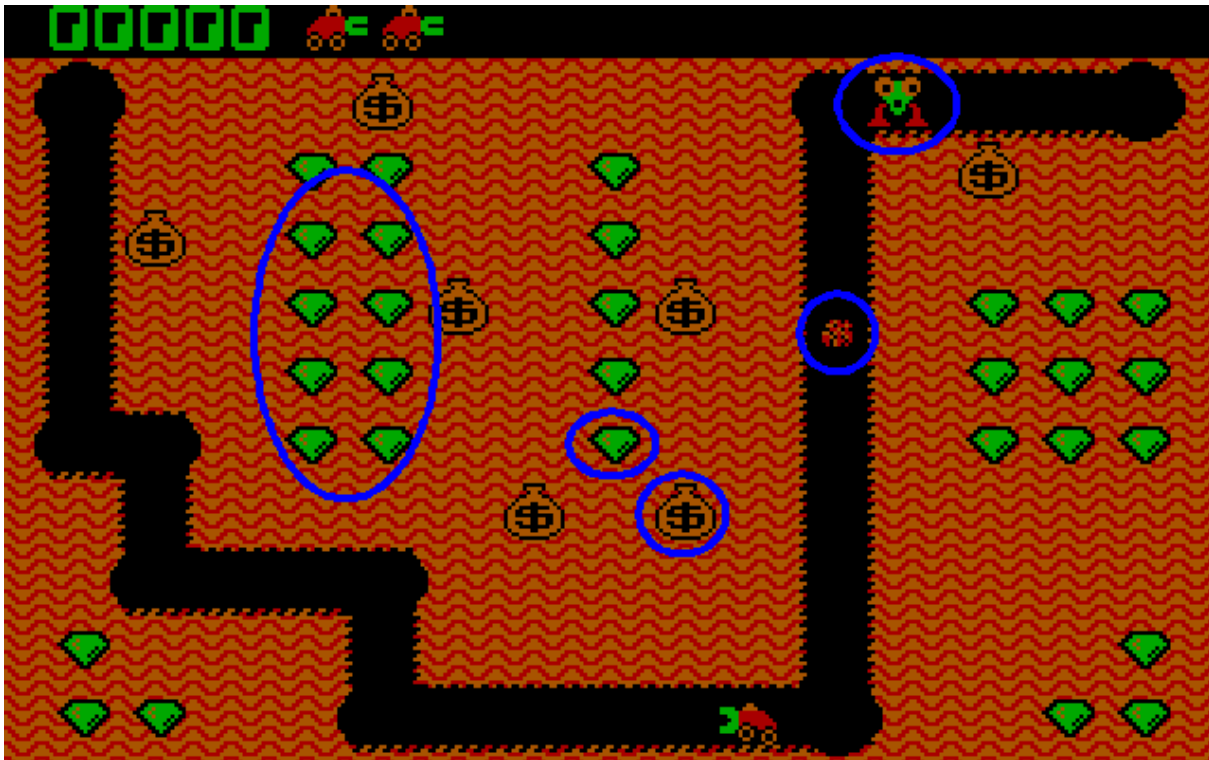
Die kleinste akustische Einheit, die ein Synthesizer oder ein Soundchip getrennt wiedergeben kann heißt Stimme. Eine Stimme entspricht z.B. einer schwingenden Saite eines Klaviers oder einer Geige.

Einstimmige Geräte werden monophon genannt, mehrstimmige polyphon.

2.1 Der IBM-PC



Der Anfang der 80er Jahre konstruierte IBM-PC besaß als Bürocomputer nur äußerst rudimentäre akustische Fähigkeiten. Um mit dem eingebauten Kleinlautsprecher Warntöne ausgeben zu können wurde dieser einfach an den noch freien zweiten Ausgang des Timerchips (Programmable Intervall Timer, PIT) angeschlossen. Als reiner Digitalbaustein konstruiert konnte dieser natürlich nur einfache Rechteckwellen wiedergeben, auch die Lautstärke konnte nicht verändert sondern nur abgeschaltet werden. Immerhin deckte diese Minimalversion eines monophonen Synthesizers den gesamten Frequenzbereich des Ohres ab.



Eines der ersten Spiele auf dem IBM-PC war Digger, das mit den beschränkten Mitteln des Ur-PCs schon alle wesentlichen akustischen Einsatzmöglichkeiten aufwies:

- Drei Soundtracks für normales Spielgeschehen, Bonusmodus und Begräbnismarsch
- Feedback beim abschießen eines Feuerballs und beim Einsammeln eines Edelsteines, Goldklumpen oder Gegners
- Information über den Zustand eines wackelnden Goldsacks, der Flugweite eines Feuerballs als auch der Anzahl mehrerer am Stück eingesammelter Edelsteine

2.2 Der C64: SID

Im Gegensatz zum PC wurde der C64 als Homecomputer schon von Anfang an mit einem richtigen spielfähigen Soundchip, dem SID (Sound Interface Device), ausgestattet. In Anlehnung an richtige Synthesizer konstruiert besaß er 3 Stimmen, die neben einer Rechteckwelle auch eine Pulswelle, einen Sägezahn oder Rauschen mit jeweils beliebiger Lautstärke wiedergeben konnten. Um die klanglichen Fähigkeiten zu erweitern konnte man mittels eines nachgebildeten Ringmodulators zwei Stimmen zu einer einzigen Stimme mit neuem Klang zusammemmultiplizieren oder mittels einer Hüllkurve den Klang eines im Laufe der Zeit ausklingenden Instrumentes nachbilden. Ein Filter zum Nachbearbeiten des Klanges rundeten den Funktionsumfang des SIDs ab.

Zwar konnten mit diesem Chip sowohl Musik als auch Soundeffekte gleichzeitig wiedergegeben

werden, dennoch reichte die geringe Stimmenzahl kaum aus. Um dies zu umgehen, wurden z.B. die Noten eines Akkorden hintereinander abgespielt (Arpeggio) oder unbedeutendere Klänge von wichtigeren verdrängt. Auch die Handhabung der einzelnen Stimmen erwies sich als trickreich, so daß die damaligen Musiker mehr Musik-Programmierer waren.

2.3 AdLib: OPL2

Mit dem in der zweiten Hälfte der 80er vorgestellten AdLib kam zum ersten Mal eine Soundkarte auf den Markt, mit der auch ein normaler Musiker annehmbar arbeiten konnte. Der OPL-Chip auf der AdLib enthielt dieselbe Technologie, die Yamaha in den damaligen Keyboards verwendete: Die FM(FrequentModulation)-Synthese. Bei dieser besitzt jede Stimme mehrere Oszillatoren, von denen einer den eigentlichen Ton wiedergibt während die anderen diesen in der Frequenz und Lautstärke modulieren und somit die Obertöne für den gewünschten Klang erzeugen. Mit insgesamt 9 Stimmen (oder alternativ 6 Stimmen + 5 Schlagzeugstimmen) wurde auch die Wiedergabe mehrere Noten und Effekte erleichtert. Wie schon vom SID bekannt gab es mehrere Wellenformen und Hüllkurven. Zur leichteren Handhabung wurden Presets zur Nachbildung vertrauter Instrumente geliefert.

Da der OPL2 allerdings nur 2 Operatoren anstatt 4, 6 oder 8 wie in professionellen Keyboards besaß blieben die klanglichen Fähigkeiten auf typische Synthesizerklänge beschränkt. Insbesondere die Wiedergabe von Sprache und komplexen Effekten ließen sich auch hiermit nicht realisieren.

Kapitel 3

Digitalisierte Wiedergabe: Sampling

Der andere Ansatz für die Tonwiedergabe auf dem PC entspricht der Aufnahme auf einem Tonband. Um ein Audiosignal zu digitalisieren muß es in zweifacher Hinsicht diskretisiert werden:

- Die zeitliche Abtastrate (Samplingrate) bestimmt die maximale Tonhöhe. Nach dem Nyquist-Theorem muß die Samplingrate mindestens doppelt so hoch sein wie die höchste wiederzugebende Frequenz.
- Die Amplitudenauflösung bestimmt den Dynamikbereich. Ein Bit entspricht einem Umfang von 6dB

Hieraus ergeben sich für eine qualitativ gute Wiedergabe hohe Anforderungen:

- Für für CD-Qualität (ca 20kHz max.) beträgt die Abtastrate 44100Hz, die Amplitudenauflösung für den vollen Dynamikbereich benötigt 16Bit: Der daraus resultierende Speicherbedarf ist enorm und beträgt rund das Tausendfache einer notenbasierten Wiedergabe.
- Zudem müssen die einzelnen abgetasteten Werte zeitlich im exakten Abstand voneinander wiedergegeben werden um Verzerrungen und Störgeräusche zu vermeiden.

Die ersten Versuche, mittels des PC-Lautsprechers (bis zu effektiv 6Bit durch höherfrequente serielle Ausgabe) oder eines an die parallele Schnittstelle anschließbare Digital-Analog-Wandler gesamplete Klänge wiederzugeben waren dementsprechend nur eingeschränkt einsetzbar, da sie einen Großteil der Rechenzeit für sich beanspruchten.

3.1 SoundBlaster und Verwandte

Behoben wurde das Problem durch Soundchips, welche die einzelnen gesampleten Werte selbstständig per DMA (Direct Memory Access) aus dem Speicher holen können. Hierbei kann ein kompletter Block an Audiodaten auf einmal in den Speicher geschrieben werden, so daß anschließend während dessen Wiedergabe der Prozessor frei für restliche Aufgaben

bleibt. Das Ende des abgespielten Blocks kann mittels eines Interrupts signalisiert werden, so daß der nächste Block vorbereitet werden kann.

Das einfachste Verfahren zur kontinuierlichen Wiedergabe ist der Double Buffer: Hierbei werden zwei Blöcke benötigt, von denen einer abgespielt und der andere mit neuen Daten gefüllt wird. Wird das Ende des abgespielten Blockes mittels Interrupt signalisiert vertauschen die Blöcke ihre Rollen und das Verfahren beginnt von vorn.

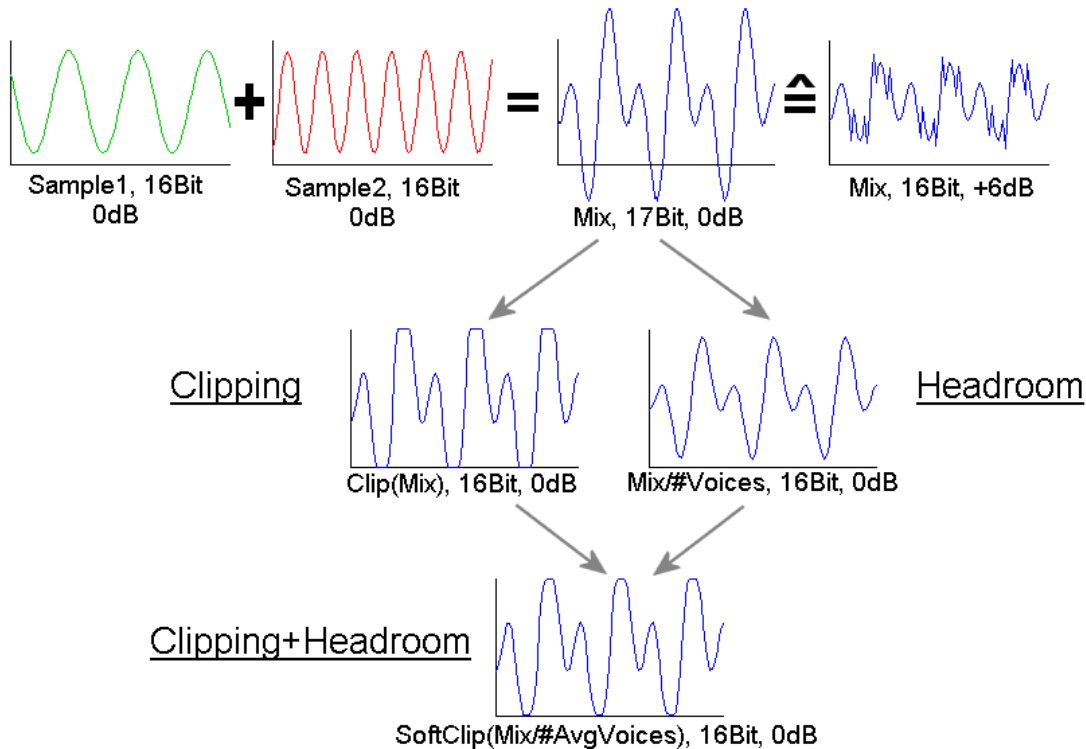
Hierbei läßt sich ein neues Phänomen beobachten: Da die Schreibposition für neue Daten nicht mit der Wiedergabeposition des Audiochips identisch werden Effekte mit einer gewissen Latenzzeit verzögert wiedergegeben. Diese läßt sich verringern, indem statt des DoubleBuffer-Verfahrens nur ein einziger, kontinuierlich wiedergegebener Puffer verwendet wird. Neue Daten werden in regelmäßigen Abständen zwischen der letzten Schreibposition und der aktuellen Wiedergabeposition eingefügt werden, für neue Effekte kann auch der Teil vor der Wiedergabeposition zuzüglich eines gewissen Sicherheitsbestandes verwendet werden. Dieser Abstand kann jedoch nicht beliebig klein gewählt werden, da die Anfälligkeit für Störungen durch zu spät geschriebene Daten dadurch stark zunimmt.

3.2 Samplemixing

Ein Großteil der Soundchips verfügt nur über einen Digital-Analog-Wandler pro Kanal, so daß mehrere Effekte vor der Wiedergabe per Software zusammengemischt werden müssen. Neben der eigentlichen Addition sollten dabei auch weitere Möglichkeiten zur Effektmanipulation bereitstehen (nach steigendem Rechenaufwand sortiert):

- Variable Lautstärke pro wiedergegebenem Effekt
- Stereopositionierung (Panning), im Idealfall unter Berücksichtigung des quadratischen Zusammenhangs von Amplitude und wahrgenommener Lautstärke zur Vermeidung von Lautstärkeschwankungen
- Änderung der Wiedergabefrequenz mit möglichst interpolierten Samples zur Verringerung von Verzerrungen
- Modulation der Tonhöhe(Vibrato) und der Lautstärke(Tremolo)
- Aufwändige Effekte wie Hall, Filter,Chorus, etc.

3.2.1 Mixing und Clipping



Bei der Addition mehrerer Stimmen kann es passieren, daß der Dynamikbereich des Audiochips überschritten wird. Da nur die unteren Bits verarbeitet werden resultiert dies in starken Verzerrungen. Diese lassen sich verringern, indem das resultierende Signal auf die Grenzen des Dynamikbereichs beschnitten wird. Allerdings entstehen dadurch neue störende Obertöne. Alternativ kann durch Division mit der Anzahl der Stimmen eine Überschreitung vermieden werden. Hierbei geht allerdings pro Verdoppelung der Stimmenzahl ein Bit an Dynamikumfang verloren.

Da in der Praxis nur äußerst selten sich alle Stimmen zum maximal möglichen Pegel addieren wird meist ein Mischverfahren verwendet, bei dem einige Bits per Division (üblicherweise ein einfaches Shift) als sog. Headroom reserviert werden, so daß nur selten auf Clipping zurückgegriffen werden muß. Dabei lassen sich durch eine abgerundete Clippingfunktion störende Oberwellen weiter reduzieren.

3.2.2 Stimmen- und Samplemanagement

Wie man sieht wird die Anzahl aktiver Stimmen sowohl durch die verfügbare Rechenzeit als auch die Clippingproblematik begrenzt. Um dennoch beliebig viele virtuell aktive Stimmen zu ermöglichen besitzen die meisten Audiobibliotheken eingebaute Verwaltungsfunktionen für die Zuordnung virtueller Stimmen zu tatsächlich verfügbaren Stimmen.

Die einfachste Methode ist die feste Zuordnung von Effekten zu den vorhandenen Stimmen oder das Ersetzen eines alten Effektes durch einen Neuen. Verfeinern läßt sich diese Verdrängung, indem verschiedenen Effekten Prioritäten zugeordnet werden, so daß nur bei genügend hoher Prioritätsstufe ein Ersetzen stattfindet. Statt einer Priorität kann auch die Lautstärke oder die Entfernung eines Effektes verwendet werden.

Die Vorteile der verdrängungsbasierten Methode zeigen sich insbesondere bei einer Menge aus mehreren verfügbaren Stimmen, da hier stets die aktuell unwichtigsten Effekte verdrängt werden. Durch Verwendung mehrerer Verfahren läßt sich deren Einsatz gezielt an die Anforderungen eines Spieles anpassen.

3.2.3 Audiokompression

Um den enormen Speicherbedarf zu verringern gehören auch Kompressionsverfahren zum Leistungsumfang jeder besseren Audiobibliothek. Im Gegensatz zu Verfahren wie ZIP oder RAR liegt der Schwerpunkt hierbei jedoch bei Verfahren, die sich mit wenig Speicher schnell zur Laufzeit an beliebigen Punkten im Datenstrom anwenden lassen, nicht auf exakter Reproduktion der Quelldaten.

Leicht in Hardware implementierbare Formate sind vor allem ADPCM (Adaptive Differentielle PulsCodeModulation), das anstelle absoluter Werte nur die Differenz benachbarter Samplewerte betrachtet. Zur weiteren Datenverringern wird dabei noch über die vorherigen Werte extrapoliert.

uLaw und aLaw setzen dagegen statt einer linearen eine logarithmische Diskretisierung ein, so daß bei großen Amplituden keine übermäßig feine Abtastung anfällt. Beide Formate erreichen je nach Qualität Kompressionsraten von 50-75% und eignen sich speziell für Effekte.

Aufwendiger sind Formate wie mp3 (Mpeg1 Audio Layer3) und OggVorbis, die zuerst das Audiosignal auf kurzen Intervallen in eine Frequenz-Amplituden-Darstellung umwandeln und anhand eines psychoakustischen Modells sämtliche Töne entfernen, die das durchschnittliche menschliche Ohr ohnehin nicht wahrnehmen würde. Aufgrund der erzielbaren Kompressionsrate von über 90% eignet es sich vor allem bei Sprache und Musik. Für allgemeine Effekte eignet es sich weniger, da durch die Nachbearbeitung beim Mischen das angewendete psychoakustische Modell nicht mehr gültig ist.

3.3 Abstraktion

Ein weiterer wichtiger Grund für den Einsatz von fertigen Audiobibliotheken ist die Abstraktion der darunterliegenden Hardware und Betriebssystems. Im Idealfall unterstützt eine Bibliothek sämtliche Variationen einer Plattform transparent, bei der Portierung auf andere Plattformen sollte von Seiten der Bibliothek eine einfache Neuübersetzung des Programmes genügen.

3.4 MIDI und Module

Um den geringen Speicherbedarf der notenbasierten Wiedergabe mit der hohen Qualität des Samplings zu vereinen bietet es sich an, einzelne Klänge zu digitalisieren und anhand der Noten wiederzugeben. Dies führte zur Entwicklung der Wavetable-Karten.

Einfachere Geräte nach dem GeneralMIDI-Standard bieten hierbei 127 verschiedene Instrumente zuzüglich eines Schlagzeugsatzes, deren Samples insgesamt einen Umfang von 1MB bis 8MB haben. Da die Stimmen in Hardware gemixt werden bleibt der Bedarf an Rechenzeit gering, allerdings läßt sich trotz vieler Tricks kaum ein überzeugendes und flexibles Klangbild in derart wenig Speicher unterbringen.

Im Gegensatz dazu bringen die Module-Formate (*mod,s3m,xm,it,669,far,...*) die einzelnen Klänge in der Datei selber mit. Dadurch wird das beste Qualitäts-Speicherbedarf-Verhältnis erzielt, allerdings kann auch der Rechenaufwand schnell enorm werden. Einige Soundkarten wie die Gravis Ultrasound oder die Soundblaster-AWE-Serie können dank eines RAM-basierten Wavetables den Rechenaufwand stark senken. Mangels Verbreitung hat sich bisher allerdings außer dem Audiochip des Commodore Amiga keiner durchsetzen können. Mit zunehmender Verbreitung von Soundkarten nach DLS-Standard wird diese Technik aber wieder interessanter.

Kapitel 4

Hardwarebeschleunigung - Hardwaremixing

RAM-basierte Wavetablehardware wird oft auch mit dem Schlagwort "hardwarebeschleunigt" angeboten. Dabei gibt es einerseits Geräte mit kleinem (256kb - 28MB) lokalen RAM-Speicher und hohen Zugriffszeiten als auch Hardware, die den Hauptspeicher des Systems verwendet. Letztere überzeugen mit niedrigem Preis und schnellen Zugriffszeiten auf große Speicherbereiche, können aber auch einen erheblichen Teil der Systembandbreite für sich beanspruchen. Daher ist die hardwarebeschleunigte Wiedergabe weniger aus Sicht der eingesparten Rechenzeit sondern vor allem wegen der reduzierten Latenzzeit und in Hardware implementierten Effekten wie Hall oder Chorus interessant.

Kapitel 5

CD-ROM + CD-DA

Bei Software, die auf CD-ROM ausgeliefert wird bietet es sich an, einen Teil der CD für normale CD-Audiospuren zu verwenden. Diese werden nach Abschicken eines einfachen Kommandos vom CD-Laufwerk selbstständig und mit voller CD-Qualität wiedergegeben. Auch seitens der Musiker liegt ein in der Handhabung vollkommen vertrautes Speichermedium vor.

Problematisch sind allerdings die geringe Spielzeit bei hohem Platzbedarf und die geringe Flexibilität einer CD. Ein Nachladen von Daten während der Wiedergabe ist nicht mehr möglich, vor allem stören die langen Zugriffszeiten und das Fehlen einer exakten und zugleich schnellen Bestimmung der Wiedergabeposition.

Kapitel 6

Raumklang

Die Simulation eines räumlichen Klangbildes erfordert zwei Bestandteile: Die Positionierung von Schallquellen in einem 2- oder 3-dimensionalen Raum als auch die Nachbildung der akustischen Charakteristik der Umgebung.

6.1 Verschiedene Varianten

6.1.1 Kunstkopfstereophonie, QSound, Virtual Surround

Die Ortung von Schallquellen erfolgt anhand der absoluten Lautstärke, der Lautstärke- und Phasendifferenz sowie der Frequenzcharakteristik an beiden Ohren. Demnach genügen zwei Schallwandler, um ein räumliches Klangbild zu vermitteln. Dies wurde bei der Kunstkopfstereophonie zur Aufnahme mit Erfolg praktiziert. Problematisch ist aber, daß dieses Verfahren in Verbindung mit Kopfhörern bei Kopfdrehungen die Schallquellen realitätsfremd mitdreht, während bei Verwendung von Lautsprechern die fehlende Trennung zwischen den Schallquellen beider Ohren die wahrzunehmenden Schallquellen nicht sauber reproduziert werden können. Für die Positionierung in Echtzeit kommt erschwerend hinzu daß die zur Simulation der richtungsabhängigen Abschattung eingesetzten HRTFs (Head Related Transfer Functions) trotz hohem Rechenaufwand nur grob an die tatsächlichen Gegebenheiten des Hörers angepaßt sind.

6.1.2 Dolby Surround

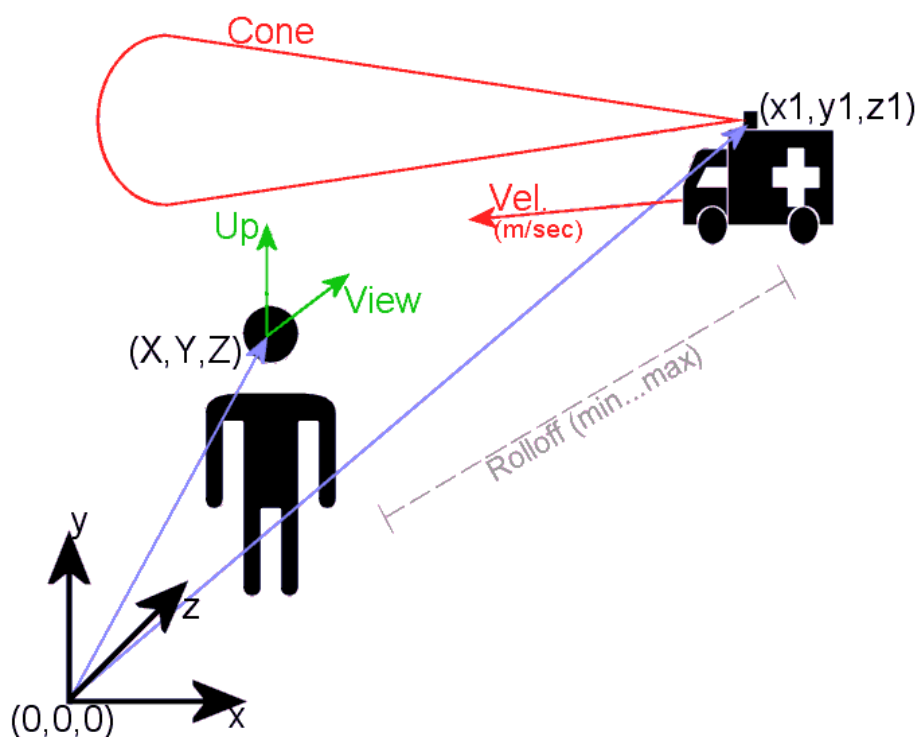
Das aus der Kinotechnik stammende Verfahren besitzt den Vorteil, vollkommen kompatibel zu normalen Stereoton zu sein. Es handelt sich allerdings um kein vollständiges 2D-Verfahren, sondern nur um eine Erweiterung des Stereotons um Umgebungsgeräusche, die im Idealfall nicht zu lokalisieren sind.

6.1.3 Quadrophonie, 5.1 / 6.1 / 7.1 etc.

Durch Verwendung von 4 und mehr Lautsprechern in einer Ebene ist eine echte 2D-Positionierung möglich. Die konkrete Anzahl an Lautsprechern ist variabel, die Verteilung

der Signale wird kann vom API oder der Hardware transparent erledigt werden. Zur Lokalisierung in der Höhe können HRTFs eingesetzt werden, welche aufgrund der in der Höhe weitgehend konstanten Kopfposition weitgehend unkritischer sind als bei reinen HRTF-basiertem Raumklang.

6.2 DirectSound3D und Co.



Zur Berechnung der Wiedergabeposition einer Schallquelle wird sowohl die Position als auch Orientierung des Hörers im 2D- der 3D-Koordinatensystem bekannt sein. Ebenso wird die Position der Schallquelle benötigt. Für Dopplereffekte wird darüber hinaus die Geschwindigkeit der Quelle in absoluten Angaben benötigt. Da die APIs nur die Schallausbreitung, nicht aber die Bewegung der Schallquellen berechnen müssen bewegte Schallquellen entweder regelmäßig in Position und Geschwindigkeit aktualisiert werden oder durch eine zeitliche Folge mehrerer solcher Schallquellen nachgebildet werden. Viele Schallquellen haben eine spezielle, nicht kugelförmige Abstrahlcharakteristik. Diese kann mit Abstrahlkegeln nachgebildet werden. Keine physikalische, sondern eine technische Bedeutung besitzt die Angabe der minimalen und der maximalen Distanz: Mit diesen läßt sich die absolute Lautstärke der auf den Dynamikumfang des Samplingverfahrens angepaßten Effekte nachbilden. Dabei kennzeichnet die minimale Distanz die Entfernung, ab der ein Effekt seine volle Lautstärke erreicht hat, während die maximale Distanz angibt ab wann ein Effekt unhörbar wird.

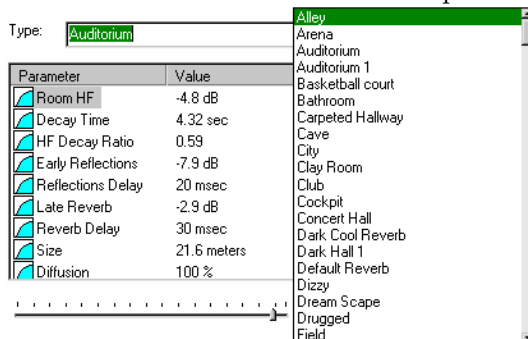
6.3 Akustischer Nachbau von Räumen

Der Klang eines Raumes läßt sich anhand der Berechnung der Schallausbreitung und Reflektion von Schallwellen in diesem Raum nachbilden. Dazu wird der Raum, ähnlich wie in der 3D-Graphik, mit all seinen schalldurchlässigen, reflektierenden und schallabsorbierenden Flächen nachgebaut. Aufgrund der großen Schallwellenlänge im Meterbereich genügt eine grobe Nachbildung, wobei allerdings aufgrund möglicher Mehrfachreflexionen dennoch ein hoher Rechenaufwand entsteht.



6.3.1 Vereinfachte Raumakustik

Einen weniger realistischen, aber dennoch guten Effekt bei weitaus geringerem Erstellungs- und Rechenaufwand liefert die Nachbildung eines Raumes durch dessen Hallmodell. Durch die große Zahl an Einstellungen lassen sich praktisch alle Umgebungen nachbilden. Zur vereinfachten Handhabung existieren zu den wichtigsten Umgebungen fertige Presets, so daß von den Spieleentwicklern nur noch die Angabe des Presets samt Effektstärke benötigt werden. Da die Bestimmung des aktiven Hallmodells nur eine grobe Positionsbestimmung benötigt eignet sich das Verfahren gut zur einfachen Integration in bestehenden Quelltext. Zudem können bessere Soundkarten den Hall komplett in Hardware berechnen.



Kapitel 7

Interaktive Musik

Zunehmend von Bedeutung wird auch die Anpassung nicht nur der Effekte, sondern auch der Hintergrundmusik an das momentane Spielgeschehen.

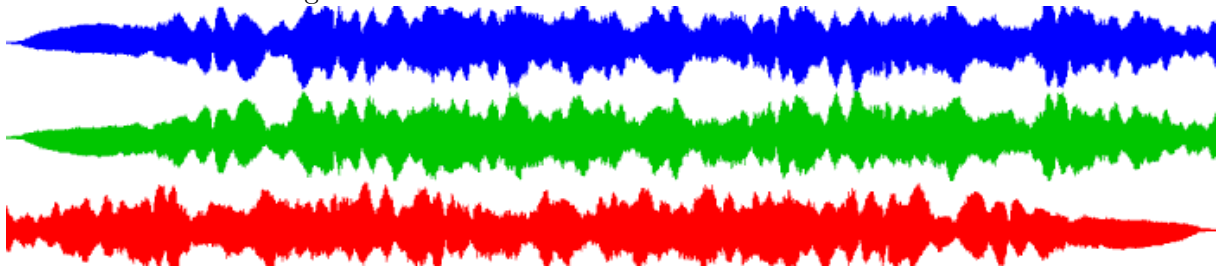
7.1 Sequentielle Titelwiedergabe

Ein Verfahren, daß mit jeder Art von Musikwiedergabe möglich ist stellt die sequentielle Wiedergabe verschiedener Titel dar. Problematisch sind jedoch ausklingende Noten, die mitunter abgeschnitten werden können als auch die fehlende Bindung der Sprungmarken an die Musikdateien beim nachträglichen Ändern der Musik.

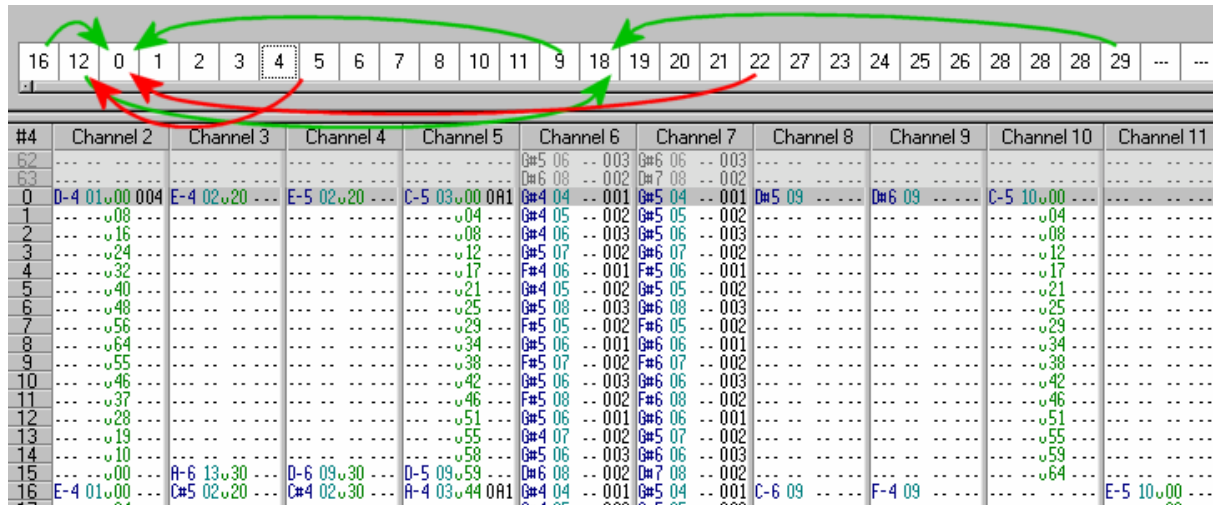


7.2 Parallele Titelwiedergabe

Mehrere Tonkanäle gleichzeitig erlauben ein freies Zusammensetzen verschiedener Titel und Elemente wie bei einem Mehrkanal-Mischpult. Der dafür nötige hohe Datendurchsatz als auch die feste zeitliche Bindung der einzelnen Kanäle schränken die Einsatzgebiete insbesondere bei umfangreichen Melodien ein.

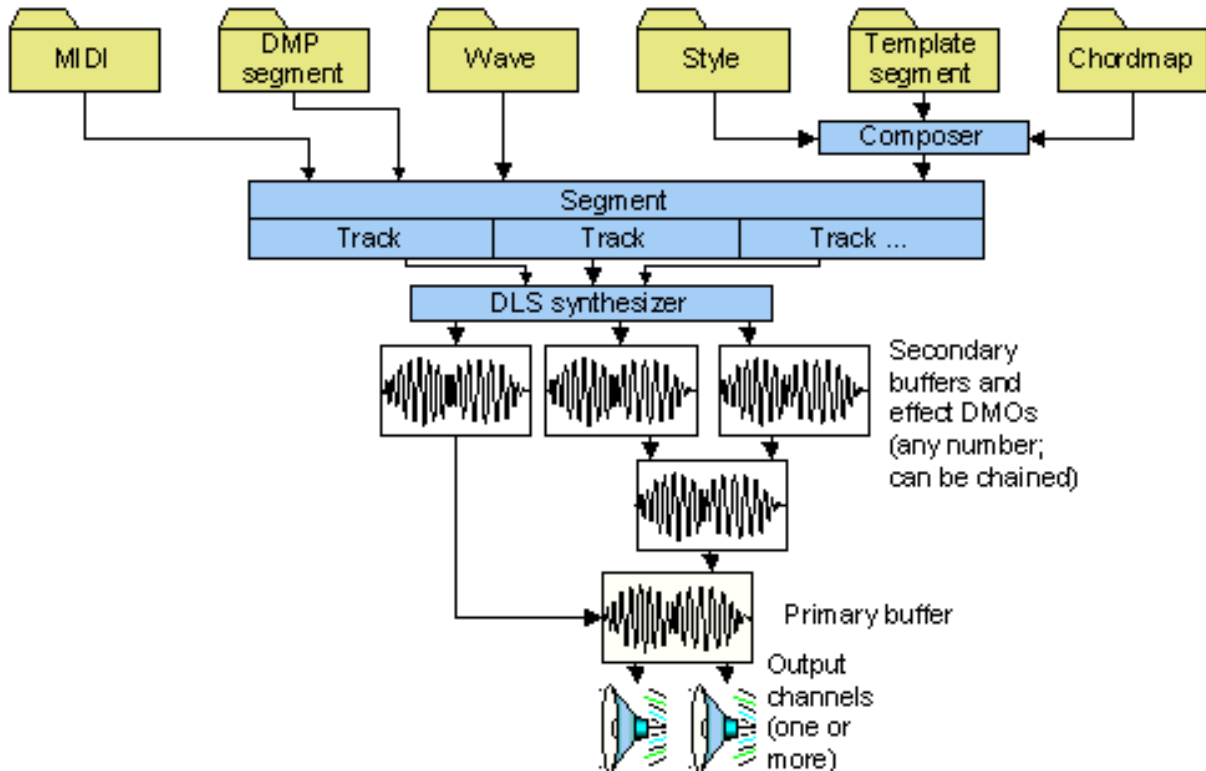


7.3 Module



Von Haus aus geeignete Voraussetzungen für interaktive Musik bilden die Module-Formate. Bei diesen sind die Noten schon in einzelnen Blöcke, den Patterns, angeordnet. Die Patterns können in beliebiger Reihenfolge angeordnet werden, zudem ist es möglich zwischen Patterns zu springen. Desweiteren sind die Noten in den einzelnen Patterns wieder in Spuren (Tracks) angeordnet, so daß diese wie bei der parallelen Wiedergabe beeinflusst werden können. Zur Synchronisation können Callbacks bei Erreichen des Ende eines Patterns oder in Verbindung mit speziellen Zxx-Markierungen gesetzt werden. Dieses Verfahren ist noch relativ gut handhabbar bei gleichzeitig relativ guter Flexibilität.

7.4 DirectMusic



Am konsequentesten ist der Ansatz von DirectMusic. Hierbei wird die Musik aus Sequenzen und Motiven aufgebaut. Motive sind spezielle Sequenzen, die als z.B. Erkennungsmerkmal dynamisch hinzugefügt werden. Sequenzen können wiederum Sequenzen enthalten, auch besteht die Möglichkeit, Sequenzen mit verschiedener Grundtonhöhe abzuspielen. Als Material für Sequenzen sind neben speziellen Formaten auch MIDI-Dateien oder samplebasierte Formate geeignet. Mit Skripten und Zufallsgeneratoren sind weitere Variationen möglich.

Neben den derzeit vielfältigsten interaktiven Möglichkeiten bietet DirectMusic auch die Integration von Soundeffekten und Raumklang. Neben dem hohen programmier-technischen Aufwand ist allerdings auch eine außergewöhnlich hohe Zusammenarbeit von Musikern und Spieledesignern nötig, da die endgültige Komposition nicht mehr im Tonstudio, sondern zur Laufzeit stattfindet.

Kapitel 8

Adressen

- Skript zur Digitalen Audioverarbeitung von der Universität Hamburg:
<http://tech-www.informatik.uni-hamburg.de>
- Entwicklung der PC-Soundkarten:
www.crossfire-designs.de/link/soundcards
- FAQs und Tutorials:
<http://thorkildsen.no/faqsys/>
- Housemarque Audio System:
www.s2.org/hmqaudio
- FMOD:
www.fmod.org
- BASS:
www.un4seen.com
- DirectSound/DirectMusic:
www.msdn.microsoft.com
- OggVorbis:
www.vorbis.com
- ModPlug:
www.modplug.com
- Renoise:
www.renoise.com
- 4Players:
www.4players.de