

# **Ludoaudiovisuelle Harmonie**

-

## **Die komplexe Verbindung von Musik und Computerspiel**

Schriftliche Hausarbeit  
für die Masterprüfung der Fakultät für Kulturwissenschaften  
im Fach Medienwissenschaft  
an der Universität Paderborn  
(Prüfungsordnung vom 14.06.2019)

vorgelegt von

Wulf, Christopher

21.06.2024

Prof. Dr. Ralf Adelman  
Prof. Dr. Christina Bartz

# Inhaltsverzeichnis

1 Einleitung.....	3
2 Methodik.....	5
3 Audiovisuelle Harmonie.....	7
3.1 Herleitung und Diskussion.....	7
3.1.1 Bild und Ton zusammen denken.....	7
3.1.2 Harmonie der vertikalen Ebene.....	8
3.1.3 „Noten“ und Synchronisationspunkte.....	9
3.1.4 Bedeutungserlangung.....	10
3.1.5 Kadenzen, Antizipation und Erfüllung.....	13
3.1.6 Definition <i>audiovisuelle Harmonie</i> .....	16
3.2 Audiovisuelle Harmonie im Film.....	17
3.3 Audiovisuelle Harmonie im Fernsehen.....	19
3.4 Audiovisuelle Harmonie in Musikvideos.....	21
4 Ludoaudiovisuelle Harmonie.....	24
4.1 Ludo-Ebene.....	24
4.2 Definition: ludoaudiovisuelle Harmonie.....	25
4.3 Diskussion.....	27
4.3.1 Unterschiede audiovisuelle Sequenz Film und Computerspiel.....	27
4.3.2 Gegenwärtigkeit, Subjektivität und Immersion.....	30
4.3.3 Gemeinsamkeiten Musik und Computerspiel.....	33
5 Untersuchung.....	34
5.1 Musik generiert Spielaspekte.....	36
5.1.1 Beat Hazard 2.....	36
5.1.2 Audiosurf 2.....	39
5.1.3 Diskussion: Musik generiert Spielaspekte.....	43
5.2 Rhythmus als Fundament für Gamedesign.....	46
5.2.1 Metal: Hellsinger.....	46
5.2.2 Hi-Fi RUSH.....	49
5.2.3 Diskussion: Rhythmus als Fundament für Gamedesign.....	54
5.3 Synästhesie als Designziel.....	56
5.3.1 Rez.....	56
5.3.2 Tetris Effect.....	59
5.3.3 Diskussion: Synästhesie als Designziel.....	64
6 Fazit und Ausblick.....	66
Quellenverzeichnis.....	69
Abbildungsverzeichnis.....	79
Eigenständigkeitserklärung.....	81

## 1 Einleitung

Musik wird in Computerspielen auf sehr vielfältige Weise verwendet; sei es als Hintergrundmusik, die die Spielwelt charakterisiert wie in *JET SET RADIO FUTURE* (Japan 2002, Sega, Xbox), als taktgebende Spielmechanik in Rhythmusspielen wie *PaRappa the Rapper* (Japan 1996, Sony Computer Entertainment, Playstation) oder als individuelle Komposition, die sich aus den Handlungen des Spielers und Objekten in der Spielwelt generiert wie in *Proteus* (UK 2013, Twisted Tree, Windows PC). Generell lassen sich die Techniken zur Musikverwendung bzw. -integration in Spielen auf einer Skala der Linearität anordnen, bei der das eine Ende von linearer Musik und das andere Ende von durch den Akt des Spielens selbst hervorgerufene generative Musik markiert ist (Sites/Potter 2018). Dynamische Musik, verstanden als Oberkategorie für Musik, die sowohl auf Handlungen des Spielers und/oder Veränderungen des Spielstatus reagiert (Collins 2008, 4), ist hierbei als das Ziel von gutem Game- und Sound-Design anzusehen (Harrer 2018, 77). Im Laufe der Jahre haben Entwickler eine Bandbreite an unterschiedlichsten Techniken entwickelt, wie „branching, layering, transitions, generative“ (van Geelen 2008, 96), um dieses Ziel zu erreichen. Wie noch gezeigt wird, ist aber nicht nur die Reaktion der Musik auf das Spiel eine Möglichkeit Musik und Spiel miteinander zu verbinden, sondern auch eine in beide Richtungen agierende Wechselwirkung kann erreicht werden.

In allen Fällen jedoch liegt eine Schwierigkeit vor, die in der Logik der beiden Medien begründet ist: Musik ist im westlichen Kulturverständnis linear, Spiele jedoch non-linear und interaktiv (Herber 2008, 104). Vor allem der Faktor Zeit spielt hier eine Rolle, denn i.d.R. ist die Spiellänge nicht vordefiniert (Collins 2008, 140), und eine Spielsitzung hat von Spieler zu Spieler eine unterschiedliche Länge, wodurch auch die Musikkwiedergabe jedes Mal anders ist (Summers 2016, 25).

Wenn also das Game-Design immer auch das Sound-Design vorgibt (Collins 2008, 90) und die Modularität von Musik und Spiel Grundlage für dynamische Musik ist (Medina-Gray 2016, 53), wie kann Musik dann konkret in die Logik von Spielen implementiert werden und wie kann diese Verbindung wissenschaftlich erfasst und beschrieben werden?

Im Idealfall liegt zwischen Musik und Spiel eine echtzeitliche und wechselwirkende Verbindung vor, die mittlerweile auch von Spielern gewissermaßen erwartet wird (Summers 2016, 137), und mitunter im Phänomen des *Musicking* mündet, bei dem Spielen gleichbedeutend mit Aufführen, Komponieren und Improvisieren von Musik ist (Svec 2008, 2). Die Hauptthese dieser Arbeit lautet daher wie folgt:

Musik kann in Spielen so genutzt werden, dass sie in Synchronizität mit an-

deren Spielelementen vorkommt, somit eine Harmonie mit diesen eingeht und so letztendlich die Harmonie der Musik erweitert, indem Aspekte des Spiels in diese mit eingeschlossen werden. So entstehen Harmonien, die teilweise auch im Film vorkommen (Bild und Ton), andere hingegen sind nur durch die interaktive Form des Computerspielmediums und/oder seiner besonderen Benutzerschnittstellen möglich (Gameplay, Kinetik, Haptik). Im Idealfall entstehen Harmonien auf mehreren Ebenen, die dementsprechend noch wirkungsvoller sind. Diese echtzeitlichen Verbindungen zwischen Musik und Spiel nenne ich *ludoaudiovisuelle Harmonien* und sie ergeben sich wenn Elemente der Musik in Synchronizität mit Elementen des Spiels und seiner multimodalen<sup>1</sup> Darstellung treten, um so die Harmonie der Musik zu erweitern, wodurch sich Musik und Spiel „zu einer unverwechselbaren Signatur“ (Fischer/Schlüter 2009, 355) miteinander verbinden. Grundlage für den Begriff der *ludoaudiovisuellen Harmonie* ist der von Michel Chion beschriebene *audiovisuelle Vertrag* (Chion 1994), der die tiefe Beziehung zwischen Ton und Bild in audiovisuellen Medien wie dem Film beschreibt. Auch wenn es bedeutende Unterschiede zwischen nicht-interaktiven und interaktiven audiovisuellen Medien gibt, auf die ich konkret eingehen werde, so wurde bereits mehrfach festgestellt, dass Ton und Musik in Spielen durchaus viele Gemeinsamkeiten mit der Verwendung und den Produktionstechniken im Film haben (Stockburger 2003, 1; Huiberts/van Tol 2008; Summers 2016, 57), weswegen eine vertiefte Auseinandersetzung mit nicht-interaktiven audiovisuellen Medien hier durchaus sinnvoll ist.

Diese Arbeit beginnt daher mit einem theoretischen Teil, in dem es um die Herleitung und Diskussion des Prinzips *audiovisuelle* und *ludoaudiovisuelle Harmonie* geht, sowie der Herausstellung von Gemeinsamkeiten und Unterschieden des Prinzips bei nicht-interaktiven und interaktiven Medien. Zum Schluss folgt eine ausführliche Untersuchung von sechs Spielen, die wie wohl keine Anderen besondere, echtzeitliche Verbindungen von Musik und Spiel aufweisen und so das Prinzip der *ludoaudiovisuellen Harmonie* und ihre Entstehung verdeutlichen: BEAT HAZARD 2 (UK 2018, Cold Beam Games, Windows PC), AUDIOSURF 2 (USA 2015, Dylan Fitterer, Windows PC), METAL: HELLSINGER (Norwegen 2022, Funcom, Windows PC), HI-FI RUSH (USA 2023, Bethesda, Windows PC), REZ (Japan 2001, Sega, Dreamcast) und TETRIS EFFECT (Japan 2018, Enhance, Windows PC). Zu der Untersuchung der Spiele werden zwei weitere Thesen aufgestellt:

Rhythmus eignet sich besonders gut zur Herstellung *ludoaudiovisueller Harmonien*, da sich der Rhythmus der Musik selten(er) ändert und Spiele so eine Konstante in der Musik nutzen können, unter anderem auch als Fundament für Gameplaysysteme.

---

1 „Im Kontext der Wahrnehmungspsychologie bezieht sich Multimodalität auf die Gleichzeitigkeit mehrerer Wahrnehmungssinne; betont wird dabei auch die Untrennbarkeit von Sehen, Hören, Tasten, Riechen und Schmecken“ (Herwig 2017, 79).

Eigens für ein Spiel komponierte und produzierte Musik, die sehr adaptiv auf den Spielverlauf reagieren kann, eignet sich deutlich besser für *ludoaudiovisuelle Harmonie* als ein System bei dem aus vom Spieler frei auswählbarer Musik Spielaspekte per Algorithmus generiert werden.

## 2 Methodik

Eine streng musikwissenschaftliche Perspektive auf Ton und Musik in Spielen ist problematisch, da sie mit engen Musikdefinitionen arbeitet und so die besonderen Eigenschaften von Spielen mitunter nicht immer berücksichtigt (Fischer/Schlüter 2009, 353). Stattdessen soll hier die Erfahrung des Spielers und die Rezeption der Spiele im Fokus stehen. Trotzdem ist es wichtig vorher zu definieren was überhaupt unter Begriffen wie Musik und Harmonie zu verstehen ist.

Auch wenn Spiele hier als Untersuchungsgegenstand dienen, wäre eine zu sehr an Spielen orientierte Definition zu eng. So fasst Stockburger in seiner auf objektbasierter Programmierung anlehnenen Typologie Musik als eine Sammlung von *sound-objects* auf, die immer zum nicht-diegetischen<sup>2</sup> Teil der Spielumgebung gehören (Stockburger 2003, 6). Dass sich Musik aber auch aus anderen, auch diegetischen Tönen zusammensetzen kann oder durch direkte Verbindungen bspw. mit dem Gameplay als diegetisch betrachtet werden kann, wird in dieser Arbeit später noch gezeigt. Eine allgemeinere Definition von Musik bietet hingegen der Duden: „Kunst, Töne in bestimmter (geschichtlich bedingter) Gesetzmäßigkeit hinsichtlich Rhythmus, Melodie, Harmonie zu einer Gruppe von Klängen und zu einer stilistisch eigenständigen Komposition zu ordnen“ (Duden.de 2024a). Wichtig sind vor allem die drei Grundkomponenten der Musik, die auch von Arsenault in seiner Untersuchung zum speziellen Controller von GUITAR HERO (USA 2005, RedOctane, Playstation 2) herausgestellt werden: „the basic components of music need to be distinguished in order to properly evaluate the accuracy of *Guitar Hero* and its controller. At the core, they are *melody*, *harmony*, and *rhythm*“ (Arsenault 2008, 2, kursiv im Original). Speziell Rhythmus und vor allem Harmonie sind für diese Untersuchung wichtig und werden noch vertieft besprochen. Zunächst reicht eine kurze Definition wie Wedel sie von Lefebvre ableitet: „die *Melodie* als lineare Sequenz von Tönen in zeitlicher Abfolge; die *Harmonie* als simultanes Ineinanderklingen von Tönen; [...] den *Rhythmus* als der Platzierung der Töne und ihrer relativen Länge zueinander“ (Wedel 2020, 115; kursiv i.O.).

Bevor jedoch ein Blick auf die komplexe Verbindung von Musik und Spiel

---

2 „Diegese wird oft verkürzt als Bezeichnung der raumzeitlichen Beziehungen der erzählten Welt, als Bezeichnung ihrer modellhaften Einheit, als räumlich-zeitliches Universum der Figuren bzw. Charaktere im Sinne des weichen Begriffs der erzählten Welt' gebraucht.“ (Wulff 2022a).

gerichtet wird, sollen in Kapitel drei basierend auf Chions „specific perceptual mode of reception [...] audio-vision“ (Chion 1994, XXV) und in Anlehnung an Bakels Begriff der *audiovisuellen Rhythmen* (Bakels 2016) die Kernaspekte einer *audiovisuellen Harmonie* in nicht-interaktiven Medien wie dem Film bestimmt werden. Chions Überlegungen zum audiovisuellen Bild Anfang der 1990er Jahre waren besonders, da er anders als zuvor in der Filmwissenschaft die Bedeutung der Einheit von Bild und Ton hervorgehoben hat (Chion 1994, 97). Er erkannte, dass Bild und Ton in audiovisuellen Medien nicht getrennt voneinander analysiert werden können. Chion betont die Bedeutung von Momenten in denen Bild und Ton in Synchronisation treten und wie Zuschauer diese Momente antizipieren. Auch wenn er nur selten auf Musik im Film eingeht, ist seine Verwendung von musiktheoretischen Begriffen wie Kadenz und Akkord, mit denen er immer wieder metaphorisch das audiovisuelle Bild beschreibt, auffällig (ebd. 55). Das Kapitel wird abgeschlossen mit einer Definition *audiovisueller Harmonie* und Beispielen aus Film, Fernsehen und Musikvideos.

Kapitel vier nutzt dann die Grundlagen aus Kapitel drei um den Begriff der *ludoaudiovisuellen Harmonie* zu etablieren. So wird nach einer Definition und Beispielen zunächst auf die Unterschiede der audiovisuellen Sequenz in Film und Spiel eingegangen und warum die Einheit von Bild und Ton in Spielen sogar noch wichtiger ist. Auch die besonderen Eigenschaften des Computerspielmediums werden besprochen, da diese das Verhältnis zur Musik entscheidend prägen. Das Kapitel wird abgeschlossen mit einer Verdeutlichung der Gemeinsamkeiten von Musik und Spiel und warum eine enge Verbindung der Beiden besonders für Spiele sinnvoll ist.

Abschließend werden in Kapitel fünf drei Spielpaare untersucht, die das Prinzip der *ludoaudiovisuellen Harmonie* für ihr Gameplaydesign nutzen. Beim ersten Spielpaar bestehend aus BEAT HAZARD 2 und AUDIOSURF 2 steht die Einbindung von linearer Musik in Spielstruktur und -darstellung anhand von Algorithmen im Vordergrund. Das zweite Spielpaar, METAL: HELLSINGER und HI-FI RUSH, hingegen zeigt wie Rhythmus und Harmonie auch außerhalb von klassischen Rhythmusspielen als Fundament für Gameplay genutzt werden können. Die Spiele REZ und TETRIS EFFECT bilden das letzte Paar und zeigen wie generative Musik mit Spielaspekten verbunden werden kann, um durch *ludoaudiovisuelle Harmonien* eine Simulation von Synästhesie zu erreichen. Als theoretische Grundlagen für die Untersuchung, die im Detail erst in Kapitel fünf parallel zur Untersuchung besprochen werden, dienen die Unterscheidung ludologischer Ziele (Debus et al. 2020), die Segmentierung von Gameplay (Zagal et al. 2008) und ein Verständnis von Gameplay basierend auf dem Affordanzbegriff (Linderoth 2011).

Die praktische Untersuchung von Spielen hinsichtlich Verbindungen von Musik und Spielaspekten kann mitunter sehr schwierig sein, da man seine Aufmerksamkeit sowohl auf das Spielen als auch auf die Musik aufteilen muss

(Summers 2016, 35). Hierzu wurde eine Abwandlung der Masking-Methode Chions angewandt (vgl. Chion 1994, 187). Glücklicherweise sind Spiele bereits seit den 1990er Jahren in der Regel mit weitreichenden Einstellungsmöglichkeiten auch bzgl. des Tons ausgestattet. So wurden bei der Untersuchung bestimmte Spielabschnitte oder Level mehrfach mit unterschiedlichen Einstellungen gespielt, bspw. einmal ohne Musik oder ohne Toneffekte, während eine Aufzeichnung der Sequenzen (visuell, auditiv, kinetisch) mit parallel ausgeführten Programmen<sup>3</sup> stattfand. Hinterher konnten so die unterschiedlichen Aufzeichnungen verglichen werden.

Alle Spiele wurden mit den höchsten Grafikeinstellungen bei konstanter Bildrate von 60hz mit 1080p Auflösung gespielt. Für das Durchspielen der Kampagnenmodi, soweit vorhanden, wurden Kopfhörer verwendet, sowie von den Spielen vorgegebene Standardeinstellungen für Ton und Musik. Als Eingabegerät wurde ein Microsoft Xbox Series X Controller genutzt, da dieser weites gehend als Standard angesehen werden kann (Steam 2021). Leider ist die Angabe von Testsystemen bei Untersuchungen in der Medienwissenschaft bzw. Computerspielforschung immer noch keine Selbstverständlichkeit, obwohl bspw. die Nutzerumgebung erheblichen Einfluss auf Wahrnehmung und Rezeption haben kann (Stockburger 2003, 2).

### **3 Audiovisuelle Harmonie**

#### **3.1 Herleitung und Diskussion**

##### **3.1.1 Bild und Ton zusammen denken**

Egal ob wir den Fernseher einschalten, ins Kino gehen, Computerspiele spielen oder sogar selbst mit unserem Smartphone ein Video drehen, um es umgehend auf einer sozialen Plattform hochzuladen, in unserer heutigen Medienwelt des 21. Jahrhunderts erscheint die Vereinigung von akustischen und visuellen Artefakten etwas Selbstverständliches zu sein (Daniels/Nauermann 2013, 217). Diese Vereinigung von Bild und Ton in Medien wirkt auf uns so natürlich, dass wir sie in der Regel nicht hinterfragen, auch wenn erst die Medientechnologien des 19. Jahrhunderts es ermöglichten, dass Töne speicherbar und Bilder bewegbar wurden (ebd. 218). In der wissenschaftlichen Auseinandersetzung, bspw. mit der Filmmusik, ist es bis heute „grundlegenden Prämissen – Bild und Ton als zwei disparate Felder zu betrachten“ (Bakels 2016, 33). Michel Chion war Anfang der 1990er Jahre einer der ersten Filmwissenschaftler, der mit seinem Konzept der *audio-vision*

---

<sup>3</sup> Open Broadcasting Software (2012-2024, Anonym/Open Source) für das Visuelle, Audacity (2000-2024, Anonym/Open Source) für die Wellenform des Ausgabetons und GamePad Viewer (USA 2024, Christopher Rodriguez) zur Darstellung der kinetischen Spielereingaben. Ein Programm zur Darstellung bzw. Erfassung von Vibration/Haptik ist bisher nicht entwickelt worden.

der Einheit von Bild und Ton eine Bedeutung zusprach, die mehr war als die Summe seiner Einzelteile. Chion suchte nach einem „genuinen Effekt audiovisueller Wahrnehmung [der] Wahrnehmungseindrücke ermöglicht, die keine der beiden beteiligten Sinnesmodalitäten für sich genommen zu vermitteln vermag“ (ebd. 34). Sicherlich ist auch die separate Untersuchung der Modalitäten sinnvoll, aber im Film und auch anderen audiovisuellen Medien sind Bild und Ton nicht zufällig miteinander verbunden, sondern werden zu einer bestimmten Komposition vereinigt. Daher wäre es auch zu kurzfristig sich bei der Untersuchung von Wirkungsweisen bspw. der Filmmusik allein auf die physiologischen Eigenschaften der Musik zu fokussieren, denn auch wenn diese nicht unwichtig sind, so ist „ihre Einbindung in einen melodischen, harmonischen oder rhythmischen Wahrnehmungszusammenhang“ (ebd. 30) entscheidend. Musik spielt hier eine besondere Rolle, denn ihre Dynamiken brauchen Zeit um sich aufzubauen und zu entfalten. Das audiovisuelle Bild jedoch ist durch das gleichzeitige Wahrnehmen beider Modalitäten bestimmt, was Bakels wie folgt ausdrückt: „Die Wahrnehmung filmischer Bilder setzt voraus, dass der Zuschauer Bild und Ton im Sinne eines gemeinsamen Ganzen, dem Film, aufeinander bezieht“ (ebd. 159). Die Wahrnehmung von Bild und Ton geschieht synchron, wodurch „Bewegungen – des Bilds, im Bild, der Musik – als *eine* Bewegung wahrgenommen werden“ (ebd. 160, kursiv i.O.). Das Zusammenwirken der Modalitäten erfolgt also synchron und kontinuierlich. Bakels erfasste bereits ausführlich den kontinuierlichen Teil dieses Zusammenwirkens von Bild und Ton im Film und prägte den Begriff des *audiovisuellen Rhythmus*, mit dem er Filmsequenzen, die mit populärer Musik unterlegt sind, auf eine gemeinsame kinetische Gestalt von Musik und Bild untersuchte. Interessanter Weise jedoch schenkte er dem synchronen Teil der audiovisuellen Wahrnehmung weitaus weniger Beachtung, was aber der Kern dieser vorliegenden Arbeit sein soll.

### 3.1.2 Harmonie der vertikalen Ebene

Grundsätzlich lässt sich Musik in eine horizontale und eine vertikale Achse herunterbrechen. Die horizontale Achse beschreibt den zeitlichen Verlauf von Dynamiken mit Rhythmus und Melodie, während die vertikale Achse den Moment der Gleichzeitigkeit beschreibt, wenn Töne synchron auftreten und Harmonien bilden (Bakels 2016, 25). Sowohl in seinen Überlegungen als auch in seinen Filmen übertrug bereits der sowjetische Filmemacher Sergej M. Eisenstein dieses Verständnis der Musik auf den Film (Wulff 2023). Auch der Film hat eine horizontale, zeitliche Ebene mit Dynamiken wie den Bewegungen im Bild bzw. des Bilds oder Schnittfolgen, aber auch eine vertikale Ebene, in der Bild und Ton zusammen in Gleichzeitigkeit erscheinen. Während Bakels sich wie bereits erwähnt auf die horizontale Ebene konzentriert, um den Begriff des *audiovisuellen Rhythmus* zu definieren,

so ist für Chion die vertikale Ebene das Entscheidende beim Wahrnehmungszusammenhang des filmischen Bildes. Der Harmoniebegriff erscheint bereits bei Chion, jedoch in rein analogischer Weise zur Beschreibung von Verhältnissen zwischen Bild und Ton (Chion 1994, 36-37). Generell nutzt Chion auffällig viele Musikmetaphern zur Beschreibung des audiovisuellen Bildes. Entscheidend ist, dass Chion der vertikalen Ebene des Films mehr Bedeutung zuschreibt als der horizontalen Ebene: „in the cinema, harmonic and vertical relations [...] are generally more salient — i.e., the relations between a given sound and what is happening at that moment in the image“ (ebd. 36). Die Bedeutungserlangung des audiovisuellen Bildes sieht Chion also mehr in der Gleichzeitigkeit von Bild und Ton und nicht im Verlauf der beiden Modalitäten auf der horizontalen Ebene. Es sei also wichtiger was zu einem bestimmten Zeitpunkt zu sehen und zu hören ist und wie beide Modalitäten in einem Moment zusammenfallen und wirken. Auch wenn Chion der Filmmusik durchaus Beachtung schenkt, so verbindet er seine Erkenntnisse zur vertikalen Ebene des Films nicht mit der vertikalen Ebene der Musik, also ihrer Harmonie. Doch ebenso wie die Tonebene allgemein, kann auch die Musik eine besondere synchrone Verbindung mit dem Bild eingehen und so Bedeutung generieren. Wenn dies geschieht, wird das audiovisuelle Bild nicht einfach nur durch Musik erweitert und untermalt, stattdessen nutzt das audiovisuelle Bild eine bereits in der Musik vorkommende Harmonie und erweitert diese durch die visuelle Modalität.

### **3.1.3 „Noten“ und Synchronisationspunkte**

Doch wenn man davon ausgeht, dass ein audiovisuelles Medium wie der Film die Harmonie der Musik aufgreifen und erweitern kann, wann genau bildet sich dann diese erweiterte Harmonie und aus welchen Elementen setzt sie sich zusammen? Zunächst einmal ist wie bei Chion von Synchronisationspunkten auszugehen, die er wie folgt beschreibt: „a salient moment of an audiovisual sequence during which a sound event and a visual event meet in synchrony“ (Chion 1994, 58). An einem bestimmten Zeitpunkt treffen also ein Bild- und Tonereignis zusammen. In Analogie zur Musik kann man diese einzelnen Ereignisse der zwei unterschiedlichen Ebenen auch als Noten bezeichnen. Bakels spricht bei den Elementen des filmischen Bildes, die einen *audiovisuellen Rhythmus* bilden können, auch bereits von Noten und meint damit schließlich „jegliche Akzentuierungen im audiovisuellen Wahrnehmungsstrom – Vokalisierung, Geräusche und musikalische Akzente ebenso wie Schnitte, Bewegungsrhythmen von Figuren oder Objekten im Bild, Licht- und Farbakzente, usw.“ (Bakels 2016, 100). Im Prinzip können also alle Elemente des audiovisuellen Bildes als Noten dazu genutzt werden, um in einer Synchronizität mit Elementen der Musik eine Harmonie zu bilden, die eine Bedeutung generiert, die über die Bedeutung der einzel-

nen Elemente hinausgeht. Auf der visuellen Ebene kann dies ein Schnitt oder eine Blende sein, der/die gleichzeitig zum Wechsel der Musik erfolgt. Auch eine Bewegung des Bildes bspw. durch eine Kamerafahrt oder einen Zoom kann als Note oder Akkord im weiteren Sinne dienen und in Synchronizität bspw. mit einer auf- oder absteigenden Tonfolge in der Musik eine harmonische Verbindung eingehen. Weitere Noten der visuellen Ebene sind: Licht-, Schatten- oder Farbdynamiken; Bewegungen im Bild durch Figuren, Objekte oder Explosionen; Einblendungen von Schrift bspw. bei Credits an Filmanfang oder -ende. Doch auch auf akustischer Ebene, die nicht direkt zur Musik zugeordnet werden kann, können einzelne Töne als Noten zur *audiovisuellen Harmonie* beitragen. Dazu zählen kurz gesagt alle Töne, seien es Effekte wie eine zufallende Tür oder auch das gesprochene Wort und simple Ausrufe wie Interjektionen. Wie so oft nutzt Chion zwar musikalische Begriffe, aber keine musikalischen Beispiele. Stattdessen schreibt er über das Einschlagsgeräusch in Synchronizität mit dem Bild des Einschlags als obligatorisches Beispiel aus Film und Fernsehen (Chion 1994, 60), was er dann als Pendant zum Akkord in der Musik begreift (ebd. 62). Doch wie können diese *audiovisuellen Harmonien* Bedeutung erlangen?

### 3.1.4 Bedeutungserlangung

Chion selbst beschreibt die Bedeutungserlangung von synchron auftretenden Bild- und Tonereignissen mit den Begriffen *added-value* und *synchresis*. Mit *added value* meint Chion, dass Bild und Ton sich vor allem informativ gegenseitig aufwerten können und zwar so, dass man als Zuschauer das Gefühl hat, die Bedeutung oder Information eines (tonalen oder visuellen) Ereignisses komme natürlicherweise nur von dem Ereignis alleine und nicht von der Kombination der Bild- und Tonebene (Chion 1994, 5). Ein typisches Beispiel dafür wäre das Geräusch eines nicht sichtbaren aber diegetischen Objektes, einer Figur oder einer Handlung in einer Filmszene bzw. -einstellung. All dies trifft auf eine Szene in Quentin Tarantinos Film *INGLORIOUS BASTARDS* (USA/Deutschland 2009, Quentin Tarantino) zu, wenn die sogenannten Bastards, eine fiktive US-amerikanische Einheit im von Deutschland besetzten Frankreich während des zweiten Weltkriegs, einen gefangenen deutschen Wehrmachtsoldaten nach einer feindlichen Stellung befragen. Als Dieser sich jedoch weigert und dabei auf sein Herz zeigt, ertönt zwei Mal ein Schlaggeräusch mit Hall (0:30:38, Abb. 1.1), während das Bild zu dem Anführer der Bastards schneidet, der hinter sich zeigt und sagt das Geräusch komme aus dem dunklen Gang von Sgt. Donny Donowitz (0:30:47), der gerne „Nazis“ mit seinem Baseballschläger verprügelt. Im folgenden nutzt die Szene das Prinzip des *added value* noch ausführlicher und man sieht (und hört) eine Sequenz unterlegt von Westernfilmmusik, in der das Schlaggeräusch des Baseballschlägers einen steigenden Rhythmus

produziert, der zusammen mit der Musik in den Höhepunkt mündet als Sgt. Donowitz mit seinem Baseballschläger aus dem dunklen Gang hervorkommt (0:32:45, Abb. 1.2). Der Ton des lange nicht zu sehenden Baseballschlägers bereichert das Bild hier maßgeblich und trägt enorm zur Spannung der Szene bei, die sich in dieser Stärke nicht ohne das Prinzip des *added value* erklären lässt. Wichtig ist aber die Synchronizität von Bild und Ton, bei der das Prinzip des *added value* besonders hervortritt und zwar durch das Prinzip der *Synchrese* (Chion 1994, 5).



Abbildung 1: 1 Soldat zeigt auf sein Herz; 2 Sgt. Donowitz erscheint aus dunklem Gang

Das von Chion aufgestellte Kofferwort *Synchrese* setzt sich zusammen aus Synchronizität und Synthese und meint „the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time“ (ebd. 63). In einem bestimmten Moment verbinden sich also Bild- und Tonereignis zu einer unvermeidlichen und bedeutungsträchtigen Symbiose (ebd. 63), wodurch eine „gemeinsame[n] Wahrnehmungsqualität [entsteht], die über die Qualitäten der einzelnen modalen Eindrücke hinausgeht“ (Bakels 2016, 36 kursiv i.O.). Wie vielleicht kein anderer Film nutzt *THE MATRIX* (USA 1999, Laurence/Andrew Paul Wachowski) das Prinzip der *Synchrese* fast schon dauerhaft und zwar durch eine eng mit dem Bild verwobene Musikebene, die durch „eine lose Abfolge kurzer musikalischer Motive“ (Bakels 2016, 97) gekennzeichnet ist. Die musikalische Untermalung einer Szene besteht also nicht aus einem als Ganzes komponierten Musikstück, sondern aus eng ans Bild verbundenen einzelnen Motiven. Die Musikkomposition von Szenen und des gesamten Films ist somit zu einem erheblichen Teil von der Bildfolge bestimmt. Ein gutes Beispiel hierfür liefert die Szene am Ende des Films als der Held Neo scheinbar vom Antagonisten Agent Smith besiegt und getötet wird. Nachdem Neo aber in der echten Welt von Trinity geküsst wird, fängt sein Herz wenige Sekunden später wieder an zu schlagen und man sieht den Herzfrequenzmonitor, der dies visuell vermittelt (2:04:49, Abb. 2.1). Gleichzeitig ertönt in der Musik, die als klassische Hollywoodfilmmusik kategorisiert werden kann, zum ersten Mal in der Szene ein wenige Sekunden andauernder aufsteigender Chor und man sieht folgend wie Neo in der Matrix seine Augen öffnet. Die direkt folgende Einstellung des aufstehenden Neo wird auch mit leicht aufsteigendem Chor untermalt (2:04:59). Nachdem Neo die Pistolenkugeln seiner Gegner scheinbar telekinetisch aufhält, lässt

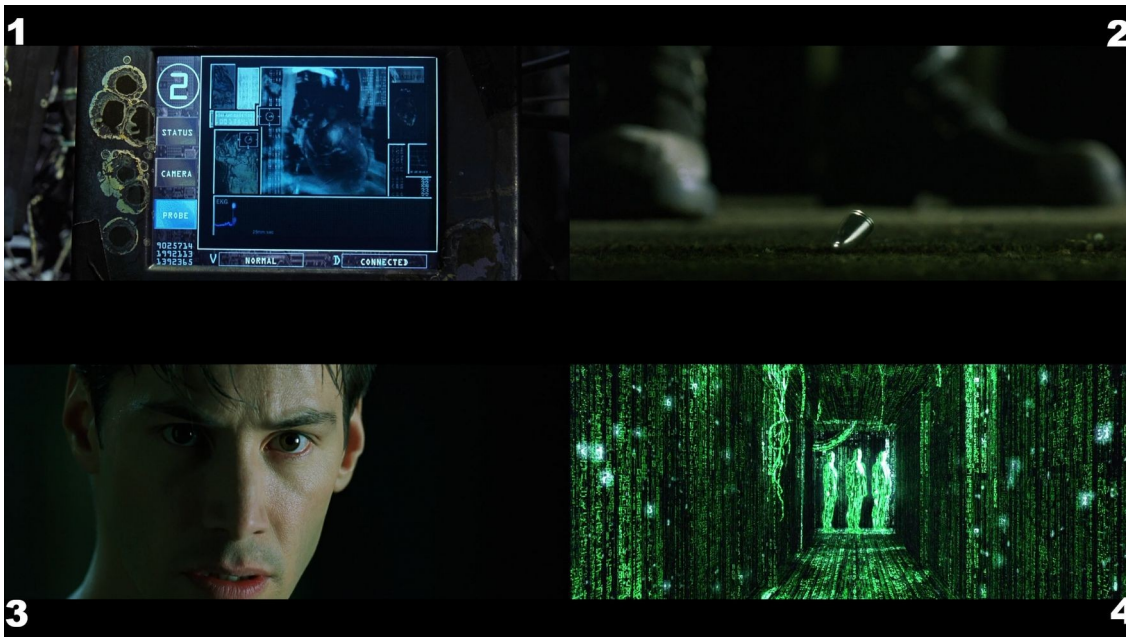


Abbildung 2: 1 Herzschlag Neos; 2 Pistolenkugel; 3 aufblickender Neo; 4 Neos Blick auf die Matrix

er sie in der Luft stehen. Es folgt ein Schnitt auf die erste fallende und auf den Boden aufkommende Pistolenkugel, mit der auf musikalischer Ebene erneut der aufsteigende Chor einsetzt (2:05:27, Abb. 2.2). Wenige Sekunden später nach dem die Helferfigur Morpheus Neo als den Auserwählten betitelt und die Geschehnisse narrativ erklärt, wechselt das Bild zum aufblickenden Neo und in der Musik ertönt erneut der immer lauter werdende Chor (2:05:35, Abb. 2.3). Nur zwei Sekunden später wechselt das Bild erneut und wir sehen Neos Perspektive, wie er die Matrixwelt nun codiert wahrnimmt, synchron dazu ertönt ein letztes Mal und nun in voller Lautstärke der Chor (2:05:37, Abb. 2.4). Die Besonderheit dieser Symbiose von Bild und Ton bzw. Musik liegt in den spezifischen Momenten in denen sie harmonisch zusammentreten und das gesamte audiovisuelle Bild durch eine Musikalität geprägt ist, die Bedeutung erlangt. Diese Momente der *audiovisuellen Harmonie* setzen sich im beschriebenen Beispiel auf bildlicher Ebene aus Schnitten und Bewegungen im Bild und auf Tonebene aus dem Chor zusammen. Die Funktion der Chors ist hier direkt mit dem jeweils Gezeigten durch das Prinzip der *Synchrese* symbiotisch verbunden. In der Synchronizität fallen Bild und Ton zusammen und bilden eine Synthese. Der Chor und die dazu gezeigten Bilder begleiten nicht nur eine Erkenntnisgewinnung sowohl der Figuren im Film als auch des Zuschauers, sondern vermitteln diese Erkenntnis gemeinsam: Neo ist durch die Liebe mit Trinity nicht nur erwacht, sondern hat nun endgültig den Status des Auserwählten erreicht, der sich dem zuvor übermächtigen Agent Smith stellen kann, die Computerwelt der Matrix entschlüsselt und so letztendlich die Menschheit retten wird. Jeder bedeutungsvolle Moment dieser Entwicklung ist durch den Chor in synchroner Verbindung mit einem Ereignis im bzw. einer Veränderung des Filmbild(s) verbunden.

*Synchrese* schafft also einen gemeinsamen aus Bild und Ton in einem spezifischen Moment entstehenden Wahrnehmungszusammenhang, der eine Bedeutung generiert, die über die Summe der einzelnen Modalitäten hinausgeht. Allerdings kann *Synchrese* auch rein zufällig entstehen, wie Chion selbst gesteht: „Play a stream of random audio and visual events, and you will find that certain ones will come together through synchresis and other combinations will not“ (Chion 1994, 63). Das liegt daran, dass wir Menschen immer versuchen unsere Wahrnehmung zu ordnen und zu strukturieren, denn „total disorder with no apparent goal is intolerable for human beings“ (ebd. 211-212). Unsere Wahrnehmung ist demnach auch durch eine Erwartungshaltung geprägt, die auch in der Rezeption von Medien gegeben ist, und wie nun gezeigt werden soll, von Filmemachern genutzt wird.

### **3.1.5 Kadenzen, Antizipation und Erfüllung**

Unser Musikgenuss ist zu einem Großteil vom Erkennen von Mustern geprägt, die wir daraufhin antizipieren können, und deren Erfüllung uns Freude bereiten kann. Rhythmen sind hierfür das einfachste Beispiel, denn das Erkennen eines rhythmischen Musters erfolgt zumeist bereits nach der ersten Wiederholung. Ein sich zu häufig und dauerhaft ändernder Rhythmus würde uns den Genuss am Hören von Musik nehmen, denn wir können die Veränderungen nicht voraussehen und die Musik hätte keine erkennbare Struktur mehr. Nicht nur Rhythmen sind wichtig, um die musikalische Form zu verstehen, sondern auch Kadenzen. Unter dem Begriff Kadenz versteht man in der Musik eine „Akkordfolge als Abschluss oder Gliederung eines Musikstücks“ (Duden.de 2024b). Auch Kadenzen sind Muster, die der Hörer erkennen kann und soll, sodass dieser den Aufbau der Musik versteht und bspw. einen Wechsel zu einem weiteren Motiv erwarten kann. Chion hat bereits herausgestellt, dass auch audiovisuelle Sequenzen vom Zuschauer permanent bewusst oder unbewusst auf Muster und deren Erfüllung hin überprüft werden (vgl. Chion 1994, 56). Diese Kadenzen in audiovisuellen Sequenzen setzen sich aus ähnlichen Elementen zusammen, die hier bereits in Kapitel 3.1.3 als Noten für *audiovisuelle Harmonie* herausgestellt wurden, also bspw. bestimmte Bewegungen der Kamera oder auch Veränderungen im Schauspiel wie Mimik und Gestik. Was dann auf diese Kadenzen hin folgt, erfüllt entweder die zuvor aufgebauten Erwartungen des Zuschauers oder er wird überrascht (ebd. 55).

Eine der letzten Szenen in Stanley Kubrick's Antikriegsfilm *FULL METAL JACKET* (UK/USA 1987, Stanley Kubrick) dient hier als gutes Beispiel. Nachdem eine US-amerikanische Truppe im Vietnamkrieg eine Scharfschützin des Vietkong erschießt, liegt diese vor den Augen der Soldaten im Sterben, fängt an zu beten und bittet schließlich um den Gnadenschuss. Der subversive Protagonist, im Film Privat Joker genannt, will der Scharfschützin die Qualen eines

langsamen Todes ersparen und erschießt diese letztendlich auch. Wichtig in dieser Szene ist der innere Kampf des Protagonisten direkt vor dem Schuss, denn dieser hat zuvor noch nie einen Menschen getötet und nimmt im Film eine gegenüber Gewalt und Krieg äußerst kritische Rolle ein. Die musikalische Untermalung in dieser Szene ist eng an diesen inneren Kampf des Protagonisten geknüpft und begleitet diesen. Sie setzt sich hauptsächlich aus Streichern<sup>4</sup> und verschiedenen dumpfen Synthesizertönen zusammen, die oft sehr langgezogen werden. Während das Bild zwischen verschiedenen Aufnahmen von Charakteren wie einigen Soldaten, der Scharfschützin und natürlich des Protagonisten wechselt, um dann anschließend auf diesem in Nahaufnahme zu verharren, werden die musikalischen Töne immer intensiver, sowohl in Lautstärke als auch Tempo und Tonhöhe. Dadurch wird eine sich immer weiter steigende Kadenz aufgebaut, die auf bildlicher Ebene synchron durch die ernster und angespannter werdende Mimik des Protagonisten erweitert wird (1:50:12, Abb. 3.1). Bild und Musik gehen hier eine über mehrere Sekunden andauernde steigende Kadenz und damit auch *audiovisuelle Harmonie* ein, die ihren gemeinsamen Abschluss findet, wenn synchron zum Schussgeräusch sich die angespannte Mimik des Protagonisten auflöst und die Musik verstummt (1:50:15, Abb. 3.2). Der Zuschauer kann diese Kadenz, sowohl musikalisch als auch bildlich leicht erkennen, und antizipiert die Entscheidung bzw. den Schuss des Protagonisten, weiß aber beim ersten Sehen der Sequenz nicht ob, wie und wann die Kadenz auch aufgelöst wird, also ob der Protagonist wirklich schießen wird. Ohne diese harmonische Kadenz, zusammengesetzt aus Bild und Musik, würde diese Sequenz nicht annähernd so eine starke emotionale Wirkung auf den Zuschauer haben.



Abbildung 3: 1 angespannte Mimik vor Schuss; 2 lockerere Mimik nach Schuss

Ein weiteres Beispiel für eine Kadenz im Film, die sowohl musikalisch als auch bildlich aufgebaut wird, findet sich in *PULP FICTION* (USA 1994, Quentin Tarantino), nachdem die beiden rivalisierenden Charaktere Butch und Wallace von zwei BDSM-Anhängern übermannt und in einem Keller gefesselt werden. Wallace wird als erstes Opfer ausgewählt und von den beiden

<sup>4</sup> Streng genommen besteht die gesamte Musik hier aus Synthesizertönen, die teilweise klassische Instrumente wie eben Streicher nachahmen.

BDSMlern in einen Hinterraum gebracht. Bildlich beginnt die Kadenz mit einer Einstellung in Zeitlupe wie der letzte BDSMler durch die Tür geht (1:42:06, Abb. 4.1), sich zur Kamera dreht (1:42:14, Abb. 4.2) und die Tür zuwirft (1:42:16, Abb. 4.3). Synchron zum Beginn der Einstellung (Abb. 4.1) startet das Lied *COMANCHE* (USA 1961, *The Revels*) mit einer Saxophonmelodie, die vier Mal wiederholt wird und sich bei jeder Wiederholung durch eine höhere Tonlage zu steigern scheint. Synchron zum Zuschlagen der Tür, die wieder in normaler Geschwindigkeit gezeigt wird, gibt es zunächst den Toneffekt der zufallenden Tür und auf musikalischer Ebene das Einsetzen von Gitarren und Schlagzeug sowie eine im Vordergrund stehende, wilde und sehr tiefe Saxophon Melodie, die das Treiben im Hinterraum suggeriert, während das Bild zum im ersten Raum weiter gefesselten Butch wechselt (1:42:20, Abb. 4.4), der sich folgend nicht seinem Schicksal, angedeutet durch die Bild-Musik-Kadenz, beugen will. Die Zeitlupe auf Bildebene und die sich steigernde Wiederholung der ersten Saxophonmelodie auf Tonebene bilden zusammen eine Kadenz, die der Zuschauer antizipieren kann und deren Teilerfüllung (auf akustischer Ebene) letztendlich eine humoristische Note bekommt, da das eigentliche Treiben der Charaktere, auf das die Musik hindeutet, eben nicht gezeigt wird.



Abbildung 4: 1 BDSMler geht durch Tür; 2 BDSMler dreht sich zur Kamera; 3 Tür ist zugeworfen; 4 Butch weiterhin gefesselt

Ob eine Kadenz und die Antizipation des Zuschauers in audiovisuellen Sequenzen wirklich erfüllt werden sollte oder nicht, kann laut Chion nicht abschließend geklärt werden: „It is often more interesting when the expectation is subverted. And, at other times, when everything ensues as anticipated, the sweetness and perfection of the rendering of the anticipation itself are sufficient to move us“ (Chion 1994, 56). Das Beispiel aus *FULL METAL*

JACKET bewegt den Zuschauer aber gerade eben wegen der Erfüllung der Antizipation<sup>5</sup>. Die Szene aus PULP FICTION hingegen erfüllt die Antizipation des Zuschauers nur auf musikalischer Ebene, erzeugt so aber einen gewissen Witz. Wie noch gezeigt wird, geht es bei Computerspielen allerdings um die perfekte Erfüllung von Kadenzen durch *ludoaudiovisuelle Harmonien*.

### 3.1.6 Definition *audiovisuelle Harmonie*

Ausgehend von den Überlegungen Michel Chions in seinem Buch Audio-Vision und in Anlehnung an den Begriff des *audiovisuellen Rhythmus* von Jan-Hendrik Bakels definiere ich den Begriff *audiovisuelle Harmonie* wie folgt:

*Audiovisuelle Harmonie* beschreibt Momente in audiovisuellen Medien, bei denen Elemente der Bild-und-Ton-Darstellung synchron zu Tönen oder Akkorden in der Musik vorkommen, wodurch die Harmonie der Musik auf Ebenen außerhalb der Musik erweitert wird. Musik und audiovisuelle Darstellung verschmelzen in diesen Synchronisationspunkten, die sich auch über wenige Sekunden strecken können, harmonisch miteinander, erlangen dadurch eine erhöhte Bedeutung und prägen die Rezeption. Eine *audiovisuelle Harmonie* kann sich aus unterschiedlichen Elementen mehrerer Ebenen zusammensetzen, die ich in Analogie zur Musik als Noten bezeichne, und in verschiedensten Kombinationen auftreten. Auf Seiten der Musik können diese Elemente bestimmte Einzeltöne aber auch Akkorde oder Tonfolgen und -leitern ebenso wie Gesang sein. Bei der visuellen Darstellung können eine bestimmte Bewegung von Objekten und Figuren im Bild oder auch die Bewegung der Kamera also des gesamten Bildes selbst wie Schwenken, Zoom oder Kamerafahrten als Noten für *audiovisuelle Harmonie* genutzt werden. Auch Schnitte, Farbwechsel sowie Licht- und Schattendynamiken können auf visueller Ebene als solche Noten fungieren und in Synchronizität mit Elementen aus der Musik treten. Auf der akustischen Ebene kann jeder einzelne Ton außerhalb der Musik in Synchronisation mit Elementen der Musik verbunden werden und so *audiovisuelle Harmonie* mit erzeugen. *Audiovisuelle Harmonie* tritt in allen audiovisuellen Medien auf, sei es Film, Fernsehen oder Musikvideos und wird genutzt, um bestimmten Momenten durch die harmonische Kombination mehrerer Modalitäten mehr Bedeutung zu geben und den Rezipienten so durch mehrere sensorische Reize gleichzeitig zu affizieren.

---

5 In LA HAINE (Frankreich 1995, Mathieu Kassovitz) gibt es eine ähnliche Kadenz, die allerdings über das lauter werdende Ticken einer Uhr und einem Zoom auf ein Gesicht funktioniert. Mit dem Schussgeräusch wird das Bild schwarz und der Zuschauer erfährt nicht wer von den beiden Charakteren, die sich mit einer Pistole bedrohen, erschossen wird.

### 3.2 Audiovisuelle Harmonie im Film

Eines der Fallbeispiele von Bakels für die Etablierung seines Begriffs des *audiovisuellen Rhythmus* ist der Prolog des Films *THE ROYAL TENENBAUMS* (USA 2001, Wes Anderson). In einer ausführlichen Analyse zeigt Bakels wie die gesamte Sequenz an eine spezielle, instrumentale und verlängerte Version des Liedes *HEY JUDE* (UK 1968, The Beatles) gebunden ist. Das Lied ist dem Bild nicht unterlegt oder begleitet dieses einfach, stattdessen stellt Bakels heraus wie sich seine musikalische Dynamik „mit Blick auf visuelle und auditive Aspekte des bewegten Bilds als audiovisuell verwoben, als ‚eingefasst‘ in einer Bild, Musik und Dialog gleichermaßen umspannenden, gemeinsamen, audiovisuellen kompositorischen Figur“ erweist (Bakels 2016, 86). Entscheidend hierfür sind die wiederkehrenden Wechsel von Strophe und Refrain, die der Vorstellung der Charaktere in Verbindung mit Schnitten und Kamerabewegungen eine Art rhythmische Struktur verleihen (ebd. 87). So interessant und richtig Bakels Analyse und die Erkenntnis eines *audiovisuellen Rhythmus* der Szene ist, so erstaunlicher ist es, dass der entscheidende Moment der Szene bei Bakels zwar Beachtung findet, aber in der Analyse selbst zu kurz kommt. Der gesamte von Bakels ausführlichst beschriebene *audiovisuelle Rhythmus* mündet nämlich in einer *audiovisuellen Harmonie* gegen Ende der Sequenz, wenn Prolog und Musik gemeinsam ihren Höhepunkt finden. Die Vorstellung der Charaktere wird abgeschlossen durch die Erzählstimme: „In fact, virtually all memory of the brilliance of the young Tenenbaums had been erased by two decades of betrayal, failure and disaster.“ (0:06:15). Direkt darauf folgt der Ausruf des Sohns Richie „Go Mordecai!“, wenn er seinen Adler vom Dach des Familienhauses aufsteigen lässt. Genau dieses Aufsteigen ist der Moment indem die zuvor rein instrumentale Version des Liedes durch seine berühmte Coda, gesungen von einem Chor, ihren Höhepunkt findet (0:06:25, Abb. 5.1). Synchron zum Einsetzen der Coda gibt es einen Schnitt hin zum nun fliegenden Adler über der Stadt (0:06:28, Abb. 5.2), die in zwei Einstellungen mehrere Sekunden anhält.



Abbildung 5: 1 Adler wird losgelassen; 2 Adler fliegt über die Stadt

Ohne diese *audiovisuelle Harmonie*, bestehend aus der Coda, die die Musik befreit und sie endlich mit Gesang bereichert, und dem Schnitt zum fliegen-

den Adler, würde der mehrere Minuten andauernde *audiovisuelle Rhythmus* nicht seinen wie Bakels sagt „ästhetischen Genuss“ erfahren, der in „einer *in Freude gegossenen Melancholie*“ liegt (Bakels 2016, 88, kursiv i.O.). Insofern muss Bakels zumindest dafür kritisiert werden, dass er der Synchronizität bei der Schaffung eines *audiovisuellen Rhythmus* keine Bedeutung zugesteht (ebd. 102).

Mit einem letzten Beispiel für das Filmmedium möchte ich noch zeigen, dass auch Gesang mit sinnhaftem Text als Note für *audiovisuelle Harmonie* genutzt werden kann, die die narrative Bedeutung einer Szene verstärkt. *CRUEL INTENTIONS* (USA 1999, Roger Kumble) popmusikalischer Soundtrack eignet sich hier sehr gut und nutzt das Phänomen der *audiovisuellen Harmonie* vor allem in der letzten Szene.

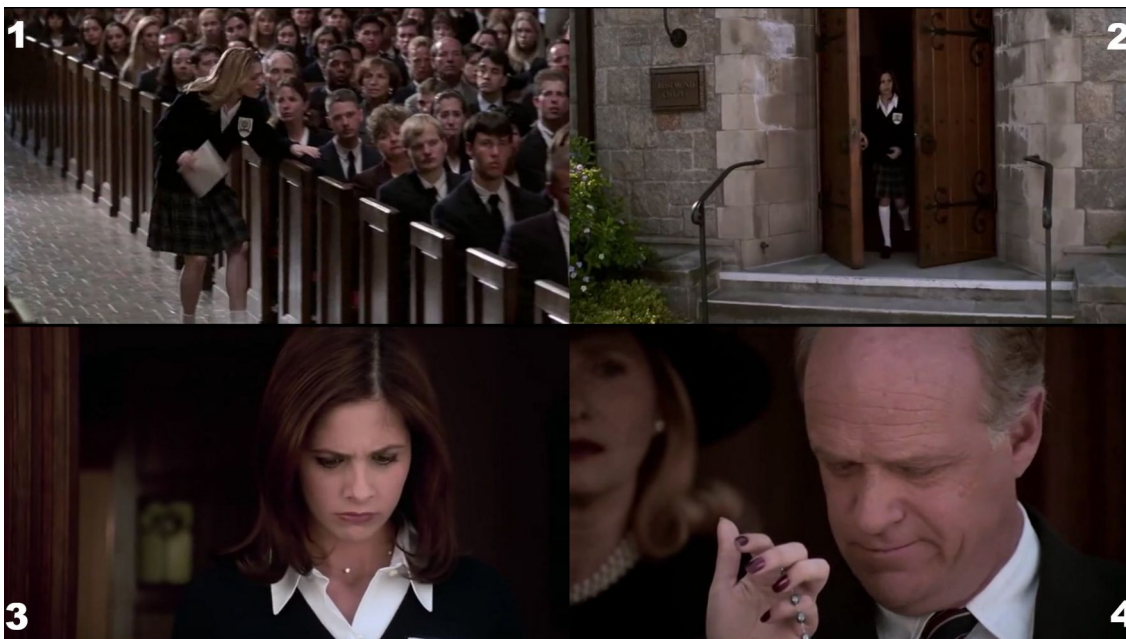


Abbildung 6: 1 Schülerin verteilt Heft; 2 Kathryn verlässt Kirche; 3 Kathryn liest aus Heft; 4 Schuldirektor findet Kokain

Zunächst sehen wir wie die Antagonistin Kathryn in einer Kirche vor der versammelten Schulgemeinschaft eine Trauerrede über ihren kürzlich verstorbenen Stiefbruder Sebastian hält, der im Film als Protagonist agiert. Während der Rede betritt jedoch eine Schülerin mit einem Heft in der Hand die Kirche und flüstert andere Mitschüler an, die daraufhin nach und nach die Kirche verlassen, was Kathryn sauer aufstößt. Mit dem Auftritt der Schülerin (1:30:44, Abb. 6.1) beginnt das Lied *BITTERSWEET SYMPHONY* (UK 1997, The Verve) mit flachen, sehr langgezogenen Streichern. Zunächst sehr leise und kaum auszumachen, erhöht sich die Lautstärke stetig parallel zum Verlassen von mehr und mehr Schülern. Die berühmte Streichermelodie des Liedes, die auf einem Sample des Liedes *THE LAST TIME* (USA 1965, The Rolling Stones) beruht (Musikexpress.de 2019), ist erst dann auszumachen, wenn sich bereits einige Schüler auf den Weg nach draußen machen

und auch Kathryn bemerkt, dass etwas nicht stimmt (1:31:08). Erst mit dem Schnitt zu Kathryn wie sie eine Tür öffnet und die Kirche verlässt, setzt synchron dazu das Schlagzeug ein (1:31:46, Abb. 6.2) und die Musik hat ihre volle Lautstärke erreicht, womit auch die zuvor musikalisch durch die Erhöhung der Lautstärke und bildlich durch eine Kamerafahrt parallel zu Kathryns Gang durch die Kirche aufgebaute Kadenz ihre Erfüllung in einer *audiovisuellen Harmonie* erlangt. Es folgen einige Einstellungen, die zeigen wie Dutzende Schüler vor der Kirche in einem mehrfach verteilten Heft lesen, das Kathryn schließlich selbst in die Hand gedrückt bekommt. Synchron zum Schnitt auf Kathryns Gesicht, bei gleichzeitigem langsamen Heranzoomen, setzt nun der Refrain des Liedes ein: „’Cause it’s a bittersweet symphony, that’s life“ (1:32:08, Abb. 6.3). Der Gesangstext ist hier direkt mit der Handlung des Films verbunden und kommentiert ironisch das nun anstehende Ende der hinterlistigen Machenschaften Kathryns. Nach dieser *audiovisuellen Harmonie* folgen Aufnahmen einiger entsetzter Schüler, Kathryns Gesicht in Nahaufnahme und Detailaufnahmen aus dem Heft selbst, das eine Collage aus Tagebucheinträgen und Fotos der intriganten Antagonistin ist. Die letzte *audiovisuelle Harmonie* setzt sich dann wenig später aus einer weiteren Textstelle des Liedes „no, no ,no“ und dem synchron dazu gezeigten Kopfschütteln des Schuldirektors zusammen (1:37:26, Abb. 6.4), der in Kathryns Kruzifix-Halskette Kokain vorfindet, genau wie es zuvor in den Detailaufnahmen des Heftes zu sehen war. Obwohl in Filmen die Sprache die Tonebene bestimmt (Chion 1994, 5-6), gilt dies in der Regel nicht für Sprache in der Musik. Dem Gesangstext wird hier in *CRUEL INTENTIONS* aber durch *audiovisuelle Harmonie* eine erhöhte Bedeutung zugestanden, denn Musik und Handlung sind so eng miteinander verbunden, wodurch die Wirkungsweise der Szene deutlich verstärkt wird, denn das Ende der Antagonistin, das der Zuschauer über den ganzen Film hin erwartet, wird hier sowohl bildlich als auch musikalisch und vor allem in Kombination der beiden Modalitäten vermittelt.

### 3.3 Audiovisuelle Harmonie im Fernsehen

Zwar spricht Chion davon, dass die Beziehung zwischen Bild und Ton im Fernsehmedium nicht von Harmonie, sondern von Kontrapunkt<sup>6</sup> geprägt ist (Chion 1994, 37), so gibt es trotzdem viele Fernsehformate, vorproduziert ebenso wie live, in denen *audiovisuelle Harmonie*, also die echtzeitliche Verbindung von Bild und Ton mit Musik, genutzt wird.

Das Intro der Zeichentrickserie *THE SIMPSONS* (USA 1989, Matt Groening) weist gleich mehrere *audiovisuelle Harmonien* auf. Direkt zu Anfang wird

6 Chion versteht hier unter Kontrapunkt, dass das Bild und der Ton eigentlich nicht zusammen passen und wählt das Beispiel einer Radrennübertragung, bei der bildlich das Rennen zu sehen ist, auf der Tonebene aber zwei Kommentatoren zu hören sind (Chion 1994, 37).

zum Erscheinen des Schriftzugs „The Simpsons“ in den Wolken, dieser auch von einem Chor gesungen (0:02, Abb. 7.1). Wenig später sieht man Bart, den Sohn der Familie, wie er in der Schule eine Strafe absolviert. Nachdem Klingeln rennt Bart aus dem Klassenraum durch die Schule und springt schließlich auf einem Skateboard aus dem Haupteingang, was alles in einer Kameraeinstellung mit einer wilden Kamerafahrt<sup>7</sup> gezeigt wird. Synchron dazu beginnen ein Xylophon und Streicher in schnellem Tempo die Tonleiter rauf und wieder runter zuspiesen (0:17). Das Aufkommen Bards mit seinem Skateboard auf dem Gehweg wird synchron durch einen Beckenschlag markiert (0:20, Abb. 7.2). Aus Platzgründen wird hier nicht auf noch weitere Beispiele aus dem Intro der Serie eingegangen, aber es sollte erkennbar sein, dass die *audiovisuellen Harmonien* hier eine humoristische Funktion erfüllen, indem sie das Gezeigte mit musikalischen Motiven unterstützen, was in gewisser Hinsicht an den Effekt des *Mickey Mousing*<sup>8</sup> erinnert.

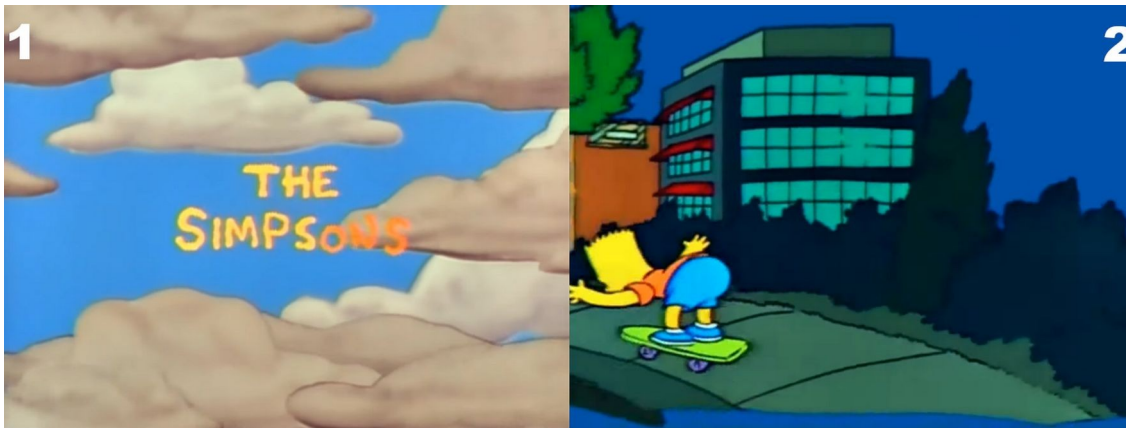


Abbildung 7: 1 Schriftzug; 2 Bard kommt mit Skateboard auf Gehweg auf

Selbst Live-Sendungen wie *WER WIRD MILLIONÄR?* (Deutschland 1999, Matthias Schwab) nutzen das Prinzip *audiovisueller Harmonie*. So gibt es direkt vor jeder neuen Frage eine fest koordinierte Kombination von Bild und Musik<sup>9</sup>. Mit dem Schnitt setzt zunächst die laut aufspielende Musik ein, während synchron dazu die Kamera von einem Scheinwerfer an der Decke des Studios nach hinten und gleichzeitig auf Höhe des Kandidaten und Moderators fährt. Diese *audiovisuelle Harmonie* bestehend aus dramatischer Musik, Schnitt im Bild und schneller Kamerafahrt schärft die Konzentration des Zuschauers nach den zuvor meist lockeren und lustigen Gesprächen bzw. Mo-

7 Natürlich wird die Kamerafahrt in Zeichentrick und Animation nur durch geschicktes Zeichnen und Animieren suggeriert.

8 „Synchronisierung oder Koordination von Filmmusik mit der im Film gezeigten Handlung, so dass diese unterstrichen bzw. herausgestellt wird. Mickey Mousing wurde und wird gern bei Zeichentrickfilmen angewandt, woher sich auch die Bezeichnung erklärt, die den Mickey-Mouse-Filmen von Walt Disney entlehnt ist. Die Bezeichnung stammt wohl von David O. Selznick, der damit eine Musik von Max Steiner, die sich allzu sehr an die Bewegung der Akteure anlehnte, verurteilte.“ (Kaczmarek/zu Hünigen 2022).

9 Leider unterbindet der TV-Sender RTL jegliche Aufzeichnungen seiner Mediathekinhalte am PC, weswegen hier keine Bilder gezeigt werden können.

nologen des Moderators der Sendung. Die synchrone Nutzung von Bild und Musik verändert innerhalb weniger Sekunden die Atmosphäre der Sendung und versetzt den Zuschauer wieder in eine angespannte Erwartungshaltung.

### 3.4 Audiovisuelle Harmonie in Musikvideos

Wie kein anderes Medienformat scheint das Musikvideo prädestiniert für eine enge Verbindung von Bild und Musik zu sein, schließlich steht die Musik hier im Vordergrund. Chion vergleicht das Musikvideo sogar mit dem Stummfilm und stellt fest, dass in beiden Medien, das Bild vom Ton befreit ist, da es keinen oder nur sehr wenig Dialog zwischen Charakteren gibt: „[...] the music video's image is fully liberated from the linearity normally imposed by sound“ (Chion 1994, 167). Allerdings ist die Produktionsweise genau entgegengesetzt, denn im Stummfilm kommt die Musik erst im Schnitt oder sogar erst bei der Vorführung hinzu, während im Musikvideo die Musik als Ausgangslage für Konzeption, Dreh und Schnitt der Bilder dient. Das Bild ist also von der Linearität, die im Film von Sprache bzw. Dialog ausgeht, befreit und kann somit im Musikvideo entweder sehr frei gestaltet oder aber sehr eng an die Musik gebunden werden. Mit zwei unterschiedlichen Beispielen werde ich nun zeigen, dass die enge Verbindung von Bild und Musik im Musikvideo ein bedeutungsvolles Phänomen ist.

Das Musikvideo zu Rammsteins Depeche Mode Cover *STRIPPED* (Deutschland 1998, Philipp Stölzl) ist zunächst besonders, da es auf Bildebene keinerlei eigens produzierte Bilder nutzt, sondern stattdessen Bilder aus Leni Riefenstahls zweiteiligem *OLYMPIA* (Deutschland 1938, Leni Riefenstahl) verwendet, die sich an zahlreichen Synchronisationspunkten mit der Harmonie der Musik verbinden. Nach einem längeren, ruhigeren Intro setzen ab 0:43 gleichzeitig E-Gitarre und Schlagzeug synchron zum Schnitt auf einen in Silhouette gezeigten Diskuswurf ein (Abb. 8.1). Diese *audiovisuelle Harmonie* wird dann in ähnlicher Form drei Mal wiederholt: Speerwurf mit einer Armbewegung von links nach rechts (0:47), Gymnastikbewegung mit Reifen von rechts nach links (0:51, Abb. 8.2), Speerwurf mit Armbewegung von rechts nach links (1:02). Wenig später setzt eine Blende von Armsilhouetten hin zum olympischen Feuer synchron mit einem alleinstehenden Gitarrenriff ein (1:20, Abb. 8.3). Bei 2:04 gibt es dann nach dem ersten Refrain einen Schnitt hin zu Gymnastinnen, der synchron zum Start des Synthesizers stattfindet. Bis 2:07 sieht man folgend verschiedene Einstellungen der Gymnastinnen, die ihre Gymnastikkeulen in schwingenden Bewegungen von einer zur anderen Seite bewegen. Diese Bewegungen sind synchron zu den Synthesizertönen, die auch am besten als schwingend bezeichnet werden können und mehrfach im Lied vorkommen. Diese *audiovisuelle Harmonie* ist außerdem ein gutes Beispiel dafür, dass die Symbiose aus Bild und Ton auch zufällig sein kann (Chion 1994, 63), denn der schwingende Synthesizerton

scheint im Video häufig mehr oder weniger synchron mit Bewegungen der unterschiedlichsten Sportler aus *OLYMPIA* zu sein. Besonders beeindruckend sind die *audiovisuellen Harmonien* gegen Ende des Videos, wenn der angesprochene Schwington in Verbindung mit den Aufnahmen der Turmspringer gebracht wird. So sind die ersten vier Sprünge (3:00, Abb. 8.4, 3:03, 3:05, 3:08), erneut als Silhouette gefilmt, synchron zum Riff der E-Gitarre, dem Becken des Schlagzeugs und eben dem angesprochenen schwingenden Synthesizerton. Auch in den folgenden Sekunden scheinen besonders die schwingenden Synthesizertöne immer wieder in Symbiose mit den vorwärts wie rückwärts abgespielten Turmsprüngen<sup>10</sup> zu stehen.



Abbildung 8: 1 Diskuswurf; 2 Gymnastin mit Reifen; 3 Blende Gymnastinnen zu olympischem Feuer; 4 Turmspringer

Die vielen *audiovisuellen Harmonien* in *STRIPPED* erzeugen eine einmalige ästhetische Verbindung zwischen Leni Riefenstahls Propagandafilm und dem rockigen Cover eines Pop-Songs, die Rammsteins Sänger Till Lindemann direkt gefielen: "wie das passte...die ganzen Schnitte auf den Rhythmus" (Rammstein Official 2019, 1:10). Über die letztendliche Bedeutung des Videos an sich und seiner *audiovisuellen Harmonien* kann hier nicht vertieft eingegangen werden, aber Regisseur Philipp Stölzl entschied sich bewusst für Riefenstahls *OLYMPIA* mit seiner „Filmästhetik [...], die eigentlich total viel mit Musikvideo zu tun gehabt hat [...]" (ebd. 7:04).

Das Musikvideo zu Bomfunk MC's Lied *FREESTYLER* (Finnland 1999, Miikka Lommi) ist bestückt mit zahlreichen *audiovisuellen Harmonien*, die entscheidende Funktion für die Metaebene der Handlung haben. Im Video sieht

<sup>10</sup> Die Aufnahmen sind bereits im original *OLYMPIA*-Film vorwärts wie rückwärts abgespielt. Generell übernimmt das Musikvideo auch einige Schnitte und Blenden aus dem Original, synchronisiert diese aber mit der Musik Rammsteins.

man einen Jugendlichen, der andere Menschen in seiner Nähe über die Steuerung seines MiniDisk-Players parallel zur Musikwiedergabe kontrollieren kann. Die erste *audiovisuelle Harmonie* ist zunächst allerdings nur ein Mitsingen und Gestikulieren des Jungen synchron zum Rap und Beat der Musik (0:49, Abb. 9.1), die im Laufe des Videos mehrfach wiederholt wird (1:35, 2:34, 2:40). Nachdem der Junge einem der Bandmitglieder gegenüber saß und scheinbar einen elektrischen Schock vom MiniDisk-Player bekommen hat, bemerkt er wie ein Mann in der Nähe direkt von der Steuerung seiner Musikwiedergabe kontrolliert wird und sich zunächst in der U-Bahn hinsetzt, um dann synchron zur Betätigung der Pausen-Funktion des MiniDisk-Players in seiner Bewegung einzufrieren (1:26, Abb. 9.2). Neben der *audiovisuellen Harmonie* wird dies auch durch mehrere Detailaufnahmen des MiniDisk-Players narrativ verdeutlicht.



Abbildung 9: 1 Gestikulation zur Musik; 2 Bewegung des Mannes wird eingefroren; 3 Bewegung des Breakdancers wird eingefroren; 4 Scanlines bei Vorspuleffekt

Als der Junge später an einer Gruppe Breakdancer vorbei geht sieht man wie deren Drehungen synchron zum Turntables scratching in der Musik vor und zurück abgespielt werden (2:03, Abb. 9.3). Nachdem der Junge mit seinem MiniDisk-Player auf die Gruppe zeigt, folgt eine Einstellung, die das scheinbare Vorspulen der Tanzbewegungen zeigt, was visuell mit Scanlines<sup>11</sup> im Bild und doppelter Geschwindigkeit der Videowiedergabe vermittelt wird (2:09, Abb. 9.4). Synchron dazu ertönt in der Musik ein schneller Beat, der die Schnelligkeit der Videowiedergabe ergänzt. Solche *audiovisuellen Harmonien* finden sich noch mehrfach im Verlauf des Musikvideos, bspw. beim

<sup>11</sup> Die Scanlines dienen hier natürlich auch als kulturell-bildlicher Hinweis für die Spulfunktion von Videokassettenrekordern, die zum Zeitpunkt der Veröffentlichung des Videos noch zahlreich verwendet wurden.

mehrfachen Einfrieren von Tänzerinnen synchron zum Beat der Musik (3:08, 3:11, 3:14). Gegen Ende des Videos kommt der Junge zu einem Live-Auftritt des DJs der Band und einer dazu tanzenden Gruppe von Breakdancern. Allerdings kann der Junge hier nicht mehr die Musikkwiedergabe seines MiniDisk-Players nutzen, um die Personen zu steuern (4:13). Es folgt jedoch noch eine letzte entscheidende *audiovisuelle Harmonie*, nachdem das bereits am Anfang gezeigte Bandmitglied auf den verzweifelnden Jungen zugeht. Synchron zum Beginn des Outros des Liedes, das unter anderem aus einem lang gezogenen Turntablescratch mit Echo besteht, der sich rhythmisch mehrfach wiederholt, wird das gesamte Video zurückgespult, auch hier mit einem Scanlineeffekt auf bildlicher Ebene, bis der Junge wieder am Anfangspunkt des Videos zurückgekehrt ist. Die *audiovisuellen Harmonien* im Video zu FREESTYLER sind entscheidend für die Vermittlung der Metaebene der Handlung und verbinden die Musik mit der diegetischen Welt der Narration.

Es ist nicht so, dass Musikvideos durch eine ständige Synchronisation von Bild und Musik gekennzeichnet wären. Die meiste Zeit gehen beide ihren eigenen Weg (Chion 1994, 167), aber wie gezeigt wurde, haben die Synchronisationspunkte zwischen Bild und Musik mit ihren *audiovisuellen Harmonien* bestimmte Funktionen und entscheidenden Einfluss auf die Rezeption und Interpretation der Musikvideos.

## **4 Ludoaudiovisuelle Harmonie**

### **4.1 Ludo-Ebene**

Wie bereits gezeigt wurde ist das Prinzip der *audiovisuellen Harmonie* zahlreich in Filmen, Fernsehsendungen und Musikvideos zu finden und kann dort Bedeutung generieren, entscheidende Funktionen übernehmen, zu einer besonderen Ästhetik beitragen und die Rezeption beeinflussen. Das Computerspielmedium jedoch verfügt über besondere Eigenschaften mit denen dieses Prinzip noch erweitert und weitaus komplexer gestaltet werden kann, was es letztendlich zu einer noch größeren Bedeutung für die Rezeption werden lässt. Neben den Elementen der akustischen und visuellen Ebene, die wie gezeigt in audiovisuellen Medien für Harmonien mit der Musik genutzt werden können und die ich als Noten der *audiovisuellen Harmonie* beschrieben habe, kommt bei Computerspielen noch eine zusätzliche und höchst komplexe Ebene hinzu, aus der ebenfalls Elemente wie Noten in harmonischer Verbindung mit der Musik treten können, um so *ludoaudiovisuelle Harmonien* zu bilden. Diese für das Computerspielmedium besondere Ebene fasse ich mit dem lateinischen Begriff "ludo", zu Deutsch "ich spiele" (Hemetsberger 2024), zur *ludo-Ebene* zusammen. Die *ludo-Ebene* setzt sich aus den besonderen Eigenschaften, Schnittstellen und Möglichkeiten des

Computerspielmediums zusammen.

Da wären zunächst strukturelle Aspekte wie die virtuelle Spielwelt mit samt ihren Objekten und Charakteren, die vom Spieler verfolgten Ziele und die Segmentierung des Gameplays, bspw. spatial durch Level oder temporal durch Zeitvorgaben. Auch die Interaktivität bzw. der Akt des Spielens selbst, der durch vom Spiel vorgegebene Affordanzen, also Handlungsoptionen, bestimmt ist, zählt zu den strukturellen Aspekten. Aus der Interaktivität ergeben sich schließlich auch die sensorischen bzw. physiologischen Aspekte, die auch als Noten der *ludo-Ebene* in Verbindung mit der Harmonie von Musik stehen können. Hierzu zählt in erster Linie die Kinetik der Spielereingaben. Auch wenn Diese sehr unterschiedlich ausgeprägt sein kann (Collins 2008, 127), vom einfachen Drücken eines Knopfes oder dem Bewegen eines Analogsticks<sup>12</sup> bis hin zu dem gestischen Trommeln auf speziellen Musikcontrollern wie im Spiel DONKEY KONGA, (Japan 2003, Nintendo, Gamecube) ist eine gewisse Kinetik immer in Spielen vorhanden und kann in direkter und echtzeitlicher Verbindung mit der Musik stehen. In entgegengesetzter Richtung dieser Benutzerschnittstelle zwischen Computerspiel und Spieler steht die Haptik, die durch Vibrationsmotoren in Controllern Feedback an den Spieler zurück gibt und seit Mitte der 1990er Jahre ein Standard in der Computerspielindustrie geworden ist<sup>13</sup>. Auch wenn sich Spiele wie andere „kommerzielle Medien vorwiegend an Auge und Ohr richten“ (Herwig 2017, 78), so ist das Computerspielmedium neben Mobiltelefonen das einzige Massenmedium, welches haptisches Feedback in seine sinnliche mediale Immersion integriert.

#### **4.2 Definition: ludoaudiovisuelle Harmonie**

In direkter Anlehnung an das Prinzip der *audiovisuellen Harmonie* beschreibe ich mit dem Begriff der *ludoaudiovisuellen Harmonie* Momente in Computerspielen in denen Elemente der Musik in direkter und synchroner Verbindung mit Elementen der Spielstruktur, dem Gameplay und der multimodalen Darstellung stehen und so die Harmonie der Musik erweitern. Die Zusammensetzung einer *ludoaudiovisuellen Harmonie* kann aus unterschiedlich vielen Elementen bestehen, die in Analogie zur Musik als Noten verstanden werden können. Im Gegensatz zu in der Komposition bspw. eines Filmes durch den Schnitt festgeschriebenen und für den Rezipienten nicht beeinflussbaren *audiovisuellen Harmonien* sind *ludoaudiovisuelle Harmonien*

---

12 Über die Jahrzehnte gab es dutzende verschiedene Eingabegeräte für Computerspiele bspw. auch Bewegungssteuerung, die mit der Nintendo Wii Konsole (2006) für einige Jahre sehr populär wurde. Mittlerweile lässt sich aber feststellen, dass die Innovation auf diesem Level ins Stocken geraten ist und Controller bis auf wenige Nuancen funktionell fast identisch sind.

13 Zwar gibt es bei der klassischen Maus- und Tastatursteuerung am PC keine Vibration, allerdings sind Controller mittlerweile für viele Genres auch am PC die bevorzugte Eingabemethode, mit Ausnahme von Genren wie z.B. Echtzeitstrategie oder Ego-Shooter.

in Spielen nicht strikt vorgegeben, sondern ihre Bedingungen und mögliche Verbindungen werden von den Entwicklern programmiert. Sie entstehen letztendlich aber erst durch den Akt des Spielens. Da jede Spielsitzung anders ist, sind auch die *ludoaudiovisuellen Harmonien* zeitkritisch von der jeweiligen Spielsitzung und den Handlungen des Spielers abhängig und somit in gewisser Weise wie der Spieler und die Spielsitzung individuell.

Es gibt zwei grundlegende Richtungen wie *ludoaudiovisuelle Harmonien* funktionieren können. Zum Einen kann die Musik auf das Spiel und/oder seine multimodale Darstellung wirken. Ein Beispiel hierfür wären die Gegner im 2D Plattformer *NEW SUPER MARIO BROS.* (Japan 2006, Nintendo, Nintendo DS), die in manchen Leveln im Spiel Bewegungen synchron zu gesanglichen Ausrufen in der linearen Hintergrundmusik ausführen. So springen Gegner kurz hoch (Abb. 10.1) oder drehen sich zur Kamera, wenn in der Musik kurze gesangliche Ausrufe vorkommen. Bestimmte Charaktere in der Spielwelt sind hier also direkt mit Elementen der Musik verbunden und erzeugen eine *ludoaudiovisuelle Harmonie*, die das Spielgeschehen beeinflusst, wie auch Collins feststellt: „Such use of audio indicates that game sound can be a significant element of gameplay in at least some cases, and that it can function in many ways [...]“ (Collins 2008, 128). So stellen diese Bewegungen der Charaktere zunächst eine erhöhte Schwierigkeit dar, da der Spieler sie nicht sofort erkennen kann. Wenn der Spieler die Verbindung jedoch versteht, kann er die Kadenz in der Musik, die zum gesanglichen Ausruf führt, und damit auch die Bewegungen der Gegner antizipieren und passende Handlungen ausführen.



Abbildung 10: 1 Gegner springen synchron zur Musik in *NEW SUPER MARIO BROS.*;  
2 Kugelgeschwindigkeit bestimmt Musiktempo in *SUPER MARIO GALAXY*

Die andere Richtung in der *ludoaudiovisuellen Harmonien* funktionieren können ist, dass der Spieler und/oder ein Aspekt der Spielstruktur die Musikwiedergabe beeinflussen. Das Level „Rolling in the Clouds“ aus dem 3D Plattformer *SUPER MARIO GALAXY* (Japan 2007, Nintendo, Nintendo Wii) liefert hier ein gutes Beispiel. So springt der Spieler bzw. sein Avatar Mario am Anfang des Levels auf eine Kugel, die daraufhin durch das Neigen des aufrecht zuhaltenden WiiRemote-Bewegungscontrollers gesteuert wird (Abb. 10.2). Je nachdem wie stark der Controller geneigt wird, desto schneller bewegt sich nicht nur die Kugel auf der Mario läuft, sondern desto schneller

wird auch die Musik des Levels abgespielt. Der Avatar Mario bzw. seine Steuerung sind hier also direkt und in Echtzeit mit der Musikwiedergabe verbunden und erzeugen so eine *ludoaudiovisuelle Harmonie*, die dem Spieler ein dauerhaftes, musikalisches Feedback über seine Handlung gibt: „3D platforming games *Super Mario Galaxy* (2007) and its sequel (2010) feature music that tightly synchronizes with, and dynamizes, in-game events.“ (Summers 2016, 193, kursiv i.O.).

In Kapitel fünf wird noch gezeigt, dass beide Richtungen, in die *ludoaudiovisuelle Harmonien* funktionieren können, auch gleichzeitig auftreten können. Zunächst soll jedoch diskutiert werden wie genau sich Film und Spiel und damit auch die vorkommenden Harmonien des Mediums mit der Musik unterscheiden, welche weiteren besonderen Eigenschaften des Computerspielmediums sein Verhältnis zur Musik allgemein prägen und welche Gemeinsamkeiten Musik und Spiel haben.

## **4.3 Diskussion**

### **4.3.1 Unterschiede audiovisuelle Sequenz Film und Computerspiel**

Zwar gibt es in vielen Computerspielen auch reine Videosequenzen, in denen der Spieler keinen Einfluss hat, wodurch die Audioproduktion von Spiel und Film durchaus ähnlich sein kann (Collins 2008, 5), und für die das Prinzip der *audiovisuellen Harmonie* nützlich ist, allerdings geht es in dieser Arbeit um die Verbindung von Gameplay und Musik, also um interagierbare audiovisuelle Sequenzen im Spiel. Auch in Spielsequenzen jedoch ist Musik ein „microcosm of [...] cross-media interplay“ (Summers 2016, 143) und nie nur eine Imitation von Filmmusik, denn Musik in Spielen muss sich immer den besonderen Eigenschaften des Computerspielmediums unterordnen. Gleichzeitig braucht Spielmusik eine Balance zwischen den spezifischen Erfordernissen des Mediums und den Erwartungen, die sich Spieler durch anderen Medien wie dem Film angeeignet haben (ebd. 143).

Abgesehen von Dialogen ist die Audioproduktion im Film zu einem großen Teil *post-production*, findet also erst nach Dreh und Schnitt statt, wenn in der finalen Version Toneffekte passend zum Schnitt hinzukommen und die verschiedenen Stimmen perfekt gepegelt werden (Collins 2008, 89). Dies ist möglich und nötig, da die Visualität des Filmbilds sich nach dem finalen Schnitt nicht mehr ändert, wohingegen das Bild in Spielsequenzen durch die Interaktivität des Mediums höchst dynamisch ist. Zwar ist der Spieler ebenso wie der Zuschauer eines Films i.d.R. statisch, verändert also nicht seinen Platz in der realen Welt, aber der vom Spieler gesteuerte Avatar hingegen ist in Bewegung, wodurch sich auch der Ton ständig anpassen muss (ebd. 105-106), was später noch mit dem Prinzip des *Point of Audition* genauer

erklärt wird. Wie bereits mit dem Prinzip der *audiovisuellen Harmonie* festgestellt wurde, sind Musik und Bild im Film oft sehr eng miteinander verknüpft, bspw. durch Schnitte, die synchron zu Beckenschlägen in der Musik stattfinden. Laut Collins ist so eine Synchronizität von Bild und Musik in Spielen auf Grund der unvorhersehbaren zeitlichen Aspekte von Spielen schwierig bis unmöglich (ebd. 128), allerdings werden die sechs zu untersuchenden Spiele in Kapitel fünf zeigen, dass auch das Computerspielmedium diese Synchronizität auf unterschiedlichste Weise erreichen kann, nicht obwohl sondern gerade wegen der hohen Dynamik.

In Spielen muss die Beziehung zwischen Material und Programm immer berücksichtigt werden. So muss das angestrebte synchrone Verhältnis zwischen Bild und Ton zunächst aufgebrochen und „in ein offenes, von der spielerischen Interaktion abhängiges Möglichkeitsfeld“ (Fischer/Schlüter 2009, 353) überführt werden. Statt wie im Film aus einem bestehenden Material, Film- und Tonaufnahmen, einen festen Schnitt zu produzieren, muss also erst ein dynamisches System geschaffen werden, das die unterschiedlichsten Verbindungen zwischen Elementen des Spiels berücksichtigt. Auf Tonebene geschieht dies durch *sound objects*, also Tondateien, die bestimmten anderen Objekten oder Befehlen im Programm zugeteilt werden, bspw. Charakteren oder Spielerhandlungen, und somit auch Teil der komplexen Verflechtung zwischen diesen anderen Objekten und Befehlen sind:

„All the sounds or objects are part of a dynamic environment. Their qualities [...] as well as the relations among them are defined by a program that can also be influenced by user action. Every sound object can potentially enter a temporal and spatial relation with every other sound or graphic object.“ (Stockburger 2003, 3)

Genau in diesem dynamischen System aus einer Vielzahl von Objekten und Verbindungen, die durch Spielregeln bestimmt sind, liegt das Potential adaptiver Ton- und Musikebenen in Computerspielen. Der Spieler ist dabei nicht mehr nur passiver Empfänger wie im Film, sondern zum Teil auch Sender bzw. Auslöser der Tonebene (Collins 2008, 3).

Allerdings ist das Erreichen einer möglichst adaptiven musikalischen Untermalung keineswegs leicht. Denn je dynamischer das Gameplay und die damit direkt zusammenhängende, angestrebte Adaptivität der Musik ist, desto schwieriger ist es Musik wie bspw. Sinfonien einzubinden, da diese auf zeitlicher Ebene Platz zur Entfaltung brauchen. So weisen Fischer und Schlüter daraufhin, dass die epochale Musik im Fantasy-Mittelalterrollenspiel *THE ELDER SCROLLS III: MORROWIND* (USA 2002, Bethesda, Windows PC) „von der hohen Frequenzdichte, mit der ein Statuswechsel zwischen *explore* und *combat* initiiert wird, unterlaufen“ (Fischer/Schlüter 2009, 367, kursiv i.O.) wird. Die Musik bekommt also einfach nicht genügend Zeit sich zu entfalten, da sie eng an den Spielstatus geknüpft ist, denn bei einem Wechsel vom friedlichen Erkunden hin zu einem Kampf, ändert sich die Musik ebenso wie wenn

der Kampf vielleicht nach wenigen Sekunden schon vorbei ist<sup>14</sup>.

Wie bereits von Chion festgestellt und von mir in Kapitel drei bestätigt, braucht und hat jedes audiovisuelle Medium eine gewisse Synchronisation von Bild und Ton. Auch Musikvideos, die eine größere horizontale Freiheit besitzen, werden durch Synchronisationspunkte von Bild und Ton zusammengehalten, die ein harmonischen Gerüst bilden (Chion 1994, 37). Für Computerspiele jedoch braucht es mehr als nur so ein Gerüst aus vereinzelt Synchronisationspunkten. Spiele haben weitaus weniger horizontale und vertikale Freiheit, da die audiovisuelle Darstellung vom Spieler genutzt werden muss, um in Spielen Leistungen zu vollbringen und Ziele zu erreichen. Eine De-Synchronisation von Bild und Ton ist in Spielen eigentlich nur dann vorzufinden, wenn im Spielprogramm Fehler, sogenannte Bugs oder Glitches, aufkommen. Selbst Spiele mit Splitscreen<sup>15</sup> haben eine totale Synchronisation von Bild und Ton über den gesamten Verlauf der Spielsequenzen, egal ob es sich um lokale Mehrspielermodi (Abb. 11.1) dreht wie in *TONY HAWK'S PRO SKATER 1 + 2* (USA 2020, Activision, Windows PC), oder um narrative Spiele wie *METAL GEAR SOLID 4* (Japan 2008, Konami, Playstation 3), bei dem Spielsequenz und Videosequenz parallel gezeigt werden (Abb. 11.2).



Abbildung 11: 1 Mehrspielermodus in *TONY HAWK'S PRO SKATER 1 +2*; 2 Splitscreen im narrativen *METAL GEAR SOLID 4*

Immer sind Toneffekte aus beiden Splitscreenbildern gleichzeitig zu hören und die Musik wirkt als verbindendes Element zwischen den Splitscreens mit ihren unterschiedlichen Bildern. Die durchgehende Synchronisation von Bild und Ton in Spielsequenzen hat i.d.R. leistungsspezifische Gründe, liegt also in der Interaktivität bzw. Ergodizität<sup>16</sup> des Mediums begründet. Insofern müssen auch die Spielereingaben Teil dieser Synchronisation sein. Wenn im Spiel eine Handlung ausgeführt wird, muss die visuelle und akusti-

14 Interessanterweise hat auch das Weltraumrollenspiel *STARFIELD* (USA 2023, Microsoft, Windows PC) vom gleichen Entwickler Bethesda immer noch ähnliche Probleme. Gleichzeitig wird die Musik hier als verbindendes Element zwischen den vielen Ladebildschirmen im Spiel genutzt.

15 "Unter einem Split Screen (auch: Splitscreen) versteht man die Teilung des Bildes auf einem Bildschirm in mehrere Bereiche." (Filmlexikon 2024).

16 Ergodisch bedeutet, dass die Textkonstitution von non-trivialen Handlungen der Spieler abhängig ist (vgl. Aarseth 1997, 1).

sche Repräsentation dieser Handlung möglichst synchron zum Befehl des Spielers erfolgen<sup>17</sup>.

Damit spiegeln Spiele eine Einheit von Bild und Ton wieder, die auch die weltliche Wahrnehmung des Menschen bestimmt<sup>18</sup>. Filme erreichen diese Einheit nicht bzw. nicht ohne erhebliche *post-production* der Tonebene, da selbst modernste Tonaufnahmen nicht ausreichen, um bereits während des Drehs alle nötigen Töne einzufangen (Chion 1994, 95). Diese Einheit oder natürliche Harmonie von Bild und Ton hat für Chion eine fast unermessliche Bedeutung:

„[...] this question of the unity of sound and image would have no importance if it didn't turn out, through numerous films and numerous theories, to be the very signifier of the question of human unity, cinematic unity, unity itself.“  
(Chion 1994, 97)

Sehen und Hören sind als Fernsinne sicherlich die wichtigsten Sinne des Menschen und entscheidend für unsere tägliche Wahrnehmung der Realität. Insofern ist es nicht überraschend, dass diese beiden Sinne vor allem in ihrer synchronen Kombination auch eine enorme Wichtigkeit für die Wahrnehmung virtueller Welten wie in Spielen haben und Spieler sich primär an dem orientieren was sie sehen und hören. Ton kann also auch in Spielen nicht separat von den anderen Modalitäten wie der Visualität untersucht werden (Stockburger 2003, 8). Auch Musik kann hier eine wichtige Rolle spielen, diese Synchronizität der verschiedenen Modalitäten aufrecht zu halten, denn sie ist eben kein Hintergrundphänomen, sondern ein direkter Bestandteil des Gameplays und vermittelt relevante Informationen an den Spieler (Summers 2016, 140).

#### **4.3.2 Gegenwärtigkeit, Subjektivität und Immersion**

Ein bewegtes Bild kann zeitlich eingefroren werden und weiterhin als Medium fungieren und Inhalte vermitteln. Ton und Musik jedoch hören auf zu existieren wenn sie pausiert werden. Sie müssen sich immer in der Zeit fortbewegen, um als Medium Inhalte vermitteln zu können. In Anlehnung an Lack<sup>19</sup> und Langer<sup>20</sup> kommt Bakels daher zu dem Schluss, dass Musik im Film ein Gefühl von Gegenwärtigkeit vermittelt: „In der ästhetischen Ausge-

---

17 Ein weiterer Beleg hierfür sind die seit 2020 aufkommenden technischen Entwicklungen seitens Grafikkartenherstellern wie AMD mit „AntiLag“ und nVidia mit „Reflex“ zur Vermeidung von Eingabeverzögerung (Horetski 2023).

18 Zwar gibt es auch in der menschlichen Wahrnehmung asynchrone Audiovisualität, bspw. wenn das Licht eines Blitzes deutlich früher unser Auge erreicht als die Schallwellen des Donners unser Ohr, aber in unserer unmittelbaren Nähe geschehen visuelle und akustische Wahrnehmung i.d.R. gleichzeitig.

19 Lack, Russell (1997): Twenty Four Frames Under. A Buried History of Film Music, London: Quartet Books. S. 283.

20 Langer, Susanne K. (1953): Feeling and Form. A Theory of Art Developed from Philosophy in a New Key, New York: Charles Scribner. S. 109.

staltung eines spezifischen ‚jetzt‘ [...] liegt das Potential der Musik“ (Bakels 2016, 174). Dieses von der Musik vermittelte Gefühl der Gegenwärtigkeit schafft aber gleichzeitig auch eine „in der Erfahrung von Zeit grundierte[n] Subjektivität, die unmittelbares Gefühl ist“ (ebd. 174). Vereinfacht ausgedrückt kann man sagen: wer Musik hört, der fühlt nicht nur die Zeit, sondern so zwangsläufig auch gleichzeitig sich selbst als Teil der fortlaufenden Zeit. Vor allem Echtzeitcomputerspiele sind auch durch Gegenwärtigkeit und Subjektivität gekennzeichnet. Der Spieler nimmt die virtuelle Welt aktiv wahr, entscheidet sich zu bestimmten Eingaben, führt so Handlungen in der Spielwelt aus, die wiederum umgehend darauf reagiert. Ein entscheidender Punkt liegt hier bereits in der Wahrnehmung des Spielers, denn alle Modalitäten, visuell, auditiv und haptisch, sind auf den Spieler ausgerichtet und vermitteln so auch ein Gefühl der Subjektivität. Das für den Film aufgestellte Konzept *Point of View (PoV)* und das davon abgeleitete Konzept des *Point of Audition (PoA)* (Chion 1994, 89) beschreiben von welchem Punkt aus in der diegetischen Welt das audiovisuelle Bild eine Situation repräsentiert. Chion hat bereits für den Film herausgestellt, dass das Bild sowohl *PoV* als auch *PoA* vorgibt (ebd. 91). Auch bei Computerspielen legt das Bild bzw. die Kamera den *PoA* fest. So ist der Ton in einem Spiel mit third-person Ansicht nicht von der Blickrichtung des Avatars abhängig, sondern von der Kamera<sup>21</sup>. Musik hingegen ist zumeist von dieser Subjektivität der Kamera entkoppelt und funktioniert nach einem anderen Prinzip, wodurch sie allerdings eine noch stärkere Funktion für die Subjektivität erfüllt als die restliche Tonebene mit *PoA*. So beschreibt Summers am Beispiel vom Snowboardspiel *SSX3* (Kanada 2003, EA, Xbox), dass die adaptive Musik des Spiels sich der Erfahrung des Spielers unterordnet, bspw. wenn bei einem großen Sprung des Avatars Filter über die Musik gelegt werden:

„This is part of what might be termed the fantasy of synergy: the magical effect of the hero to whom all other elements either submit or synchronize. The player is ‘in tune’ with the music and world, making them feel powerful and heroic, bolstered and supported in this construct. Music places the player ‘in the groove’. Miraculously, everything is in total synergy with them.“ (Summers 2016, 191-192)

Die Musik erreicht so eine Subjektivität, die nicht über die Kamera an die Wahrnehmung gebunden ist, wie der Rest der Tonebene durch den *PoA*, sondern an die Gegenwärtigkeit der Spielerfahrung. Musik kann in Computerspielen also Subjektivität und Gegenwärtigkeit zugleich vermitteln und diese verbinden. Im Grunde beschreibt Summers somit bereits Teilaspekte des Prinzips der *ludoaudiovisuellen Harmonie*, allerdings ohne genauer darauf einzugehen wie sie strukturell entstehen und mit der Darstellung eines Spiels zusammenhängen können. Summers Beispiel aus *SSX3* kann aber bereits als eine *ludoaudiovisuelle Harmonie* verstanden werden, da das Gameplay bzw. eine Spielhandlung und Spielstatus, in diesem Fall der

---

21 Spiele nutzen komplexe Technologien, um die Verknüpfung des Tons an die Kamera zu gewährleisten (siehe Fischer/Schlüter 2009, 357-358).

Sprung, die Musik echtzeitlich beeinflusst.

Aus der durch Musik vermittelten Subjektivität und Gegenwärtigkeit lässt sich auch ihre Wichtigkeit für Immersion ableiten. Unter Immersion versteht man „[I]m Sprechen über virtuelle Realitäten [...] Art und Maß des Eintauchens des Benutzers in die virtuelle Welt“ (Wulff 2022b). Ermi und Mäyrä definieren Immersion in Computerspielen nach drei Gesichtspunkten: Immersion durch sensorische Überladung, Immersion durch ein Zusammenspiel aus Herausforderung und Gameplay sowie eine imaginative Immersion, also das Hineinversetzen in Charaktere und Fiktion (Ermi/Mäyrä 2005, 7–8). Vor allem die ersten beiden Gesichtspunkte kommen bei Musik und *ludoaudiovisueller Harmonie* zum tragen. Lange Zeit wurde dem Ton und der Musik in Fragen zur Immersion jedoch eine wenn überhaupt nur untergeordnete Rolle zugeteilt (Rupert-Kruse 2014, 11), egal ob es um Filme, Spiele oder andere audiovisuelle Medien ging (Stockburger 2003, 1). Dabei ist Hören „aufgrund des involvierenden und omni-direktionalen Charakters des Tons“ (Rupert-Kruse 2014, 11) generell immersiver als Sehen und die Unmittelbarkeit der akustischen Wahrnehmung des Menschen, die evolutionär aus der „Notwendigkeit einer ständigen Gefahrenerkennung erklärbar ist“ (ebd. 11), wird auch durch Musik in Spielen kopiert. So ertönt in vielen Spielen, bspw. *THE LEGEND OF ZELDA: OCARINA OF TIME* (Japan 1998, Nintendo, Nintendo 64), eine Kampfmusik, wenn sich Gegner dem Avatar nähern, unabhängig davon ob diese bereits im Sichtfeld der Kamera zu sehen sind. Collins verweist auf die historisch immer schon immersive Funktion von Musik, bspw. im Stummfilm um den Lärm von Projektoren zu übertönen oder in den Arcades zur Überlagerung des Tons der Außenwelt (Collins 2008, 132-133). Wharton und Collins kommen in ihrer Studie über frei auswählbare Musik in *FALLOUT 3* (USA 2008, Bethesda, Xbox 360) unter Anderem zu dem Ergebnis, dass Musik erheblichen Einfluss auf die Immersion und emotionale Einbeziehung des Spielers haben kann (Wharton/Collins 2011) und auch Schwinning fasst die Erkenntnisse mehrerer Arbeiten<sup>22</sup> zusammen, wenn er schreibt, dass

„sie [Musik] in vielen Titeln eine zentrale Konstituente eines immersiven Spielerlebnisses [ist], indem sie Kommunikationsfunktionen übernimmt [...], Emotionen hervorruft, der Spielwelt in ihrer Vielfalt eine auditive Textur gibt [...] und besondere Ereignisse unterstreicht.“ (Schwinning 2023, 38-39)

Musik funktioniert als immersives Element in Computerspielen oft durch *ludoaudiovisuelle Harmonien*, denn durch die direkte Verbindung der Musik mit dem Spielgeschehen und seiner Darstellung ist die Musik direkt an der Gegenwärtigkeit der Spielwelt und des Spielens beteiligt sowie an der Subjektivität der Darstellung und Erfahrung. Andererseits kann Musik, die nicht zum Spielgeschehen passt, also jenes durch Asynchronizität sogar auffällig kontrastiert, negative Auswirkungen auf die Immersion haben, denn der

<sup>22</sup> Schwinning bezieht sich auf folgende Arbeiten: Jørgensen 2007, 64–71; Summers 2016, 116–136; Summers 2016, 57–63).

Spieler erkennt ihre Oberflächlichkeit und die fehlende Verbindung zum Spiel (vgl. Wharton/Collins 2011). Zu Spielen bedeutet für Claussen auch immer mit Musik zu interagieren, denn erst durch den Akt des Spielens folgt eine für die jeweilige Spielsitzung individuelle Musikausgabe (Claussen 2021, 59). Durch diese direkte Verbindung zwischen Spiel und Musik, bei der Gegenwärtigkeit und Subjektivität vermittelt werden, wird Musik zu einem besonders wichtigen Faktor für Immersion in Spielen. Die Stärke der Immersion hängt hierbei sowohl von der sensorischen Breite als auch Qualität ab (Rupter-Kruse 2014, 12), also wie viele *ludoaudiovisuelle Harmonien* wie häufig auftreten und wie viele Modalitäten der Darstellung, Aspekte des Gameplays und der Spielstruktur mit der Musik verbunden sind.

### 4.3.3 Gemeinsamkeiten Musik und Computerspiel

Verschiedenste Forscher haben bereits weitreichende Gemeinsamkeiten von Musik und Computerspiel herausgearbeitet oder zumindest angedeutet, auf die im Folgenden eingegangen werden soll. So werden sowohl Spiel als auch Musik als Performanzen betrachtet, „die auf einem Set von Regeln basieren und erst dann ihre volle Wirkung entfalten können, wenn sie gespielt bzw. ‚performed‘ werden.“ (Fritsch 2021, 69). Musik und Spiel brauchen also einen Akt des Spielens, um zu existieren bzw. wie Fritsch sagt um ihr volles Wirkungspotential als Medium ausschöpfen zu können. Dabei setzt sich die Performanz laut Fritsch mit Verweis auf Carlson<sup>23</sup> erstens aus einer Leistungsebene und zweitens aus dem zur Schau stellen von kulturellen Verhaltensmustern zusammen (ebd. 71-73). Vor allem Musik- und Rhythmusspiele wie GUITAR HERO oder DANCE DANCE REVOLUTION (Japan 1998, Konami, Arcade) beweisen wie beide Eigenschaften von Performanz sowohl auf Seiten der Musik als auch des Spiels in Symbiose treten können.

Musik kann allerdings auch direkt als Spiel verstanden werden, da man Musik gut mit klassischen Spieltheorien<sup>24</sup> erfassen kann. So verweist Claussen zunächst auf Huizinga, der Musik nicht nur immer als Spiel versteht (Huizinga 2004, 178), sondern auch die Schönheit des Spiels mit musikalischen Begriffen beschreibt:

„Die Schönheit des bewegten menschlichen Körpers findet ihren höchsten Ausdruck im Spiel. In seinen höher entwickelten Formen ist das Spiel durchwoben von Rhythmus und Harmonie, jenen edelsten Gaben des ästhetischen Wahrnehmungsvermögens, die dem Menschen beschert sind. (Huizinga 2004, 15)

Da Huizingas Buch „Homo Ludens“ (Erstveröffentlichung 1938) einer der Grundpfeiler in der Computerspielforschung ist, können seine Überlegung und Vergleiche von Spiel und Musik laut Claussen also auch für die Untersu-

---

23 Carlson, Marvin (2004): Performance: a critical introduction, London: Routledge.

24 Es muss angemerkt werden, dass es sich nicht um Computerspieltheorien, sondern um Spieltheorien handelt, die aber bis heute Einfluss auf die Computerspielforschung haben.

chung von Musik in Spielen nützlich sein (Claussen 2021, 35). Auch die Spieltheorie Caillois wird von Claussen genutzt um Musik als Spiel zu verstehen. So nutzt er die vier von Caillois aufgestellten Spielkategorien, um verschiedenste Praktiken in der Musik zu beschreiben. Vor allem die Illinx-Kategorie Caillois ist hier hervorzuheben, da sie ähnlich zu dem Flow-Begriff Csikszentmihalyis ist, der eine weitere Gemeinsamkeit von Musik und Spiel ausmacht, denn sie bezeichnen einen „Moment der vollständigen Hingabe im Spiel und an das Spiel, während eine Handlung innerhalb bestimmter Regeln und in einer fiktiven Umgebung ausgeführt wird, ohne die Regeln oder den fiktionalen Status zu reflektieren.“ (ebd. 47).

Csikszentmihalyi selbst fasst mit dem Begriff Flow die typischen Eigenschaften optimaler menschlicher Erfahrung zusammen: eine Tätigkeit, die in einem regelbasierten und zielgerichteten Handlungssystem stattfindet, das Rückmeldung über Leistung liefert, wobei die eigenen Fähigkeiten genau ausreichen, um bestimmte Herausforderungen zu meistern während die Konzentration so hoch ist, dass man über nichts anderes nachdenkt, ein verzerrtes Zeitgefühl hat und keinerlei extrinsische Motivation benötigt (Csikszentmihalyi 1990, 71). Neben sportlichen Aktivitäten berichten Menschen in Studien aber auch von einem Flow-Gefühl beim Computerspielen und Musikhören (Sites/Potter 2018). Es ist daher nicht verwunderlich, dass Spieleentwickler zum Erreichen von Flow in ihren Spielen auch vermehrt auf die Soundtracks der Spiele achten (ebd.).

Wenn es also so viele Gemeinsamkeiten zwischen Musik und Spiel gibt, die zu einem großen Teil ludischer also spielerischer Natur sind (Claussen 2021, 25), dann liegt eine enge Verbindung der beiden Medien im Computerspiel nahe. Wie vor allem Kapitel fünf im Anschluss zeigen wird ist das Prinzip der *ludoaudiovisuellen Harmonie* eine Möglichkeit diese Verbindung herzustellen, denn in der Synchronizität von Musik, multimodaler Darstellung und Gameplay vereinen sich auch Performanz und Flow-Gefühl.

## 5 Untersuchung

Folgend sollen sechs Computerspiele untersucht werden, um genau aufzuzeigen wie *ludoaudiovisuelle Harmonien* entstehen, aus was für Elementen sie sich zusammensetzen und welche Wirkung sie für die Rezeption haben können. Es handelt sich hierbei um Spiele, die u.A. speziell für ihre besondere Verwendung von Musik Bekanntheit erlangt haben und nach Summers eher loser Definition gewissermaßen als Musikspiele verstanden werden können:

„While this is a very diffuse genre, games given the designation ‘music game’ are characterized by providing some kind of unusual mode of interacting with music, or making the player (at least appear) musically empowered.“ (Summers 2016, 185-186)

Andererseits stehen sie nicht in Tradition klassischer Musikspiele wie Rhythmus- und Tanzspiele, Karaoke-spiele oder Spiele mit instrumentaler Performance (ebd. 179), sondern vermischen das lose Musikspielgenre mit anderen Genren wie Shootern oder Puzzlespielen. Wichtig ist, dass diese Spiele das Prinzip der *ludoaudiovisuellen Harmonie* sowohl in hoher Quantität als auch Qualität aufzeigen.

Eine Untersuchung von Musik in Computerspielen scheint zunächst recht einfach, da moderne Spiele ganz im Sinne der *New Media* Variabilität (Manovich 2001, 40) und damit ein Zuschneiden der Rezeption durch zahlreiche Einstellungsmöglichkeiten auf den individuellen Spieler ermöglichen. Das im Film aufgezwungene aurale Feld (Chion 1994, 33) existiert in dieser Art und Weise in Spielen nicht und auch das visuelle Feld (ebd. 33) wird zumeist vom Spieler bestimmt und ist somit freier, wodurch es allerdings nicht die eine Art zu spielen, also die eine Form der Rezeption gibt, sondern fast unendliche Viele. Zudem stechen einige der hier untersuchten Spiele durch eine gezielte sensorische Überladung der Wahrnehmung des Spielers auf. All diese Punkte gestalten die Untersuchung von Musik in Spielen und speziell das Erkennen von Verbindungen der Musik zu multimodaler Darstellung, Spielstruktur und Gameplay enorm schwierig, da nicht immer erkennbar ist ob ein Synchronisationspunkt der Musik mit Elementen des Spiels so programmiert oder rein zufällig ist. Daher wird im Sinne Chions versucht sich auf die „key points of synchronization, the primary synch points that are crucial for meaning and dynamics“ (ebd. 190), zu fokussieren. Hilfreich hierfür ist die eingangs erwähnte Abwandlung der Masking-Methode Chions, bei der Spielsequenzen mit unterschiedlichen Spieleinstellungen (Bild, Ton, Musik, Vibration) gespielt, aufgezeichnet und verglichen werden.

Um die letztendlichen Verbindung der Musik zum Spiel zu erkennen, benötigt es außerdem noch ein theoretisches Grundverständnis von Spielstruktur und Gameplay wofür die Theorien der ludologischen Zielen (Debus et al. 2020), der Segmentierung von Gameplay (Zagal et al. 2008) und dem Affordanzbegriff (Linderoth 2011) genutzt werden und die parallel zur Untersuchung erläutert werden.

Die Untersuchung verlangt somit zwar einen sehr spezielles, aktives Hören von Musik in Spielen, das aber in abgeschwächter und sicherlich unbewusster Form, auch von normalen Spielern bereits praktiziert wird. Denn Spieler versuchen beim Spielen die Beziehung zwischen Musik und Spiel zu entschlüsseln indem sie die Musik nach ludischer also spielerischer Signifikanz begutachten und andererseits im Spiel Verbindungen zur Musik suchen (Summers 2016, 187). Summers spricht hier von einer musikalischen Analyse bzw. kritischem Musikhören, die/das auf Gameplay basiert (ebd. 188), ohne aber konkreter darauf einzugehen wie dies geschieht und ob es je nach Spiel oder Genre Unterschiede gibt. Um die Hörweise der Musik in Spielen zu kategorisieren, werden in der Untersuchung die Spiele basierend

auf ihrer *ludoaudiovisuellen Harmonien* den Hörmodi Csikszentmihalyis zugeordnet. Dieser hatte in seinen Überlegungen zu Flow in Musik drei Kategorien aufgestellt wie Menschen aktiv Musik hören: sensorisch, analogisch/narrativ und analytisch (Csikszentmihalyi 1990, 110-111), wobei darauf hinzuweisen ist, dass die gleichzeitige Rezeption von audiovisuellen Medien oder gar interaktiven Medien mit interaktiver Musik nicht in Csikszentmihalyis Buch erwähnt wird. Trotzdem sind diese Hörmodi gut auf Spiele anwendbar, denn in Spielen hört der Spieler besonders aktiv zu. „It is not the *hearing* that improves life, it is the *listening*. [...] As with anything else, to enjoy music one must pay attention to it.“ (ebd. 109, kursiv i.O.). Gerade durch *ludoaudiovisuelle Harmonien* und eine enge Anbindung der Musik an das Gameplay, muss der Spieler die Musik bewusst wahrnehmen und ihre Verbindungen zum Spiel erkennen.

## 5.1 Musik generiert Spielaspekte

### 5.1.1 Beat Hazard 2

BEAT HAZARD 2 (nachfolgend BH2) ist ein Twin-Stick-Shooter, der im Kern das Spielprinzip des Arcadespiels ASTEROIDS (USA 1979, Atari, Arcade) übernimmt: der Spieler steuert durch Bewegungen des linken Analogsticks ein Raumschiff im zwei-dimensionalen Raum, um Weltraummüll, Gegnern und deren Waffenfeuer auszuweichen, und muss versuchen durch Bewegungen des rechten Analogsticks Gegner abzuschießen. Erweitert wird das Prinzip durch unterschiedliche Gegnertypen, Bossgegner, Waffensysteme, Perks (Bonis für die Raumschiffe) und zufällig generierte, freischaltbare Raumschiffe mit unterschiedlichen Eigenschaften und Aussehen.

Eine erste Verbindung der Musik mit dem Spiel ergibt sich dadurch, dass die Länge der Level jeweils davon abhängen wie lange ein vom Spieler ausgewähltes Lied dauert. Ziel ist es die gesamte Länge eines Liedes zu spielen, indem man nicht alle Leben verliert. Nach Zagal et al. liegt damit eine temporale Segmentierung des Gameplays vor, bei der die Musiklänge als Ressource genutzt wird: „A game can establish its total duration and any sub-periods and their length. When duration is explicitly regulated by a game, time is being treated as a resource.“ (Zagal et al. 2008, 180). Die große Besonderheit des Spiels liegt aber darin, dass basierend auf dem vom Spieler ausgewählten Musikstück, die Wellenform des Liedes während des Gameplays in Echtzeit durch einen fast-fourier-transform Algorithmus (Hunt 2018) ausgewertet wird. Die Wellenform wird in Frequenzen aufgeteilt aus der die Intensität<sup>25</sup> der Musik bzw. der einzelnen Frequenzen gemessen wird

<sup>25</sup> Mit Musikintensität ist hier gemeint wie hoch die jeweiligen Frequenzen in ihrer Lautstärke ausschlagen, siehe Abbildung 12.1. Die Intensität ist hierbei jedoch immer von dem Ausschlag der gesamten Wellenform abhängig und kann daher nicht mit Lautstärke an sich gleichgesetzt werden.

(Abb. 12.1). Anhand dieser Daten werden dann bestimmte Aspekte des Spiels beeinflusst, die sich so synchron zur Musik verändern und dauerhaft *ludoaudiovisuelle Harmonien* ergeben.

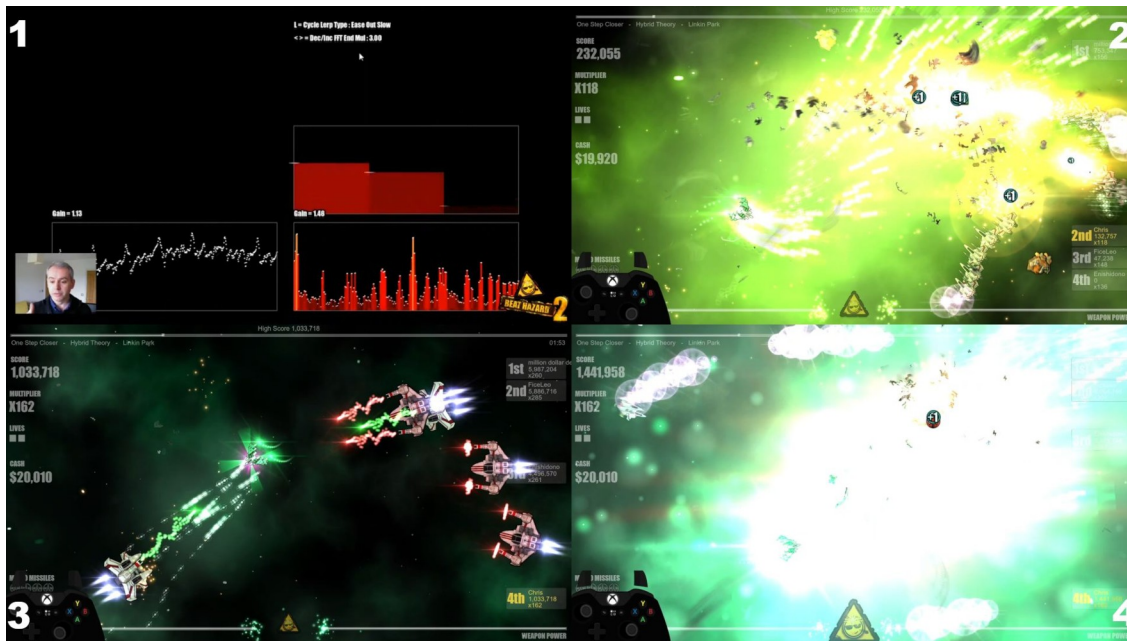


Abbildung 12: 1 Berechnung Intensität der Frequenzen; 2 & 3 visuelle Effekte abhängig von Musik; 4 überwältigende visuelle Effekte durch Musik

Die auffälligste Verbindung der Musikintensität zum Spiel ist die visuelle Darstellung. Je intensiver die Musik, desto stärker erscheinen bestimmte visuelle Effekte wie das Leuchten des eigenen Raumschiffs, der Weltraumnebel im Hintergrund, Leuchten von Powerups, Antrieb mancher Gegnertypen, Bossgeschütze und die eigenen Waffensysteme (Abb. 12.2 und 12.3). BH2 hat dadurch auf visueller Ebene eine große Ähnlichkeit zu Musikvisualisierungsplugins<sup>26</sup> für Programme wie Winamp (USA 1997, Radionomy S.A., Windows PC) oder VLC Media Player (Frankreich 2001, VideoLAN, Windows PC) und kann in Abhängigkeit von der Musikintensität eine visuelle Darstellung entwickeln, die mitunter überwältigend wirken kann (Abb. 12.4).

Die für das Gameplay bedeutsamsten Verbindungen der Musikintensität zum Spiel hängen aber mit den Affordanzen des Spielers und der Gegner zusammen. Unter Affordanz versteht Linderoth nach Gibson eine Handlung, die einem Organismus von seiner Umwelt angeboten wird (Linderoth 2011, 4). Auf Spiele übertragen heißt dies, dass ein Spiel dem Spieler und sogenannten non-playable characters (NPC) je nach Situation unterschiedliche Handlungsmöglichkeiten anbietet. Das Kerngameplay von BH2 besteht aus den Handlungsmöglichkeiten Bewegen und Schießen, die sowohl Spieler als auch Gegner nutzen können. Bis auf die Bewegung des vom Spieler gesteu-

<sup>26</sup> Das Spiel selbst hat sogar einen eigenen Modus, der genau diese Funktion erfüllt und kein Gameplay beinhaltet. Allerdings ist auch das Gameplay selbst eine Art Musikvisualisierung, durch seine vielen dauerhaften *ludoaudiovisuellen Harmonien*.

erten Raumschiffs sind alle Affordanz von der Musikintensität abhängig. Je stärker die Intensität, desto schneller bewegen sich die Gegner, ihre abgefeuerten Projektile und desto höher ist ihre Schussfrequenz. Auch die Waffenstärke des Spielers ist direkt mit der Musikintensität verbunden. Wenn die Intensität besonders hoch ist, erhöht sich auch der Schaden, den die Waffen bei den Gegnern anrichten können, was durch ein helleres Leuchten der Partikel visuell vermittelt wird. Insofern kann im Sinne Linderoths die Musik als Teil der Umwelt angesehen werden, die die Affordanz der einzelnen Entitäten im Spiel mitbestimmt. Aus diesen echtzeitlichen Verbindungen der Musik mit den Affordanz, die teilweise auch visuell stark widergespiegelt werden, ergeben sich bedeutungsvolle *ludoaudiovisuelle Harmonien*, die den Kern des Gameplays ausfüllen und aus dem sich der Spielspaß erklären lässt. Für den Spieler bedeutet dies nämlich, dass seine Handlungen direkt von der Musik abhängig sind, was am deutlichsten in der *ludoaudiovisuellen Harmonie* von Musikintensität und Waffenstärke zu erkennen ist. In Kombination mit den *ludoaudiovisuellen Harmonien* auf Seiten der Gegner ergibt sich so ein Gameplay, das am besten als eine Art *ludischer Tanz* beschrieben werden kann, der durch die Verbindungen mit der visuellen Darstellung in Echtzeit begleitet wird, fast wie die Lichtshow<sup>27</sup> in einer Discothek oder bei einem Konzert. In Abhängigkeit von der Musik bewegen sich die Gegner auf einen zu und feuern ihre Waffen ab. Je intensiver die Musik, desto schneller und gefährlicher sind die Gegner. Gleichzeitig ist aber auch der Spieler am stärksten wenn die Musikintensität besonders hoch ist.

Hier hilft es auch BH2 mit einer Perspektive auf die ludologischen Ziele zu betrachten. Unter ludologischen Ziele verstehen Cardona-Rivera et al. vorgegebene Ziele oder Konditionen innerhalb eines Spiels, die der Spieler erreichen soll: „These are designed in-game objectives or conditions players are expected to meet to succeed at a game; for example, score points, move an on-screen character, or trigger certain in-game events.“ (Cardona-Rivera et al. 2020). Dabei gibt es eine Unterscheidung zwischen dem größeren ultimativen Ziel und kleineren imperativen Zielen. Das ultimative Ziel bestimmt die Endkonditionen eines Spiels und es gibt drei Arten: „Win a game, Finish it, or Prolong the act of play“ (ebd.). In BH2 ist das ultimative Ziel eine bessere Leistung pro Level bzw. Lied aufzustellen als andere Spieler, also zu gewinnen (Win). Zum Erreichen des ultimativen Ziels geben Spiele auch eine Hierarchie von imperativen Zielen vor, die sich aus dem Design des Spiels ergeben: „Imperative goals are sub-ordinate (to ultimate) goals which more-concretely require the player effect a particular game state of affairs described by the game itself“ (ebd.). Cardona-Rivera et al. haben nach Untersuchung einer Vielzahl von Spielen zehn verschiedene imperative Ziele aufgestellt, mit denen sich die Zielhierarchie jedes Spiels be-

---

27 Programme wie DMXControl 2 (Deutschland 2024, DMXControl Projects e.V., Windows PC) können Musik analysieren und dazu Lichteffekte synchronisieren.

schreiben lassen soll, bis hin zu bestimmten Momenten des Gameplays: „Choose, Configure, Create, Find, Obtain, Optimize, Reach, Remove, Solve, and Synchronize“ (ebd.). Das Konzept des *ludischen Tanzes* ist auch in der Zielhierarchie des Spiels erkennbar (Tabelle 1), wo nur die beiden Ziele *Synchronize (Gegner, Waffenfeuer)* und *Prevent-Synchronize (Avatar, Gegner/Gegnerfeuer/Weltraummüll)* Verbindungen zur linearen Musikebene des Spiels aufweisen. Der Spieler geht also auf Grund der *ludoaudiovisuellen Harmonien* eine besondere Beziehung mit der Musik ein, die über die „musical-ludic mechanics“ (Summers 2016, 187) des Spiels bestimmt sind. Die Auswahl der Musik ist demnach ein entscheidender Faktor für das sich entfaltende Gameplay und die Rezeption des Spielers, ganz im Sinne des Werbespruchs des Spiels: „Gameplay powered by YOUR music“ (Beat Hazard 2 2018). Ein Lied mit besonders häufigen Unterschieden in der Intensität der Frequenzen, wird ein dynamischeres Gameplay hervorrufen, als ein Lied, das eher eine gleichbleibende Intensität besitzt. Die *ludoaudiovisuellen Harmonien* sind durch die angesprochenen Verbindungen zum Gameplay und dem daraus resultierenden *ludischen Tanz* auch der Grund für ein besonders starkes Flow-Gefühl in BH2. Das Gameplay spiegelt durch die *ludoaudiovisuellen Harmonien* das bereits durch das Hören der Musik aufkommende Flow-Gefühl synchron wieder, welches sich auf beiden Ebenen durch die Wechsel in der Intensität der Frequenzen ergibt.



Tabelle 1: Zielhierarchie BEAT HAZARD 2

### 5.1.2 Audiosurf 2

AUDIOSURF 2 (nachfolgend AS2) ist eine Mischung aus Puzzle-, Musik-, und Rennspiel. Der Spieler steuert ein Fahrzeug auf einer in drei Spuren aufge-

teilten Strecke und kann zwischen diesen drei Spuren wechseln. Ziel ist es die auf den drei Spuren entgegenkommende Blocks und Spikes so einzusammeln, dass das 3x7-Raster hinter dem Fahrzeug möglichst vollständig mit Blocks gefüllt wird. Das Einsammeln von Spikes führt zum Löschen eines bereits eingesammelten Blocks in der gleichen Spur. Bei mindestens drei eingesammelten und im Raster benachbarten Blocks beginnt ein Timer, der beim Ablaufen dazu führt, dass die bereits eingesammelten Blocks dem Punktekonto gut geschrieben werden und verschwinden. Ebenso werden die Blocks dem Punktekonto gut geschrieben, wenn eine Spalte des Rasters voll mit Blocks ist. Da es jedoch weitaus mehr Punkte gibt, wenn das Raster komplett gefüllt ist, ergibt sich daraus, dass es manchmal sinnvoll sein kann bestimmte Blocks nicht einzusammeln, um eine Spalte noch nicht komplett zu füllen, oder auch Spikes einzusammeln, um den Timer so kurz zu unterbrechen. Wie später noch gezeigt wird hat diese durchaus komplexe Puzzlemechanik auch Auswirkungen auf die vom Spiel erzeugten *ludoaudiovisuellen Harmonien*.

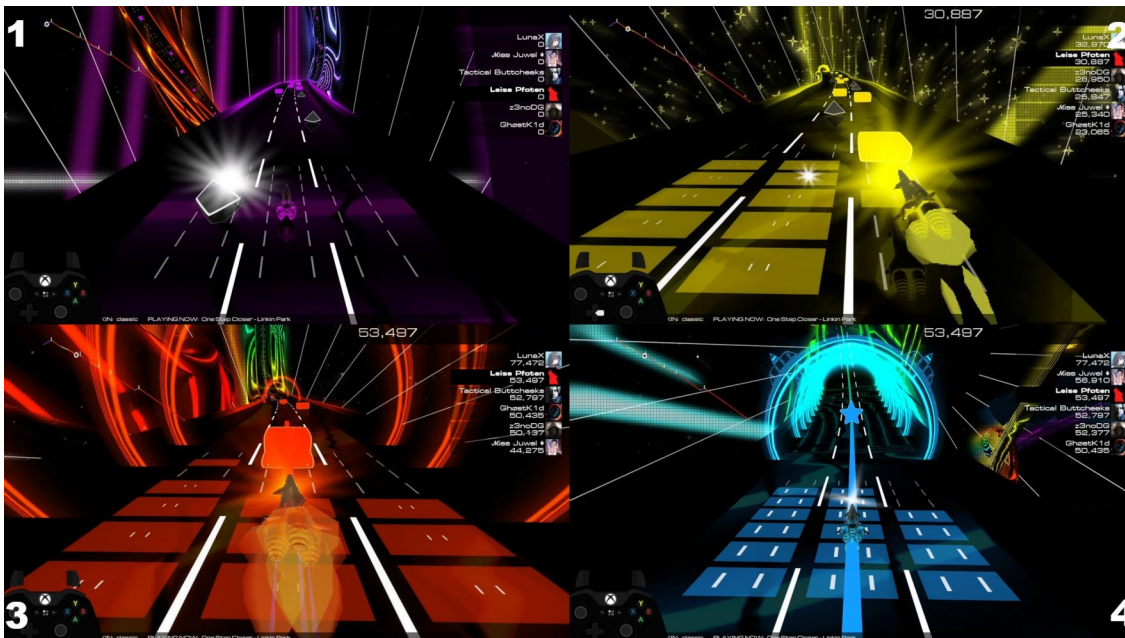


Abbildung 13: 1 blaue Blocks bei langsamem Tempo; 2 gelbe Blocks bei schnellerem Tempo; 3 rote Blocks bei sehr schnellem Tempo; 4 Powerblock bei besonderem Moment wie Refrain

Die Besonderheit von AS2 liegt wie bei BH2 darin, dass einige Spielaspekte von der vom Spieler ausgewählten Musik bestimmt werden. Zum einen legt die Länge eines Liedes auch die Länge eines Levels fest, wodurch es wie bei BH2 eine temporale Segmentierung des Gameplays durch die Dauer der Musik gibt. Zum anderen wird die Strecke, im Spiel Track genannt, von der ausgewählten Musik generiert. Jedes Lied erzeugt eine einzigartige vom Spieler zu fahrende Strecke. Somit liegt im Sinne Zagals et al. eine klassische spatiale also räumliche Segmentierung durch Level vor (Zagal et al. 2008, 182). AS2 nutzt hier auch einen speziell für das Spiel geschriebenen fast-fourier-transform Algorithmus (Macgregor 2018). Der Unterschied ist,

dass BH2 die Berechnung synchron zum Spielen durchführt, AS2 jedoch vor dem Spielen in wenigen Sekunden das Level aus der Musik berechnet. So wird anhand der Musikintensität die Geschwindigkeit des vom Spieler gesteuerten Fahrzeugs festgelegt. Langsame Sequenzen der Musik erzeugen eine Steigung und färben die einsammelbaren Blocks sowie visuelle Effekte im Hintergrund blau (Abb. 13.1). Schnellere Musikabschnitte jedoch erzeugen ein Gefälle und färben Blocks erst gelb (Abb. 13.2) und bei sehr schneller Musik mit steilem Gefälle rot ein (Abb. 13.3). Blocks und Spikes werden vor allem durch Ausschläge in den höheren Frequenzen der Musik generiert, die in der Regel von Perkussionstönen (Snare Drum, Becken, Hi-Hats etc.) kommen. Dies ist beim Spielen recht leicht erkennbar, vor allem in den langsameren Passagen von Pop- und Rock-Liedern, in denen weniger Instrumente bzw. Töne gleichzeitig zu hören sind. Dabei erfolgt der jeweilige Ton synchron zum Vorbeigleiten des Blocks auf Höhe des vom Spieler gesteuerten Fahrzeugs. Die Blocks müssen demnach also synchron zu bestimmten Tönen in der Musik eingesammelt werden wodurch sich *ludoaudiovisuelle Harmonien* ergeben. Beim erfolgreichen Einsammeln von Blocks kommt es zusätzlich zu Effekten in der multimodalen Darstellung. So leuchten die im Level verteilten Partikel farbig auf; blau während langsamer Passagen und gelb bzw. rot bei schnelleren Passagen. Gleichzeitig ertönt ein leichter Tuschton, der demnach als *effect sound object* einzuordnen ist, weil er von einem visuellen Objekt bzw. einem Ereignis in der Spielwelt ausgeht (Stockburger 2003, 5). Da diese *effect sound objects* direkt mit der Musik verbunden sind, können sie in gewisser Weise auch als Teil der vom Spieler rezipierten Musik angesehen werden. BH2 besitzt zwar auch solche *effect sound objects*, bspw. bei Explosionen von Gegnern, aber diese sind nicht mit der Musik verbunden. Des Weiteren erfolgt eine leichte Vibration des Controllers synchron zu jedem eingesammelten Block, wodurch die *ludoaudiovisuellen Harmonien* hier in ihrer sensorischen Quantität ausgeprägter sind als in BH2, da die PC-Version von BH2 anders als die Konsolenversionen keine Vibrationsfunktion besitzt. Im Gegensatz zu BH2 sind diese *ludoaudiovisuellen Harmonien* also nicht dauerhaft, aber treten sehr häufig auf und können in Abhängigkeit des Taktes der Musik auch sehr schnell hintereinander vorkommen, woraus sich eine der Herausforderungen im Gameplay ergibt.

Die Bedeutung der beschriebenen *ludoaudiovisuellen Harmonien* wird allerdings durch die Puzzlemechanik des Spiels zum Teil abgeschwächt. Um dies zu verstehen hilft es das Spiel in Hinblick auf seine ludologischen Ziele hin zu betrachten. Das ultimative Ziel in AS2 ist eine bessere Leistung pro Track aufzustellen als andere Spieler, also zu gewinnen (Win), was durch eine möglichst hohe Punktezahl geschieht. Durch die Puzzlemechanik mit seinem 3x7-Raster ergibt sich eine Zielhierarchie, die auf den unteren Ebenen eine gewisse Komplexität erlangt (Tab. 2), denn wie bereits beschrieben bringt das Auffüllen des gesamten Rasters die meisten Punkte. Dadurch wird der

Spieler dazu animiert Blocks und damit auch *ludoaudiovisuelle Harmonien* je nach konkreter Spielsituation bewusst auszulassen, um Spalten nicht zu früh aufzufüllen, sondern das gesamte Raster zu füllen, um so letztlich mehr Punkte zu generieren. Nur das kontingente Ziel<sup>28</sup> *Synchronize (Fahrzeug, Block/Spike)* weist eine Verbindung zur Musik auf und damit *ludoaudiovisuelle Harmonie* vor. Des Weiteren wird der Spieler durch die Puzzlemechanik dazu ermutigt Spikes einzusammeln, was allerdings zu einer Art *ludoaudiovisueller (Dis-)Harmonie* führt, denn beim Einsammeln von Spikes ertönt ein negativ konnotierter, elektrisch klingender Ton, fast wie eine Strafe. Die *ludoaudiovisuellen Harmonien* sind demnach nur ein Teil der unteren Ebene der Zielhierarchie, was den Spielspaß erheblich mindert, denn so gibt es nicht wie bei BH2 eine weitreichende Verknüpfung vom Flow-Gefühl des Gameplays und der Musik, sondern eine gewisse Trennung in bestimmten Momenten des Spiels.



Tabelle 2: Zielhierarchien AUDIOSURF 2

Alternativ können Spieler allerdings auch selbst Ziele definieren: „Player defined goals [...] are the goals that players bring to a game that are not necessarily considered by the game’s creators.“ (Debus et al. 2020). Wenn Spieler finden, dass das Auslösen der *ludoaudiovisuellen Harmonien* ihnen

28 Kontingente Ziele beschreiben Ziele in konkreten Momenten eines Spiels und bilden die unterste Ebene der Zielhierarchie (Debus et al. 2020).

den meisten Spielspaß ermöglicht, da so eine möglichst starke Verbindung zu ihrer zuvor ausgewählten (Lieblings-)Musik entsteht, können sie so eine alternative Zielhierarchie erstellen (Tab. 2), in der das Synchronisieren des Fahrzeugs mit den Blocks das entscheidende kontingente Ziel des Gameplays ist. Leider bietet AS2 allerdings keinen Modus, der dieses alternative Spielprinzip mit einem Fokus auf die *ludoaudiovisuellen Harmonien* ermöglicht. Trotzdem kann in Anlehnung an das *ludische Tanzen* in BH2 hier von einer Art *ludo-musikalischem Surfen* in AS2 gesprochen werden. Die Steigungen und Gefälle, die durch das Tempo der Musik generiert werden (spatiale Segmentierung), wirken wie Wellen und die trotz der Puzzlemechanik zahlreich und sehr offensichtlich auftretenden *ludoaudiovisuellen Harmonien* bieten viele synchrone Verbindungen zur vom Spieler ausgewählten Musik.

### 5.1.3 Diskussion: Musik generiert Spielaspekte

BH2 und AS2 sind mit ihrer durch den Spieler auswählbaren Musik und die dadurch beeinflussten Spielaspekte Paradebeispiele für Spiele als „co-creative media“ und der weiten Personalisierungs- und Anpassungsmöglichkeiten digitaler Medien (Wharton/Collins 2011). Auch wenn eine freie Musikauswahl durch Austricksen der Hardware schon bei bestimmten Spielen auf der ersten Sony Playstation möglich war<sup>29</sup>, so ist das Thema erst mit der sechsten und siebten Konsolengeneration wirklich populär geworden. Vor allem Microsoft war hier Vorreiter, denn bereits die erste Xbox hatte eine eingebaute Festplatte, auf der Musik gespeichert und so in manchen Spielen wiedergegeben werden konnte<sup>30</sup>. Mit der Xbox 360, auf der auch das erste BEAT HAZARD (UK 2009, Cold Beam Games, Xbox 360) erschien, gab Microsoft dann allen Spieleentwicklern vor dieses Feature in Spiele einzubauen (Wharton/Collins 2011). Generell unterscheiden Wharton und Collins bei der Anpassung von Spielen nur ob diese von den Designern geplant (*customization*) oder nicht beabsichtigt (*personalization*) ist (ebd.). Sowohl BH2 als auch AS2 haben mit ihrer frei auswählbaren Musik eine geplante Anpassung, denn nur so sind die vielen aufgezeigten Verbindung der Musik zu Spielstruktur, Gameplay und multimodaler Darstellung möglich. Allerdings muss die Dichotomie von Wharton und Collins trotzdem kritisiert werden, denn es gibt auch Spiele, die eine Zwischenlösung anbieten, bspw. viele Spiele des Entwicklers Bethesda wie FALLOUT 4 (USA 2015, Bethesda, Windows PC) bieten Modding-Unterstützung an, mit der Spieler weitreichende

29 In manchen Spielen wie RIDGE RACER (Japan 1994, Namco, Playstation) oder TWISTED METAL 4 (USA 1999, 989 Studios, Playstation) konnte man nach dem Laden eines Levels die Spiele-CD-ROM mit einer Audio-CD austauschen und so eigene Musik abspielen lassen. Möglich war dies, wenn alle für das Spiel notwendigen Daten bereits in den Arbeitsspeicher der Konsolen geladen wurden (Collins 2008, 69).

30 Allerdings unterstützten nur wenige Spiele wie SEGA GT 2002 (Japan 2002, Sega, Xbox) oder GTA SAN ANDREAS (USA 2005, Rockstar Games, Xbox) diese Funktion.

Änderungen an ihren Spielen vornehmen können, die im Detail aber nicht von den Entwicklern geplant sind. Dadurch werden beide Möglichkeiten gewissermaßen verbunden und aus der geplanten allgemeinen Anpassungsmöglichkeit wird durch Modding eine nicht vorher kalkulierbare Anpassungsmöglichkeit. Egal ob eine Anpassung geplant oder nicht geplant ist, der Spieler wird dadurch zu einer Art Teilautor des Spiels (Collins 2008, 106) und seiner möglichen Rezeption. Im Film wäre so etwas unvorstellbar, doch Spiele bieten seit jeher eine gewisse Individualisierung, die im Verlauf der Geschichte von Computerspielen nur zugenommen hat. So kann die Auswahl der Musik auch einfach als eine weitere Ebene dieser Individualisierung von Spielen angesehen werden (Wharton/Collins 2011). AS2 erweitert dieses Prinzip jedoch sogar noch, da das Spiel aktive Modding-Integration über den Steam-Workshop<sup>31</sup> bietet, wo Spieler eigene Skins erstellen und austauschen können, mit denen die visuelle Ebene der Tracks komplett überarbeitet wird. Insofern muss durchaus festgestellt werden, dass das Prinzip des Gesamtkunstwerks hier teilweise stark in Frage gestellt wird (ebd.), wenn nicht nur die Musik als bedeutender Teil des Spiels, sondern selbst die visuelle Darstellung vom Spieler ausgewählt bzw. erstellt werden kann. Trotzdem bleibt aber die Programmierung und das Design des Spielprinzips mit seinen direkten Verbindungen von Musik mit Segmentierung, Zielen, Affordanzen und multimodaler Darstellung weiter vorhanden und somit auch die bedeutsamen *ludoaudiovisuellen Harmonien*.

In Kapitel 3.1.5 wurde bereits gezeigt wie *audiovisuelle Harmonien* bedeutungsvolle Kadenzen im Film erschaffen können. BH2 und AS2 sind gute Beispiele dafür wie auch Computerspiele durch *ludoaudiovisuelle Harmonien* Kadenzen der Musik erweitern und für den Spieler bedeutungsvoller werden lassen können. In BH2 sorgen die *ludoaudiovisuellen Harmonien* aus Musik und Affordanzen, für eine doppelte Antizipation der Kadenzen in der Musik. Da die Waffenstärke des eigenen Raumschiffs, die Geschwindigkeit der Gegner und Projektile sowie ihre Schussfrequenz echtzeitlich an die Intensität der Musik gebunden sind und sich hieraus der *ludische Tanz* des Gameplays ergibt, ist es durchaus wichtig für den Spieler besonders starke Veränderungen der Musikintensität zu erkennen und zu antizipieren. Denn nur so kann der Spieler von Moment zu Moment die richtigen Entscheidungen treffen und erfolgreich sein. Bspw. kann der Spieler durch das Erkennen einer Kadenz, im Sinne einer schnellen Intensitätssteigerung, eine erhöhte Bewegungsgeschwindigkeit seiner Gegner antizipieren und so möglichst schnell ausweichen. In AS2 ist das Erkennen von Kadenzen weniger wichtig für eine gute Performanz, da der Spieler die kommenden Blocks und Spikes auf der Strecke in der Regel weit im Vorlauf sehen kann, womit die Kadenzen im Gegensatz zu BH2 auch visuell antizipiert werden können. Allerdings

---

31 Mit dem Steam-Workshop bietet Valve auf seiner Vertriebsplattform Steam eine direkte Schnittstelle für Spieleentwickler Modding-Unterstützung anzubieten, womit Spieler Mods nur noch downloaden und nicht über komplizierte Wege manuell installieren müssen.

nutzt AS2 den fast-fourier-transfrom Algorithmus, um Kadenzen in der Musik zu ermitteln und damit besondere Momente zu schaffen. So schafft es das Spiel die Anfänge von Refrains oder anderen besonderen Momenten in Liedern zu erkennen und erstellt für diese Momente Powerblocks auf der Strecke. Ein Powerblock, dargestellt durch ein Sternobjekt (Abb. 13.4), führt beim Einsammeln zu einem etwas stärkeren visuellen Leuchten der Partikel und pro Strecke gibt es einen größeren Powerblock, der beim Einsammeln einen Jubeltoneffekt auslöst, der die *ludoaudiovisuelle Harmonie* auf akustischer Ebene noch verstärkt und des Weiteren kulturell auf eine Konzertperformanz hinweist. Wichtiger ist aber, dass die Strecke für wenige Sekunden nach den Powerblocks zu einem Schraubenzieherlooping wird, indem das Einsammeln von Blocks mehr Punkte gibt, da das Einsammeln durch die starke Krümmung der Strecke deutlich schwieriger ist. Das Spiel nimmt also Höhepunkte aus der Musik und gestaltet synchron dazu auch Höhepunkte des Gameplays. Beide Spiele zeigen, dass es bei Kadenzen in Zusammenhang mit *ludoaudiovisuellen Harmonien* in Spielen um die perfekte Erfüllung der Kadenzen geht, anders als in Filmen wo das Nichterfüllen von Kadenzen überraschend und interessanter sein kann (Chion 1994, 56), was auch durch die Teilerfüllung beim Beispiel aus PULP FICTION in Kapitel 3.1.5 verdeutlicht wurde, wo das Bild nicht Teil dieser Erfüllung ist und erst dadurch der Witz der Szene entsteht. In Spielen jedoch muss es eigentlich immer eine volle Erfüllung geben, da wie gezeigt die Kadenzen der Musik durch *ludoaudiovisuelle Harmonien* mit dem Gameplay verbunden sind und Spieler diese für eine gute Performanz erkennen müssen. Die Besonderheit liegt hier aber auch darin, dass Spieler ihre eigene Musik auswählen können und so die Kadenzen bereits kennen, beim früheren Hören der Musik antizipiert haben und Erfüllung fanden. Denn gerade hier liegt ja der Reiz von Spielen wie BH2 und AS2: der Spieler bringt seine Lieblingsmusik mit ins Spiel und erlebt diese nun auf eine spielerische Weise neu, die durch das Prinzip der *ludoaudiovisuellen Harmonie* möglich wird. Eine Unterbrechung der Verknüpfung des Flow-Gefühls, bestehend aus Musik, Gameplay und multimodaler Darstellung, kann nur negativ auf die Rezeption wirken und würde wohl eher als Fehler im Programm gedeutet werden.

Insofern sind beide Spiele auch durch den gleichen Hörmodus geprägt. Da der Spieler selbst die Musik auswählt und ihren Verlauf kennt, stellt sich ein analogischer Hörmodus nach Csikszentmihalyi ein. Die Musik weckt auch durch Kadenzen im Spieler Gefühle oder auch Bilder hervor (Csikszentmihalyi 1990, 111), die der Spieler bereits vorher schon mit der Musik in Verbindung gebracht hat. Die Verbindung der Musik zum Gameplay ist zwar wie gezeigt durchaus gegeben und wird antizipiert, aber es kann trotzdem nicht von einem analytischen Hörmodus gesprochen werden, da nicht das Erkennen von musikalischen Strukturen wie Rhythmus gefordert wird; anders als bei den nächsten beiden zu untersuchenden Spielen.

## 5.2 Rhythmus als Fundament für Gamedesign

### 5.2.1 Metal: Hellsinger

METAL: HELLSINGER (nachfolgend MH) ist ein First-Person-Shooter (FPS), mit sehr starker Inspiration durch das DOOM (USA 2016, Bethesda, Windows PC) Reboot von 2016<sup>32</sup>. Die Besonderheit des Spiels ist, dass der Spieler viele Handlungen, insbesondere das Schießen der Waffen, im Rhythmus der Musik durchführen soll. Die Musik selbst ist dem Death-Metal Genre zuzuordnen und besteht aus instrumentalen Eigenkompositionen des Studios Two Feathers (Hjertberg/Björkman 2024), die gesanglich durch verschiedene bekannte Sänger aus der Metalszene komplettiert werden z.B. Tetjana Schmajljuk (Jinjer), Serj Tankian (System of a Down), etc. Solche Synergien zwischen Computerspiel- und anderen Kulturindustrien sind keine Seltenheit und bereits seit den 1980er Jahren in verschiedensten Formen vorhanden, bspw. durch Lizenzierung von Musik, um Kosten für Eigenproduktionen zu sparen (Collins 2008, 108), oder „the use of star talent in voice recording“ für Werbezwecke (ebd. 108).



Abbildung 14: 1 Fury-Multiplier auf Stufe 1; 2 Finisheranimation; 3 Bosskampf; 4 Flammen synchron zum Rhythmus

Die *ludoaudiovisuellen Harmonien* liegen primär im Rhythmus der Musik. Vier für das Spielprinzip wichtige Affordanzen soll der Spieler möglichst synchron zum Rhythmus der Musik ausführen<sup>33</sup>, wobei alle Stimmen der Musik,

<sup>32</sup> Beide Spiele sind First-Person-Shooter, besitzen ein Höllensetting, in dem man Dämonen bekämpfen muss, und haben einen Soundtrack bestehend aus Metalmusik.

<sup>33</sup> Affordanzen, die keine Verbindung zur Musik besitzen, sind das allgemeine Bewegen des Avatars, Sprünge, Kamerasteuerung und Waffenwechsel. Die Affordanzen der Gegner sind

also auch E-Gitarre, Bass und Gesang eine Vielzahl an Harmonien mit dem Rhythmus aufweisen, weswegen der Spieler sich nicht nur am Schlagzeug orientieren kann. Da ist zum einen das Schießen der Waffen, das Gegnern mehr Schaden zufügt, wenn es synchron zum Rhythmus geschieht. Auch das Nachladen der Waffen und das Ausführen einer dreifachen Dash-Bewegung<sup>34</sup> synchron zum Rhythmus wird belohnt durch ein Auffüllen des Fury-Multipliers (Abb. 14.1), dessen fünf Stufen (1x, 2x, 4x, 8x, 16x) Waffenstärke und Punktezahle beeinflussen. Der Multiplier leert sich allerdings automatisch über Zeit und muss daher ständig durch Aneinanderreihung möglichst vieler im Rhythmus durchgeführter Affordanzen wieder aufgefüllt werden. Dadurch gibt die Musik in gewisser Weise den kontinuierlichen und schnellen Bewegungsrhythmus des Spielers vor, ähnlich wie Fischer und Schlüter dies für den Metal-Soundtrack in *QUAKE 2* (USA 1998, Activision, Windows PC) beschrieben haben<sup>35</sup> (Fischer/Schlüter 2009, 363), denn was zählt ist die Synchronisation mit dem schnellen Rhythmus der Musik. Zuletzt gibt es noch Finisher, die dann ausgeführt werden können, wenn ein Gegner zuvor bereits geschwächt wurde. Die Betätigung des linken Triggers synchron zum Aufleuchten des geschwächten Gegners, löst eine Animation des Avatars aus, bei der der Gegner auf brutale Weise im Nahkampf besiegt wird (Abb. 14.2). Als Belohnung wird nicht nur der Fury-Multiplier aufgefüllt, sondern der Avatar erhält auch Lebensenergie zurück. Eine fast identische Funktion gab es bereits im 2016er *DOOM*, mit dem Unterschied, dass diese Affordanz in *DOOM* aber nicht mit dem Rhythmus der Musik synchronisiert werden musste. In MH sind die Finisher jedoch noch befriedigender, da selbst die Animation des Nahkampfs im Anschluss zum Rhythmus der Musik erfolgt und somit Teil der *ludoaudiovisuellen Harmonie* ist. Jeder Schuss, jedes korrekte Nachladen, Dashes und Finisher werden durch Vibration des Controllers begleitet, womit die *ludoaudiovisuellen Harmonien* auch eine haptische Komponente besitzen. Zusammen mit den konkaven Triggern des Xbox Series X Controllers entsteht so beim Schießen der Waffe eine *ludoaudiovisuelle Harmonie*, die den Abzug der Waffe samt Vibration nicht nur simuliert, sondern haptisch und auch kinetisch erfahrbar macht (Herwig 2017, 81).

In der Zielhierarchie (Tab. 3) stechen diese vier Affordanzen auch hervor, da sie alle kontingente Ziele für die Erhöhung des Fury-Multipliers sind. Dieser erhöht nicht nur den Schaden der Waffen und die Punktezahle, sondern beeinflusst auch direkt die Musikwiedergabe, denn je höher der Multiplier, desto mehr Ebenen der Musik sind zu hören. Bis auf die Musik in Bosskämpfen (Abb. 14.3) nutzt die Musik in MH nämlich die branching-Methode, bei der

---

generell nicht mit dem Rhythmus der Musik verbunden.

34 Mit dem Dash kann der Spieler seinen Avatar per Tastendruck schnell in eine Richtung bewegen, was zum Ausweichen von gegnerischen Angriffen hilfreich ist.

35 Der Fokus auf die Musik macht es auch unmöglich, dass der Spieler sich über den Sound in der Spielwelt orientieren kann wie in *DOOM 3* (USA 2004, Bethesda, Windows PC) (Berndt 95). Auch dadurch muss der Spieler in MH immer in Bewegung bleiben.

ein lineares Lied aus unterschiedlichen layers bzw. branches besteht, die an- und ausgeschaltet oder durch andere Versionen ausgetauscht werden können (van Geelen 2008, 96). Die Musik wird also erst dann wirklich intensiv, wenn der Spieler bereits eine gewisse Performanz gezeigt hat und die vier angesprochenen Affordanzen möglichst synchron zum Rhythmus der Musik ausgeführt hat. Erst bei der höchsten Ebene (16x) wird der Gesang der Lieder hinzu geschaltet, was gleichzeitig auch die höchste Waffenstärke sowie Punktezahl für abgeschossene Gegner bedeutet und somit eine typische Verwendung der branching- bzw. layer-Methode ist (ebd.). Durch die *ludoaudiovisuellen Harmonien* entsteht so eine Wechselwirkung zwischen Affordanzen, Zielen und Musik, die den Kern des Gameplays und Spielspaß ausmacht. Je besser der Spieler diese Affordanzen zum Rhythmus der Musik ausführt, desto intensiver wird die Musik und Waffenstärke sowie mögliche Punktezahl werden erhöht. Auch hier entsteht so ein Flow-Gefühl, dass durch die Kombination von Musik und Gameplay hervortritt und gewissermaßen zwei Flow-Gefühle miteinander verbindet.

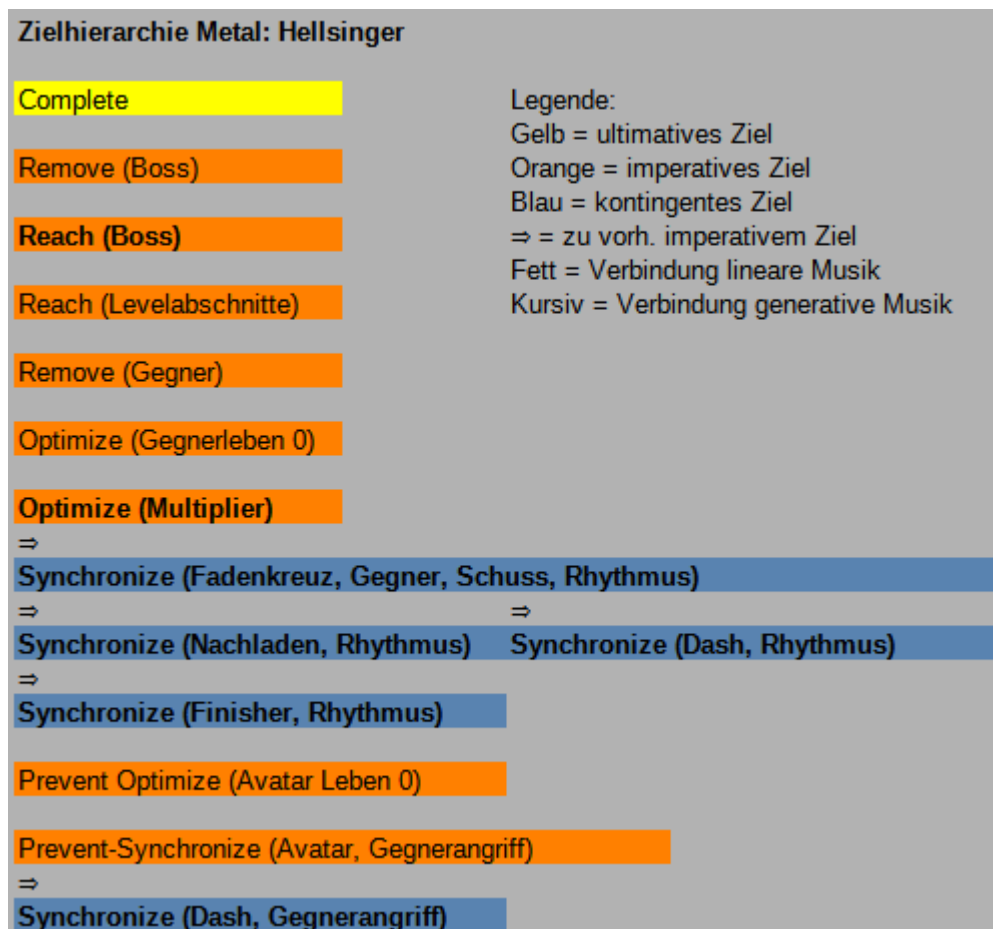


Tabelle 3: Zielhierarchie METAL: HELLSINGER

Da der Gesang nur auf der höchsten Stufe des Multipliers zu hören ist, gibt es hier auch eine Verbindung der Musik zur Narration des Spiels. Die Geschichte des Spiels dreht sich nämlich um einen gefallenen Engel, der sich

durch die Hölle kämpft, um seine Stimme wiederzuerlangen und so zum Hellsinger zu werden, was durch die höchste Ebene des Multipliers und der Einblendung des Gesangs-layers auch im Gameplay wiedergespiegelt wird. Auch die Lyrics der Musik nehmen direkt Bezug auf die Protagonistin, wahlweise mit dem Titel Unknown oder Hellsinger, und ihrem Kampf gegen die Dämonen der Hölle (Genius.com 2022). Die Fusion von Gameplay und Narration, die ein Ziel von Spieleentwicklern ist (Summers 2016, 145), wird hier durch die Musik, ihren Verbindungen zu mehreren Spielaspekten und somit durch *ludoaudiovisuelle Harmonien* erreicht. Die Musik in MH erfüllt somit auch eine Funktion des *epic texturing*:

„music [...] connects the player both to the subjectivity of the avatar and to the higher-level background narrative in which the avatar is placed. This effect immerses players through hero-izing them. It asserts both their adoption of the hero-role and their agency within a wider narrative.“ (Summers 2016, 68)

In der Musik werden die Protagonistin und ihr Kampf gegen die Dämonen der Hölle nicht nur besungen, sondern durch die Verbindungen der Affordanzen zum Rhythmus der Musik, ist auch der Spieler Teil dieser Symbiose. Durch das Spielen mit Hinblick auf die *ludoaudiovisuellen Harmonien* über den Rhythmus erhöht sich der Multiplier, was ein oberes imperative Ziel der kontingenten Ziele ist (Tab. 3), und nicht nur die Musik, sondern auch die Protagonistin bekommt ihre Stimme zurück. Die Verbindungen der Musik mit den anderen Spielaspekten schaffen so Subjektivität, die gleichzeitig aber auch an die übergeordnete Narration ankoppelt und durch den Gesang darauf verweist.

### 5.2.2 Hi-Fi RUSH

Hi-Fi RUSH (Nachfolgend HFR) verbindet gekonnt die beiden Genre Hack'n'Slash und Rhythmusspiel. In third-person Ansicht steuert der Spieler den Protagonisten Chai, einen jungen Rockmusikfan, durch schlauchartige Level, bewältigt Plattforming-Passagen und kämpft gegen unterschiedlichste Robotergegner und Bosse. Die Musik ist eine Mischung aus lizenzierten und eigens fürs Spiel komponierten Rockliedern, die teilweise auch elektronische, industrial oder klassische Komponenten aufweisen.

Ebenso wie in MH dient der Rhythmus der Musik hier als Fundament für das Gameplay, allen voran für das umfangreiche Kampfsystem. So kann Protagonist Chai eine aus Elektroschrott zusammengesetzte Gitarre nutzen, um damit leichte und schwere Nahkampfangriffe auszuführen, die in unterschiedlichste Kombos aufgehen können, wenn sie in bestimmten Takten synchron zur Musik durchgeführt werden. Auch das Blocken von gegnerischen Attacken soll zum Rhythmus der Musik erfolgen, denn selbst die Attacken der Gegner sind mit der Musik synchronisiert, anders als in MH wo die Gegner keinerlei Verbindung zur Musik aufweisen. Eine hohe Synchronizität

der Affordanzen zum Rhythmus wird ebenso wie in MH mit höherem Schaden der Attacken und mehr Punkten belohnt und ist somit auch mit den ludologischen Zielen des Spiels verbunden (Tab. 4). Selbst bei manchen der wenigen nicht mit dem Rhythmus in Verbindung stehenden Affordanzen wie dem Laufen des Avatars, erfolgt die Animation trotzdem noch im Rhythmus der Musik.



Abbildung 15: 1 Comiceffekte bei Angriffen; 2 Objekte in Spielwelt synchronisiert mit Rhythmus; 3 visuelle Effekte bei Plattforming; 4 visuelle Hilfe durch horizontale Linien nahe Avatar

HFR setzt allerdings auf eine deutlich weitreichendere Verbindung der Musik und Affordanzen mit der multimodalen Darstellung als MH. Generell bietet das Spiel mit seiner Cel-Shading Grafik eine höchst stilisierte Comicvisualität. Diese wirkt auch im Gameplay, denn die Attacken Chais rufen visuelle Comiceffekte hervor (Abb. 15.1). Gleichzeitig erklingen zu den Attacken und auch zum Blocken musikalische Geräusche wie der Akkord einer E-Gitarre bei einer Attacke oder ein einzelner Gitarrenton beim Blocken, weshalb hier das erste Spiel dieser Untersuchung mit einer Art generativer Musikebene vorliegt, also musikalischen Tönen und Motiven, die durch bestimmte Handlungen des Spielers hervorgerufen werden. Außerdem wird jede Attacke auch von der Vibration des Controllers begleitet, die den Spieler in HFR im Sinne eines transmedialen Storytellings haptisch affizieren will (Herwig 2017, 85) und enorm die Spieler-Avatar-Beziehung festigt, denn der Spieler fühlt so quasi alles was auch Chai fühlt<sup>36</sup>. Alle diese Effekte wirken wie die Einschlagsgeräusche, die Chion für die audiovisuelle Sprache des Films mit dem Akkord in der Musik verglichen hat (Chion 1994, 62) und verstärken

<sup>36</sup> Die Vibration gekoppelt an bestimmte Ereignisse setzt sich auch in Videosequenzen fort, bspw. wenn Chai stürzt, wird dies synchron durch den Vibrationseffekt haptisch an den Spieler vermittelt.

dementsprechend die *ludoaudiovisuellen Harmonien* bei Attacken. Im Spiel hilft diese multimodale Kombination von Effekten nämlich dem Spieler erkennen zu lassen ob er einen Gegner getroffen hat und lassen ihn so gleichzeitig auch ermächtigt fühlen, da seine Attacken visuelles, akustisches, musikalisches und haptisches Feedback zur Folge haben. Bei erfolgreichen Kombos, also drei in einem bestimmten Takt ausgeführte Attacken auf einen Gegner, erklingt sogar ein kurzes Gitarrenriff. Durch den akustischen Teil dieser *ludoaudiovisuellen Harmonien* wird somit die Musikwiedergabe durch thematisch passende Klänge noch erweitert und der Spieler beeinflusst die Musik noch direkter als wie mit der branching-Methode und dem Fury-Multiplier in MH. Auch wenn der Spieler durch gewisse Affordanzen musikalische Klänge hervorbringen kann, die zur Musik des Spiels passen und sogar mit dem Rhythmus synchronisiert sind, wäre ein Verständnis des Spielers als „mimetischer Performer eines musikalischen Spiels“ (Claussen 2021, 60) oder gar dem Spiel als Musikinstrument (ebd. 62) hier unzureichend und verkürzt. Man kann allerdings sagen, dass HFR dem Spieler dies durch seine *ludoaudiovisuellen Harmonien* durchaus suggerieren will und dort auch ein nicht unerheblicher Teil des Spielspaß liegt.

Besonders hervorzuheben ist, dass die Musik in HFR durchgehend diegetisch zu sein scheint. Bereits in der Introsequenz des Spiels wird die Musik nämlich schlagartig leiser, sobald Chai seine Ohrstöpsel herausnimmt. Wenig später erfolgt dann das Ereignis mit dem auch die gesamte Synchronizität der Affordanzen mit der Musik erklärt wird, denn Chai fällt aus Versehen in eine Maschine, die ihm sein Musikabspielgerät ins Herz verpflanzt. Dies hat zur Folge, dass der Rhythmus der Musik nun Chais Herzschlag ist, was im Spiel auch ganz konkret so durch einen Monolog Chais erklärt wird. Doch die Musik hat auch eine synchrone Verbindung zur Spielwelt. So gibt es in jedem Level zahlreiche Objekte, die sich zum Rhythmus der Musik bewegen, bspw. wippende Bäume oder auf Fließbändern und Rohrsystemen fahrende Container (Abb. 15.2). Somit sind auch die Sprungpassagen vom Rhythmus bestimmt und weisen *ludoaudiovisuelle Harmonien* auf, wenn Kisten auf einem Fließband sich ruckartig synchron zum Rhythmus der Musik fortbewegen und der Spieler auf diese springen muss, um im Level voran zu kommen. Auch hier gibt es audiovisuelle Begleiteffekte, wenn Plattformen mit Beckenschlägen erscheinen und das Landen auf solchen Plattformen ebenfalls mit Perkussionstönen und einem visuellen Comiceffekt begleitet wird (Abb. 15.3). In der Zielhierarchie ist dies durch das kontingente Ziel *Synchronize (Avatar, Umgebung)* markiert (Tab. 4). Insofern gilt in HFR sehr stark das bereits in Kapitel 4.3.2 angesprochene Prinzip der *fantastic synergy*, bei dem sich alle Elemente der Spielwelt, auch die Musik, mit dem Avatar synchronisieren (Summers 2016, 191-192) und die in HFR zumindest teilweise über die Narration erklärt wird.

Die Musik in HFR ist Teil der spatialen Segmentierung des Spiels in Level und Checkpoints. Jedes Level hat eine Auswahl an Liedern, die an bestimm-

te Abschnitte gekoppelt sind und durch spatiale Checkpoints aktiviert werden. Ein spatialer Checkpoint bedeutet, dass es bestimmte Bereiche in einem Level gibt, die kontinuierlich verbunden und erreichbar sind, aber an denen sich das Gameplay ändert (Zagal et al. 2008, 184-185). Dies ist vor allem dann ersichtlich, wenn es zu Kämpfen mit Gegnern kommt, die nur an vorher festgelegten Orten aufkommen und dann auch abgesperrt werden, bis alle Gegner besiegt sind. Die Musik wird hier deutlich schneller und intensiver. Zudem setzt hier meist der Gesang der Lieder ein, der also anders als in MH nicht leistungstechnisch an Affordanzen und den dazu gehörenden ludologischen Zielen gekoppelt, sondern an die spatiale Segmentierung des Gameplays und den damit zusammenhängenden Zielen gebunden ist: *Reach (Boss)* und *Reach (Levelabschnitte)* (Tab. 4).



Tabelle 4: Zielhierarchie Hi-Fi RUSH

Besonders ein Kampf fällt hier durch seine musikalische Untermalung auf. Wenn Chai in Level 10 „A Masterplan“ ein Gebäude infiltriert und dort unverhofft in die Mensa der Robotergegner einfällt, ertönt das durchaus populäre Lied INVADERS MUST DIE! (UK 2009, The Prodigy). Die roboterartige Stimme, die mehrfach den Titel des Liedes singt, begleitet den folgenden Kampf gegen dutzende aufgebrachte, da beim Essen gestörte, Roboter äußerst ironisch und ist ein Höhepunkt des Spiels (Abb. 16.1 und 16.2). Die Nutzung vorkomponierter bzw. lizenzierter Musik an bestimmten Punkten im Spielverlauf kann wie Collins anmerkt durchaus problematisch sein, da dies

Spieler ablenken oder das Lied die Bedeutung des Spiels bzw. andersherum das Spiel die Bedeutung des Liedes ändern kann (Collins 2008, 118). Nach Gibbons wird lizenzierte Musik in narrativen Spielen generell ironisch verwendet (Gibbons 2011), allerdings ist seine Untersuchung des Spiels *BIOSHOCK* (USA 2007, 2K Games, Windows PC) mitunter nicht korrekt, da es sich bei den meisten Liedern in *BIOSHOCK* um instrumentale Versionen handelt, die bei einem normalen Durchspielen wenn überhaupt nur für wenige Sekunden zu hören sind. Zudem gibt es auch narrative Spiele, die lizenzierte Lieder ernst verwenden wie in der *METAL GEAR SOLID* Reihe (Wulf 2021, 25-26), um bspw. Figuren zu charakterisieren. *INVADERS MUST DIE!* in HFR ist allerdings klar ironisch und die *ludoaudiovisuellen Harmonien* des Kampfsystems werden hier Teil eben jener ironischer Bedeutungsgebung, wenn der Spieler im Rhythmus der Musik genau das Gegenteilige bezweckt als der Gesang des Liedes es fordert. Der Spieler besiegt durch synchrone Attacken zum Rhythmus des Liedes alle Robotergegner und ermächtigt sich so durch *ludoaudiovisuelle Harmonien* basierend auf Affordanzen quasi der Musik selbst und richtet diese scheinbar gegen den im Lied zu hörenden singenden Roboter.



Abbildung 16: 1 Roboter beim Essen in der Mensa; 2 Kampf gegen Roboter und Einblendung des Musiktitels

Weitere äußerst detaillierte Einblicke in die auf Rhythmus basierende Spielmechanik HFRs und deren Entwicklung bietet eine Präsentation des Designers John Johanas bei der Game Developers Conference 2024. Johanas spricht hier unter anderem davon, dass die gesamte Entwicklung des Spiels rückwärts gerichtet von im Gameplay vorkommenden „impact“ Momenten aus ging. Mit „impact“ betitelt Johanas im Kern genau das Prinzip der *ludoaudiovisuellen Harmonie*, bei dem Elemente von Spielaspekten wie Spielerhandlungen, visuellen Effekten oder Vibrationsfeedback mit Elementen der Musik synchronisiert sein sollen:

„So in my mind this made sense, having the impact land with the music as it would basically make the impact feel stronger. But not only that [...] having [...] the beat interact with the impact as well would also reinforce synchronization with the music. And honestly I just thought this would feel really good in general. [...] The core idea was that you were the star of a music video and if you look at any music video or movie or trailer that’s edited to the music or does something similar [...] they al-

ways do where the hit of the action lands with the music, giving it that extra power.“ (Johanas 2024, 0:10:29)

### 5.2.3 Diskussion: Rhythmus als Fundament für Gamedesign

Wie Bakels bereits in seiner Arbeit zum *audiovisuellen Rhythmus* festgestellt hat, besitzen Filme in der Regel kein an Metrum oder zeitliche Maßeinheiten gebundenes Verständnis von Rhythmus (Bakels 2016, 96-97), sondern nutzen „dynamische Temporelationen [aller] Akzentuierungen im audiovisuellen Wahrnehmungsstrom“ (ebd. 100). Einige Computerspiele jedoch, allen voran Rhythmusspiele, funktionieren nur durch ein an Zeit gebundenes musiktheoretisches Rhythmusverständnis, bei dem in Synchronizität zum Rhythmus passende Befehle des Spielers ausgeführt werden sollen: „Rhythm-action games are video games in which the player must respond in some way to the rhythm or melody being presented [...]“ (Collins 2008, 74). Die Musik wird hier wortwörtlich zum Taktgeber des gesamten Spielprinzips- und designs, was in der Untersuchung zu MH und HFR deutlich wurde. Entscheidende Affordanzen des Spielers sollen, auch durch Verbindungen zur Zielhierarchie der Spiele und insbesondere den kontingenten Zielen, an den Rhythmus der Musik gekoppelt werden (Tab. 3 und 4). Ein Spielen mit ausgeschalteter Musik wäre fast unvorstellbar, auch wenn beide Spiele visuelle Hilfen bieten, um den Rhythmus der Musik zu erkennen. So besteht das Fadenkreuz in MH in der Mitte des Bildschirms aus jeweils drei Pfeilspitzen auf linker und rechter Seite, die horizontal auf die Mitte zulaufen und von dem die mittigen Pfeilspitzen synchron zum Rhythmus der Musik aufleuchten (Abb. 14.1), was sehr an traditionelle Rhythmusspiele wie DANCE DANCE REVOLUTION erinnert. Auch HFR bietet eine visuelle Hilfe in Form einer Roboterkatze, die sich zu einer Kugel formt, ständig neben dem Avatar her schwebt und vertikale oder horizontale Linien<sup>37</sup> synchron zum Rhythmus aussendet (Abb. 15.4). Auch die vielen bereits erwähnten Objekte in der Spielwelt HFRs, Bäume, Fließbänder, NPCs etc., erfüllen diese Funktion. MH bietet hier nur in den Leveln verteilte Flammen, die synchron zum Rhythmus in die Höhe schießen (Abb. 14.4), und Powerup-Objekte, die sich zum Rhythmus der Musik drehen. Solche visuellen Hilfen sind sicherlich wichtig, da Entwickler nicht von allen Spielern ein gut geschultes Rhythmusgefühl erwarten können. Die Vermittlung des Rhythmus wird damit in beiden Spielen audiovisuell vermittelt und dient als Voraussetzung für *ludoaudiovisuelle Harmonie*, hervorgebracht durch den Akt des Spielens.

Der Rhythmus der Musik in MH und HFR kann daher auch als ein Muster für das Gameplay verstanden werden. Erst mit der Annahme und Ausführung dieses Musters kann der Spieler nicht nur erfolgreich sein, sondern über-

---

<sup>37</sup> Das Optionsmeü in HFR bietet hier drei unterschiedliche Möglichkeiten für die visuelle Hilfe, was zeigt wie wichtig Diese und der Rhythmus allgemein für das Spielprinzip sind. In MH lassen sich die Farben der visuellen Hilfe anpassen.

hauptsächlich erst handeln, da wichtige Affordanzen nur in Synchronizität mit dem Rhythmus Wirkung zeigen und die nötige Interaktivität mit der Spielwelt hier teilweise vom Rhythmus und *ludoaudiovisuellen Harmonien* abhängt. Rhythmusspiele nutzen damit eine der wohl ältesten Funktionen der Musik, denn Musik lenkt die Konzentration des Hörers auf Muster, die bestimmte Gefühlslagen hervorrufen und als Basis für Handlungen dienen sollen (Csikszentmihalyi 1990, 108-109). Die Auswahl des Metal- bzw. Rockgenres in MH und HFR ist nicht zufällig. Sicherlich passen diese Genre auch thematisch sehr gut zum Setting der Spiele, Kämpfe in der Hölle in MH und Rebellion gegen ein gigantisches Technikunternehmen in HFR, doch sie besitzen vor allem das nötige Tempo, um das schnelle und fesselnde Gameplay der Spiele zu ermöglichen. Alle in den Spielen vorkommenden Lieder, die parallel zum Gameplay laufen, haben ein Tempo von über 100 BPM (beats per minute). In MH wird die BPM-Anzahl sogar im Musikauswahlmenü angezeigt. Der höchste Wert liegt in HFR laut des Designers des Spiels, John Johanas, sogar bei 160BPM (Wilhelm 2023). Das Tempo von Musik kann physiologische Auswirkungen auf den Menschen haben, denn 72 BPM gilt als der durchschnittliche Herzschlag, alles darunter als beruhigend und alles darüber als erregend (Bakels 2016, 28). Die Musik in MH und HFR ist daher definitiv als erregend einzustufen und der Rhythmus somit ein Muster, dass den Spieler zu hoher Konzentration und schnellen Handlungen antreibt.

Bei so hohen BPM-Werten muss darauf verwiesen werden, dass die *ludoaudiovisuellen Harmonien*, ausgelöst durch Gameplay, nicht zu jedem Beat der Musik vorkommen sollen, da dies viel zu schnelle Eingaben des Spielers erfordern würde. Stattdessen basieren die *ludoaudiovisuellen Harmonien* in erster Linie, trotz der engen Verbindung zum Rhythmus der Musik, auf der Spielstruktur bzw. den spezifischen Spielstaten. Der Spieler muss immer noch die Spielwelt, Gegner und seine eigene Position berücksichtigen, bevor Affordanzen wie das Schießen in MH synchron zum Rhythmus ausgeführt werden und genau hier liegt die Schwierigkeit aber auch der Spielspaß. Allerdings kann die Musik nicht vernachlässigt oder gar ausgeblendet werden, sondern muss in die Überlegungen des Spielers aufgenommen werden. Im Sinne Csikszentmihalyis liegt demnach in beiden Spielen ein analytischer Hörmodus vor. Die Musik muss hinsichtlich strukturaler Merkmale wie dem Rhythmus analytisch gehört werden (Csikszentmihalyi 1990, 111), um so die eigenen Handlungen synchron auf den Rhythmus anzupassen, wodurch *ludoaudiovisuelle Harmonien* entstehen.

## 5.3 Synästhesie als Designziel

### 5.3.1 Rez

REZ<sup>38</sup> (Japan 2001, Sega, Dreamcast) ist ein klassischer On-Rail-Shooter<sup>39</sup> von Spieledesigner Tetsuya Mizuguchi, bei dem der Spieler einen durch verschiedenste 3D-Level fliegenden Avatar begleitet und dessen Ziel- und Schießfunktion steuert, um verschiedene Gegnertypen zu besiegen. Die Musik des Spiels ist eine Mischung aus eigens komponierter und lizenzierter Electronic Dance Music (EDM), die durch das Herunterbrechen bis auf einzelne Noten geloopt und gelayert wurde, um so dynamisch mit der Spielstruktur verbunden zu werden (Hermida 2002). Visuell besticht das Spiel durch seine besondere Wireframe-Grafik, die wohl bewusst an die ersten 3D-Grafiken erinnert, da es in der Narration des Spiels um einen Hacker geht, der in Computersysteme eingreift.

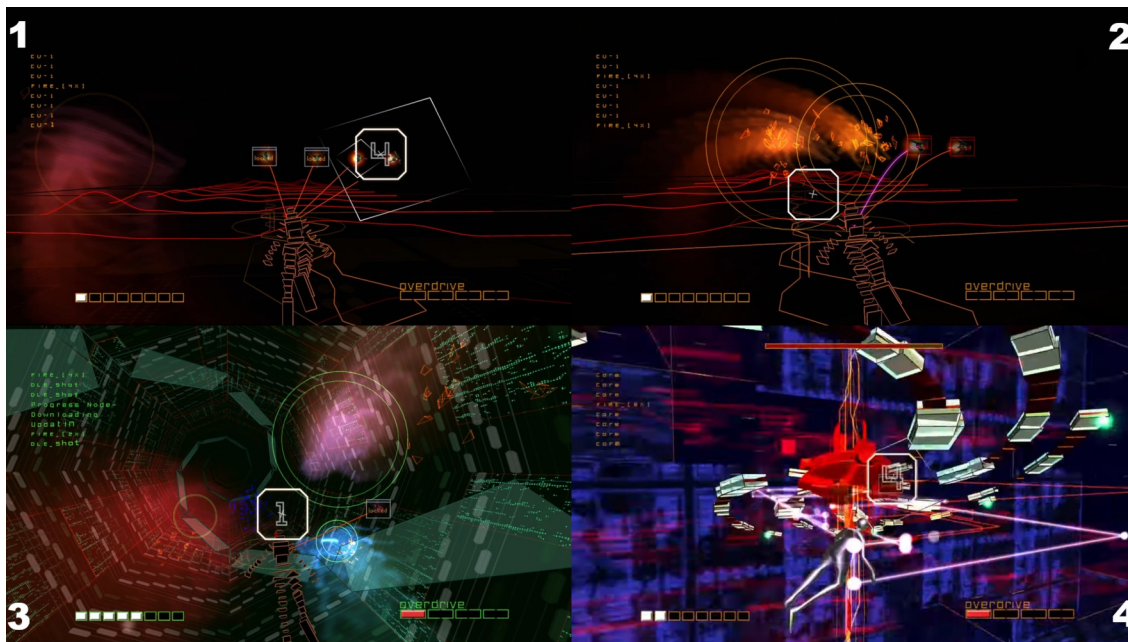


Abbildung 17: 1 rotierende Quadrate beim Anvisieren; 2 farbige Kreise nach Abschuss; 3 Vielzahl an Effekten; 4 Bossgegner

Die Musik ist aufgeteilt in zwei von einander unabhängige Ebenen. Die erste Ebene ist die lineare Hintergrundmusik, die aus dem EDM-Soundtrack besteht und mit der Segmentierung des Gameplays zusammenhängt. So hat

38 Für diese Untersuchung wurde auf das Spiel REZ INFINITE (2017, Enhance, Windows PC) zurückgegriffen, das das Originalspiel beinhaltet und sich lediglich durch höhere Auflösungen vom Original unterscheidet.

39 Das Genre ist dadurch gekennzeichnet, dass die Bewegung des Avatars gar nicht oder nur sehr eingeschränkt möglich ist, daher die Bezeichnung "On-Rail", also auf Schienen. Manche Genrevertreter wie STAR FOX 64 (Japan 1997, Nintendo, Nintendo 64) erlauben allerdings auch das Auswählen unterschiedlicher Pfade. Im Vordergrund steht jedoch immer das Abschießen von Gegnern.

jedes Level ein eigenes Lied, wodurch die Musik hier Teil der spatialen Segmentierung ist. Die Level selbst sind zwar in Abschnitte, genannt Level Layer, unterteilt, allerdings handelt es sich hier um challenge Segmentierung und nicht spatiale Checkpoints, denn der Avatar ist nicht frei bewegbar und die Level Layer können nur verlassen werden, wenn alle darin vorkommenden Gegnerwellen überstanden sind (Zagal et al. 2008, 187). Die Lieder der Level ändern sich fließend von Level Layer zu Level Layer und werden in der Regel parallel zur Reise des Avatars intensiver. Bei den Bossgegnern ändert sich die Musik erneut. Es wird kein anderes Lied abgespielt, sondern quasi eine andere Version des gleichen Liedes geloopt bis der Spieler den Boss besiegt hat. Zu erkennen ist dies auch am fortlaufenden Rhythmus der Musik und dem synchronen Vibrieren des Controllers, das über die Level Layer hinweg ohne Unterbrechung weiter geführt wird. Der Rhythmus der Musik wird hier primär nicht über das Akustische, sondern über das Haptische vermittelt und ist somit im Sinne Chions transsensorisch, also kann über mehr als nur einen Sinn erfasst werden (Chion 1994, 136-137).

Die zweite Ebene der musikalischen Untermalung in REZ ist ein generatives Musiksystem, mit dem das Spiel die Simulation von Synästhesie<sup>40</sup> zu erreichen versucht, was Mizuguchis Designziel für REZ war (Hermida 2002). Diese generative Ebene setzt sich aber nicht aus einem vom Computer berechneten Algorithmus zusammen (van Geelen 2008, 96), sondern sie wird durch Affordanzen des Spielers erzeugt, bzw. durch davon beeinflusste Objekte in der Spielwelt wie Gegner. Diese Spielerhandlungen resultieren in musikalischen Töne, zu denen wiederum synchron visuelle Effekte und Vibrationen ausgelöst werden. So ertönt beim Drücken des A-Knopfes zum Anvisieren der Gegner ein Klatschgeräusch. Mit gedrückt gehaltenem A-Knopf werden dann per linkem Analog-Stick Gegner anvisiert, was Perkussionstöne wiedergibt. Gleichzeitig erscheinen bei jedem anvisierten Gegner Quadrate, die sich rotierend auf diesen hinzubewegen und verkleinern (Abb. 17.1). Mit Loslassen des A-Knopfes werden dann alle anvisierten Gegner abgeschossen, was visuell vom Zentrum der Gegner ausgehende farbige<sup>41</sup> Kreise erscheinen lässt, sowie einen leichten gleichfarbigen Nebелеffekt hervorruft (Abb. 17.2). Synchron zum Abschießen von Gegnern erfolgt ein Vibrationseffekt und es ertönen je nach Gegnertyp unterschiedliche Synthesizertöne. Abgeschossene Projektile von Gegnern hingegen erzeugen Perkussionstöne. Es können maximal acht Gegner pro Spielerangriff anvisiert werden. Je mehr Gegner anvisiert und abgeschossen werden, desto länger sind mögliche Melodien, die durch die Töne auf der generativen Musikebene erzeugt werden. Dies ist sehr leicht erkennbar, wenn man per Optionsmenü

40 „Das Wort Synästhesie ist abgeleitet von den altgriechischen Wörtern syn (= zusammen) und aisthesis (= Empfinden), [...] die gleichzeitige Miterregung eines Sinnesorgans bei Reizung eines anderen. Synästhesie beruht auf einem zusätzlichen Kanal der Wahrnehmung. Manche Synästhetiker können Buchstaben fühlen oder Worte schmecken. Andere können Laute, d.h. Töne, Geräusche oder auch Sprache [...] in Farben wahrnehmen (auch → „Farbenhören“ genannt.“ (Synaesthesia.org 2024)

41 Die Farbe der Effekte hängt vom jeweiligen Level ab.

das Spiel ohne die lineare Musikebene spielt. Allein die durch Spielerhandlungen hervorgerufenen Toneffekte erzeugen eine recht vollständig klingende Musik (EDM), die das Spielgeschehen synchron begleitet.

Das recht simple Gameplay des Spiels ist somit geprägt von einer Vielzahl an *ludoaudiovisuellen Harmonien*, bestehend aus akustischen, visuellen und haptischen Modalitäten, die zusammengenommen ein multimodales Musikerlebnis gestalten, das aber erst durch die generative Musikebene und die damit verbundenen Affordanzen ermöglicht wird. Töne, Farben, Formen, Bewegungen und haptisches Fühlen werden hier zu *ludoaudiovisuellen Harmonien* verbunden und synchron für den Spieler erfahrbar gemacht, egal ob dieser eine spezielle Verbindung von Gehirnarealen, also Synästhesie, besitzt oder nicht: „COLOR [sic], brightness, shape, movement, beats, notes, sounds, even the pulsing vibrations of the controller in hand come together in *Rez* in a veritable synesthesia, a blending of the senses intent on melding the gamer with the game.“ (Wark 2007, 128, kursiv i.O.). Daniels und Naumann halten Synästhesie sogar für ein grundsätzliches Bedürfnis des Menschen und finden dieses in zahlreichen kulturellen Praxen in der Geschichte vor wie Fackeltänze, die Kombination von Orgelmusik und Lichteinfall in gotischen Kirchen oder auch „visual performances in techno club[s]“ (Daniels/Naumann 2013, 218), dessen persönliche Besuche in Europa eben auch Inspiration für Mizuguchis *REZ* waren (Hermida 2002). Ähnlich wie bei der Kunst Wassily Kandinskys, eine weitere Inspirationsquelle Mizuguchis (Wark 2007, 146), sollen Farben und Formen als Musik gehört werden bzw. aus der Musik sollen farbliche Formen entstehen (hamburger-kunsthalle.de 2024). Die *ludoaudiovisuellen Harmonien* in Zusammenhang mit dem generativen Musiksystem ermöglichen dies in einer interaktiven Form und überwältigen den Spieler gewissermaßen mit einer Kombination aus „psychedelic visuals and pulsating dance beats“ (Post 2006, 216), wodurch eine Vielzahl an jederzeit simulierten synästhetischen Erfahrungen hervorgerufen wird (Abb. 17.3).

Ein Blick auf die Zielhierarchie des Spiels (Tab. 5) zeigt eine, vor allem im Vergleich zu den anderen hier untersuchten Spielen, recht eindeutige hierarchische Aufteilung der linearen und generativen Musikebenen. So ist die lineare Musikebene nur mit zwei in der Zielhierarchie hohen imperativen Zielen verbunden, *Reach (Boss)* und *Reach (Level/Level Layer)*, die eng mit der beschriebenen challenge Segmentierung des Gameplays zusammenhängen. Die generative Musikebene hingegen ist nur mit weiter unten stehenden Zielen verbunden: dem imperativen Ziel *Remove (Gegner)*, sowie den beiden kontingenten Zielen, *Synchronize (Fadenkreuz, Gegner, Schuss)* und *Synchronize (Fadenkreuz, Gegnerangriff, Schuss)*, die das Gameplay selbst, also die ausführbaren Affordanzen umfassen.



Tabelle 5: Zielhierarchie Rez

### 5.3.2 Tetris Effect

Zwar gab es seit Erscheinen des ersten TETRIS (Sowjetunion 1985, Alexey Pajitnov, Elektronika 60) eine Vielzahl von Neuauflagen, die mitunter auch Neues wagten<sup>42</sup>, Tetsuya Mizuguchis TETRIS EFFECT (nachfolgend TE) sticht jedoch aus allen Tetris-Variationen hervor, weil es eine meines Wissens nach bis dahin unerreichte Verbindung von Musik und Computerspiel erreicht, die ich im Folgenden aufzeigen werde. Der Spieler muss zufällig fallende Puzzleteile, Tetraminos genannt, so ordnen, dass sich aus ihnen volle Reihen bilden, die sich dann auflösen und Punkte einbringen. Erreichen die gestapelten Blocks jedoch den oberen Rand des Puzzlerasters ist das Spiel vorbei. TE hat jedoch eine vor allem für ein Puzzlespiel einmalige audiovisuelle und haptische Darstellung samt Musik, die bis ins letzte Detail mit dem Gameplay und dem Spielgeschehen verbunden ist.

Die Musik des Spiels besteht aus einer Vielzahl von Eigenproduktionen und ist ebenso wie bereits in Rez in eine lineare Hintergrundmusikebene und eine generative Musikebene aufgeteilt, die sich beim und durch das Spielen ergänzen. Im Haupteinzelspielermodus „Journey“ liegt zunächst eine spatia-

<sup>42</sup> Beispielsweise versuchte Nintendo mit 3D TETRIS (Japan 1996, Nintendo, Virtual Boy) den stereoskopischen 3D Effekt der Virtual Boy Konsole zu nutzen, um auch das Spielprinzip in den drei-dimensionalen Raum zu übertragen.

le Segmentierung vor, da es 27 Level gibt, die jeweils ein eigenes Lied als lineare Hintergrundmusik aufweisen. Auch die generative Musikebene ist in jedem Level individuell. Innerhalb der Level gibt es zudem eine challenge Segmentierung, die mit beiden Musikebenen verbunden ist. Ziel in den Levels ist es 36 Reihen von Blöcken aufzulösen, um das Level abzuschließen<sup>43</sup>. Beide Musikebenen ändern sich sobald der Spieler ein und zwei Drittel dieses quantitativen Ziels erreicht hat, also bei 12 und 24 aufgelösten Reihen. Insgesamt hat jedes Level also drei Stufen der linearen Musikebene und drei Stufen der generativen Musikebene (0-11, 12-23, 24-35). Beim Erreichen dieser Schwellen nimmt die Fallgeschwindigkeit der Tetraminos i.d.R. zu und parallel dazu wird die lineare Musik intensiver und schneller, was zu dem schnellerem Gameplay passt und vergleichbar mit der Nutzung von schneller Musik in Filmen ist: "These finding are consistent with research on the perception of pacing of film shots along with different songs, where faster tempo music is consistently influential on creating a perception of faster paced images" (Wharton/Collins 2011). Gleichzeitig kommt es durch die Stufen auch bei der generativen Musikebene zu Veränderungen der musikalischen Tönen, die für verschiedenste Affordanzen abgespielt werden. Hier gibt es daher bereits die erste Verbindung der Musik mit imperativen Zielen des Spiels (Tab. 6). Das erste imperative Ziel *Optimize (Anzahl Reihen)* ist an die challenge Segmentierung innerhalb der Level geknüpft und damit auch an die Veränderung der Stufen beider Musikebenen. Die Verbindungen der Affordanzen mit der Musik, die wiederum mit der audiovisuellen und haptischen Darstellung des Spiels zusammen hängen, machen TE so besonders. Jede performative<sup>44</sup> Affordanz des Spielers hat einen Ton zur Folge. Es handelt sich um Töne, die nicht per se musikalisch sind, aber im Setting des Levels Sinn ergeben und durch den Kontext der generativen Musikebene musikalisch interpretierbar sind, denn „[...] the distinction between music and noise is completely relative, and has to do with what we are listening for“ (Chion 1994, 206). Wie bereits angesprochen hat jedes Level eine eigene lineare und generative Musikebene. Auch die visuelle Darstellung ist von Level zu Level unterschiedlich und hängt thematisch mit den Musikebenen zusammen. Folgend wird das erste Level des Journey-Modus genauer beleuchtet und die Zusammenhänge zwischen Musik, Darstellung und Gameplay, sowie die daraus entstehenden *ludoaudiovisuellen Harmonien* im Detail betrachtet.

Das erste Level „The Deep“ steht zu Anfang im Zeichen einer Unterwasserwelt. Generell ist die Spielwelt wie in allen anderen Levels auch in Vorder- und Hintergrund aufgeteilt. Im Vordergrund steht das für Tetris typische Puzzleraster, indem das eigentliche Spielgeschehen stattfindet. Dies wird zu

43 Die Zahl variiert leicht je nach ausgewähltem Schwierigkeitsgrad. Hier wird vom "normalen" Schwierigkeitsgrad ausgegangen.

44 Linderoth unterscheidet zwischen exploratorischen Affordanzen, mit denen Menschen erörtern welche Handlungsmöglichkeiten sie haben, die wiederum performative Affordanzen genannt werden (Linderoth 2011, 5).

den Seiten noch durch eine Art Heads Up Display (HUD)<sup>45</sup> leicht erweitert, das die folgenden Tetraminos, sowie gespielte Zeit, Punktezahl oder Anzahl aufgelöster Reihen anzeigt. Der Hintergrund zeigt das Thema des Levels und scheint auf den ersten Blick auch nichts weiter zu sein als das: ein Hintergrund, der mit dem Puzzleraster im Vordergrund und dem Spielgeschehen nicht zusammenhängt.

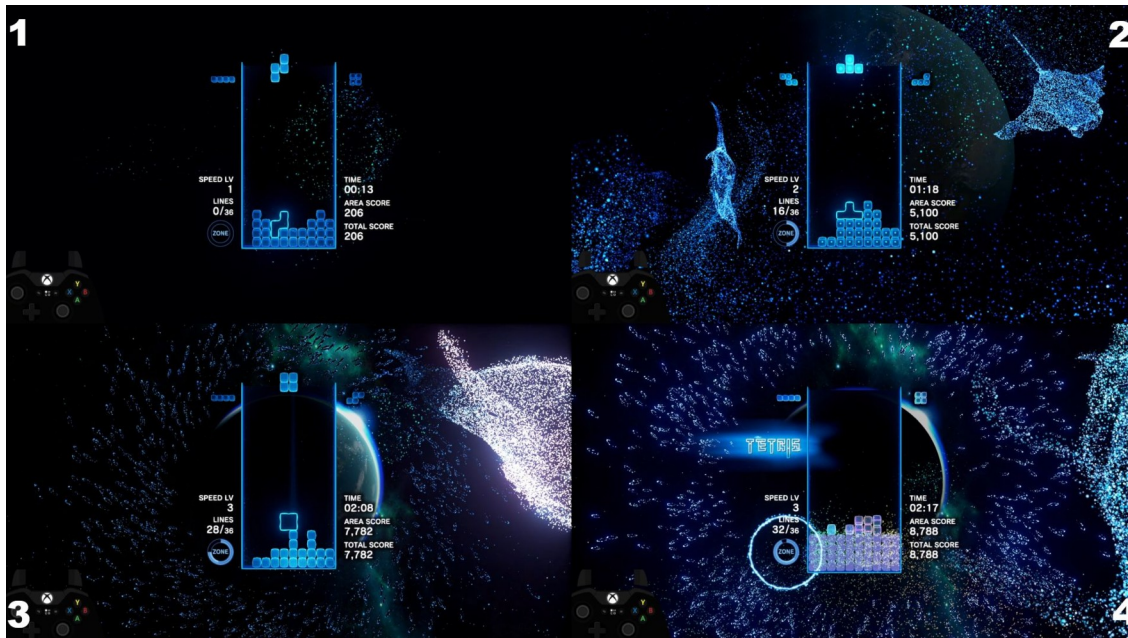


Abbildung 18: 1 Partikelwirbel; 2 Meerestiere und Aufleuchten von Punkten innerhalb der Blocks; 3 Leuchten Planet Erde; 4 Partikeleffekt bei Vervollständigung von Reihen

Allerdings gibt es mehrere Verbindungen des Hintergrunds mit beiden Musikebenen, sowie der Darstellung des Vordergrunds. Beim Drehen eines Tetraminos kommt es nicht nur zu einem musikalischen Ton auf der generativen Musikebene (Ausruf der Sängerin und Tropfgeräusch in den ersten beiden Stufen, sowie Pianotöne auf der letzten Stufe), es entsteht zeitgleich zur Aktion des Spielers auch ein Partikelwirbel im Hintergrund (Abb. 18.1). Somit sind die Affordanz des Tetraminodrehens, einer von mehreren musikalischen Tönen und der visuelle Effekt im Hintergrund in einem Moment des Spielgeschehens miteinander verbunden und es entsteht eine *ludoaudiovisuelle Harmonie*. Auch auf der linearen Musikebene gibt es solche Verbindungen. So leuchten während der ersten Stufe der Musik Felsen im Hintergrund synchron zu Streichern, die im Takt der Musik spielen. Ab Stufe zwei verschwindet der Felsen und es gibt stattdessen ein Aufleuchten von Punkten innerhalb der Blöcke im Puzzleraster synchron zu einem Perkussionston im Takt der Musik (Abb. 18.2) und direkt danach ein aufsteigendes Leuchten einiger Blöcke im Puzzleraster. Der Hintergrund während Stufe

<sup>45</sup> Das HUD ist ein Teilbereich des Bildschirms, der für das Spiel relevante Informationen anzeigt, bspw. gesammelte Punkte oder eine Karte der Spielwelt. Nur selten sind HUD-Elemente diegetisch wie in METROID PRIME (Japan 2002, Nintendo, Gamecube).

zwei ist gekennzeichnet durch Meerestiere, zusammengesetzt aus Partikeln, die durch die Spielwelt tauchen und synchron zum Takt der Musik aufleuchten (Abb. 18.2). In der dritten und letzten Stufe ändern sich erneut Musik und Darstellung. Zusätzlich zum Leuchteffekt von Stufe zwei erfolgt noch ein Aufleuchten und Drehen von Partikeln im Hintergrund der Spielwelt und der direkt hinter dem Puzzleraster aufgetauchte Planet Erde erfährt ein leichtes Leuchten synchron zum Takt der Musik (Abb. 18.3). Je nach Stufe entsteht also auch auf der linearen Musikebene durch die Verbindung dieser mit anderen Spielaspekten wie der Darstellung des Vorder- und Hintergrunds eine *ludoaudiovisuelle Harmonie*, die ab der zweiten Stufe (also ab zwölf aufgelösten Reihen) auch noch durch einen Vibrationseffekt synchron zum Takt der Musik erweitert wird. Wie bei REZ bewirken die *ludoaudiovisuellen Harmonien* durch ihre Multimodalität und Echtzeitlichkeit also eine beachtlich ausgeprägte Simulation von Synästhesie, die hier in schriftlicher Form nur sehr schwierig zu beschreiben ist.

Zielhierarchie Tetris Effect	
Complete	Legende: Gelb = ultimates Ziel Orange = imperatives Ziel Blau = kontingentes Ziel ⇒ = zu vorh. imperativem Ziel Fett = Verbindung lineare Musik Kursiv = Verbindung generative Musik
Optimize (Anzahl Reihen 36/36)	
Configure (Reihen, Blocks)	
Reach (Tetraminos, Reihen)	
⇒ Synchronize (Affordanzen, Tetraminos, Puzzleraster)	

Tabelle 6: Zielhierarchie TETRIS EFFECT

Auch die ludologischen Ziele weisen Verbindungen zur Musik auf und helfen die *ludoaudiovisuellen Harmonien* zu fassen. Das zweite imperative Ziel (Tab. 6) ist *Configure (Reihen, Blocks)* und beschreibt das Formieren von vollen Reihen. Sobald eine Reihe vervollständigt wird, erfolgt ein Partikeleffekt (Abb. 18.4) und auf der generativen Musikebene ist einer von mehreren Tönen zu hören: Walgesang, Synthesizer, Glocken oder Streicher. Auch hier wird durch die direkte Verbindung mit der Musik eine *ludoaudiovisuelle Harmonie* erzeugt, die durch ein Vibrieren des Controllers noch haptisch erweitert wird. Das unterste imperative Ziel kann mit *Reach (Tetraminos, Reihen)* beschrieben werden und beschreibt das Einordnen der Tetraminos in Reihen. Daraus ergibt sich das komplexe kontingente Ziel *Synchronize (Affordanzen, Tetraminos, Puzzleraster)*, welches letztendlich alle performativen Affordanzen, die in einem konkreten Spielmoment dazu genutzt werden können ein Tetramino so zu platzieren, dass es für das Vervollständigen von Reihen nützlich ist, zusammenfasst. Alle diese Affordanzen haben eine Verbindung zur generativen Musikebene. Wie bereits beschrieben erzeugt das Drehen eines Tetraminos einen Ausruf der Sängerin und ein Tropfgeräusch

in den ersten beiden Stufen, sowie Pianotöne auf der letzten Stufe. Wird ein Tetramino durch Eingabe mit dem Steuerkreuz des Controllers nach links oder rechts bewegt, ist zeitgleich auch ein kurzer Ausruf der Sängerin zu hören bzw. auf letzter Stufe ein Perkussionston. Die Hold Que-Funktion, mit dem ein Tetramino gegen das nächste Tetramino ausgetauscht werden kann, hat die gleichen akustischen und visuellen Effekte wie das Bewegen von Tetraminos. Selbst das bloße Aufkommen eines Tetraminos, das auch durch Eingabe anhand des Steuerkreuzes vom Spieler ausgelöst werden kann, gibt einen dumpfen Perkussionston ab, der visuell durch einen minimalen Partikeleffekt unterstützt wird. Bei so vielen direkten Verbindungen der Affordanzen und Ziele mit musikalischen oder zumindest thematisch passenden Tönen auf der generativen Musikebene, wird allein durch den Akt des Spielens eine für sich theoretisch auch alleinstehende Musik erzeugt, die im Vergleich zur generativen Musikebene in REZ aber weitaus komplexer ist. REZ und TEs generative Musikebene übernehmen die objektorientierte Logik von Spielen, die bereits in Kapitel 4.3.1 besprochen wurde, und erzeugen eine höchst dynamische und adaptive Musikuntermalung: „Deploying small units of object-associated music accommodates the non-linear game structures.“ (Summers 2016, 161).

Nimmt man all diese Erkenntnisse zusammen, wird klar, dass der Verlauf der Musik in TE, bestehend aus den beiden sich ergänzenden Musikebenen, synchron und parallel zur Entwicklung des Gameplays verläuft, denn wenn ein System für generative Musik Veränderungen im Spiel und Eingaben des Spielers nutzt, dann kann das System so immer synchron zum Gameplay des Spielers sein (Sites/Potter 2018). Die challenge Segmentierung in den jeweiligen Leveln, die dazu führt, dass beide Musikebenen sich durch das Erreichen von bestimmten Schwellen (12 und 24 aufgelöste Reihen) verändern und steigern, führt auch zu einer Erhöhung des Schwierigkeitsgrades innerhalb eines Levels, da die Schwellwerte auch die Fallgeschwindigkeit der Tetraminos erhöhen. So ist der Anfang eines Levels i.d.R. durch einen erforschenden Aspekt des Gameplays geprägt (Linderoth 2011, 5), bei dem der Spieler genügend Zeit hat sich auf die beste Platzierung der Tetraminos zu konzentrieren. Die Musik ist dementsprechend auch vergleichsweise ruhig und in einigen Leveln ist die erste Stufe der linearen Musikebene sogar nur minimal wahrnehmbar. Später jedoch steigt die Fallgeschwindigkeit mit der zweiten und dritten Stufe, wodurch die letztendliche Platzierung der Tetraminos eher nach performativem Aspekt erfolgt und ein längeres Überlegen hier nicht mehr möglich ist. Die gestiegene Fülle der Musik spiegelt dies nicht nur sehr gut wieder, sondern kann bei dem ein oder anderen Spieler den Stressfaktor eventuell sogar noch erhöhen. Auch die jeweiligen *ludoaudiovisuellen Harmonien* werden im Verlauf der challenge Segmentierung sensorisch stärker, da wie gezeigt visuelle Effekte zunehmen und akustische Elemente in höherer Quantität vorkommen. Dies ist für die akustischen Effekte auch leicht erkennbar, wenn man sich eine parallele Aufzeichnung der

Wellenform der Tonausgabe des Spiels anschaut. Anhand der in Audacity aufgezeichneten Wellenform des Spielverlaufs ist auch eine stetige Zunahme der Tonkulisse, bestehend aus linearer und generativer Musikebene, erkennbar (Abb. 19). Je höher die Stufe, desto gefüllter ist die Wellenform der beiden Musikebenen.

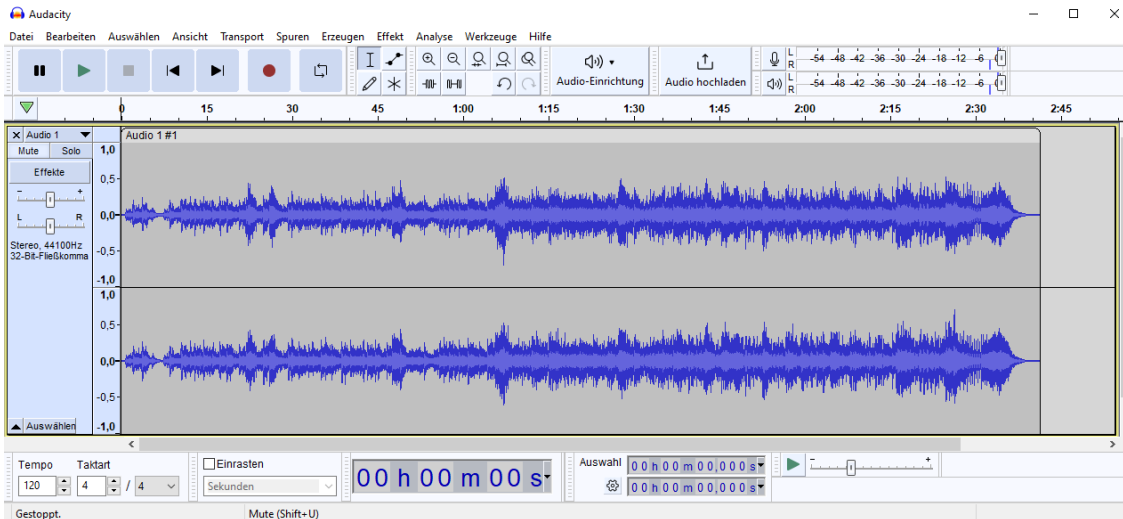


Abbildung 19: Wellenform des Tons einer Spielsitzung des Levels „The Deep“ in *TETRIS EFFECT*

### 5.3.3 Diskussion: Synästhesie als Designziel

Sowohl REZ als auch TE schaffen es vor allem durch ihre generativen Musikebenen zahlreiche *ludoaudiovisuelle Harmonien* hervorzubringen, die in eine beeindruckende und teilweise sogar überwältigende Simulation von Synästhesie münden. Die enge Verknüpfung von Spielaspekten wie ludologischen Zielen, Gameplay-Segmentierung und Affordanzen mit dem generativen Musiksistem erzeugt einen für jede Spielsitzung individuellen Soundtrack. Speziell in TE ist jede Handlung des Spielers mit einer Tonausgabe verbunden, die wiederum visuelle und haptische Effekte auslöst.

Doch kann das Spielen von REZ und TE dank ihrer generativen Musikebenen als eine Form von Musizieren angesehen werden? Immerhin spricht Wark davon, dass in REZ der Controller das Piano ersetzen würde (Wark 2007, 146), und auch Svec beschreibt das Spielen von REZ im Vergleich zu Musikspielen wie *GUITAR HERO* als virtuos (Svec 2008, 8). Bereits die Fremdnutzung des Plattenspielers in Remix- und Hip-Hop-Kultur wird mittlerweile von der Musikwissenschaft als eine Form des Musizierens und der Plattenspieler somit als Musikinstrument anerkannt (Claussen 2021, 65). Zum einen gibt es nicht nur in REZ und TE sondern allgemein in allen vergleichbaren Spielen aber das Problem, dass es im Vergleich zu „echten“ Instrumenten ein gewisses Ungleichgewicht zwischen Input und Output gibt: „Mangelnde Anerkennung erfahren musikalische Performances immer dann,

wenn die Bedienung des Instruments vergleichsweise trivial im Vergleich zur Komplexität seines klanglichen Ergebnisses ist.“ (ebd. 66). Durch wenige, simple Tastendrücke auf dem Controller kann der Spieler dank der generativen Musiksysteme in REZ und TE bereits eine vielschichtige Musikwiedergabe auslösen. Zu bedenken ist aber, dass die Eingaben des Spielers nicht auf Grund musikalischer Kompositionsabsichten erfolgen. Stattdessen handelt der Spieler nach der Logik des Spiels und versucht dieses zu meistern, was ebenso das Problem in Musikspielen ist, wo nur der korrekte Input Musikausgabe zur Folge hat (Svec 2008, 6). Die musikalische Ausgabe in REZ und TE ist ja nur eine Folge des auch ohne speziellem generativem Musiksystems vorhandenen Gameplays<sup>46</sup>. Beide Spiele bieten jedoch gesonderte Modi, in denen eben jene Spiellogik mit ludologischen Zielen, Gegnern und Punktezahlen ausgeblendet wird. REZ Level lassen sich nämlich auch so spielen, dass der Avatar keinen Schaden nehmen kann und ein Scheitern somit nicht mehr möglich ist. Das Abschießen der Gegner und Bosse (Abb. 17.4) ist aber weiterhin nötig um voranzuschreiten, doch hier kann der Spieler seine Affordanzen freier und in Hinblick auf eine musikalisch, kompositorische Sinnhaftigkeit ausführen. Die *ludoaudiovisuellen Harmonien* und Simulation von Synästhesie bleiben dementsprechend auch weiterhin vorhanden. TE geht mit dem Theatre-Mode sogar einen Schritt weiter, denn hier kann sogar das Puzzleraster ausgeblendet werden. Übrig bleiben die Musik, die eben teils durch die von Handlungen ausgelösten Tönen auf der generativen Musikebene stammt, und die damit synchron vorkommenden visuellen und haptischen Effekte, also die *ludoaudiovisuellen Harmonien*.

Dem entsprechend ist es sogar egal ob man diese speziellen oder die normalen Einzelspielermodi in REZ und TE spielt. Im Vordergrund stehen immer die *ludoaudiovisuellen Harmonien* und somit auch das Simulieren von Synästhesie. Beide Spiele weisen daher einen sensorischen Hörmodus auf. Die Musik ist Dreh- und Angelpunkt einer „sensory experience“ (Csikszentmihalyi 1990, 110). Sie hilft nicht wie bei MH und HFR für eine bessere Performance und wird daher nicht analytisch gehört, sondern kann eher im Gegenteil das Spiel schwieriger machen, da die Musik wie gezeigt parallel zum Spielgeschehen an Intensität zunimmt. In Verbindung mit den vielen anderen sensorischen Reizen, wird die Musik hier auch visuell, haptisch und ludisch erlebt, ja mitunter durch die generative Musikebene sogar erst erschaffen. Die vielen direkten Verbindungen der Musik, sowohl auf linearer aber vor allem auf generativer Ebene, mit den sensorischen Reizen, Spielerhandlungen, ludologischen Zielen und Segmentierungen des Gameplays erzeugen etliche *ludoaudiovisuelle Harmonien* in jedem Moment des Spielens und ermöglichen so eine beeindruckende Simulation des Synästhesieprinzips.

---

46 Allerdings gibt es Spieleentwickler wie Yoshiaki Koizumi, die den Akt des Spielens von Computerspielen direkt mit dem Spielen eines Musikinstrument vergleichen und eine allgemeine Musikalität im Computerspiel sehen (Iwata 2010).

## 6 Fazit und Ausblick

Diese Arbeit hat gezeigt, dass Musik in Computerspielen in bestimmten Momenten des Spielens eine Synchronizität mit verschiedensten Spielaspekten eingehen kann, um so bedeutende Verbindungen zwischen Musik und Spiel herzustellen, die ich *ludoaudiovisuelle Harmonien* nenne. Vor allem visuelle Effekte, haptisches Feedback und Spielereingaben wirken so wie weitere Noten der Musik selbst und erweitern deren Harmonie. Dadurch entsteht eine erhöhte Subjektivität und Gegenwärtigkeitsgefühl, sowie eine vertiefte Immersion und Flow beim Spieler. Auch wenn es wie gezeigt mit dem Prinzip der *audiovisuellen Harmonie* ähnliche Phänomene in nicht-interaktiven audiovisuellen Medien gibt, so orientieren sich die synchronen Verbindung von Musik und Spiel nicht primär an Bild und Montage wie in Filmen, Fernsehsendungen und Musikvideos, sondern an der Spielstruktur mit der Segmentierung von Gameplay, ludologischen Zielen und den konkreten Affordanzen des Spielers. Visuelle Effekte als Teil einer *ludoaudiovisuellen Harmonie* sind kausal gesehen meist eine Folge der Verbindungen zwischen Musik und Spielstruktur.

Die Musik ist durch das Prinzip der *ludoaudiovisuellen Harmonie* nicht mehr nur im Hintergrund, sondern wird Teil der Struktur des Computerspiels, das aus Code eine „audiovisuelle[n] Substanz“ (Daniels/Naumann 2013, 234) hervorbringt, die es so zuvor in Massemedien noch nicht gab. *Ludoaudiovisuelle Harmonien* helfen somit bei der Schaffung von interaktiven, immersiven, sensorischen, audiovisuellen Räumen und können daher im besten Fall sogar als Möglichkeit angesehen werden ein „digitales multimedia Gesamtwerk“ zu schaffen (ebd. 233), das den Spieler als Co-Autor hinzuzieht. Wie in anderen Studien nachgewiesen, suchen Spieler genau so eine Art von Verschmelzung von Musik, Visualität, Haptik, Spielstruktur und Gameplay, die als wichtig für die Spielerfahrung angesehen wird (Wharton/Collins 2011). Es kann daher nicht wie von Manchen davon ausgegangen werden, dass Musik per se immer bewusst oder unbewusst wahrgenommen werden sollte (Berndt 2014, 96), wichtig ist, zumindest für das Computerspielmedium, eine direkte Verbindung der Musik zu den anderen Aspekten des Spiels.

Die Untersuchung der Spiele METAL: HELLSINGER und HI-FI RUSH hat zudem gezeigt, dass Rhythmus ein ideales Fundament für die Symbiose von Musik und Gameplay sein kann. Anders als im Film kann Rhythmus in Spielen bzw. Spielrhythmus klaren Zeiteinheiten zugeordnet werden, was eben einem klassischen musiktheoretischen Verständnis von Rhythmus entspricht (Bakels 2016, 38). In beiden Spielen sollen Affordanzen synchron zum Rhythmus der Musik ausgeführt werden, was dann zeitgleich auch ein multimodales Feedback auslöst. Musik bzw. Rhythmus fungieren hier also als Taktgeber bzw. Muster für das gesamte Spielprinzip (vgl. Csikszentmihalyi

1990, 108-109). Auch die Spielwelt wird durch *ludoaudiovisuelle Harmonien* direkt mit der Musik verbunden. Die Spiele werden so lesbar und Musik samt Rhythmus zu einem wichtigen Werkzeug für gute Performanz und das Erreichen von ludologischen Zielen.

Des Weiteren konnte festgestellt werden, dass sich eigens für Spiele komponierte Musik durch generative Musikebenen besser für *ludoaudiovisuelle Harmonien* eignet als Algorithmen, die aus bestehender Musik Spielaspekte generieren. Sowohl BEAT HAZARD 2 als auch AUDIOSURF 2 bieten dem Spieler durch die Möglichkeit bereits bekannte Musik zu nutzen eine dem Musikhören ähnliche Erfahrung von Antizipation und Erfüllung von Kadenzen, die aber nur bedingt gut mit den Affordanzen des Spielers verbunden werden. *Ludoaudiovisuelle Harmonien* beschränken sich so zumeist auf die Verbindung von Musik und multimodaler Darstellung, aber integrieren nur teilweise die Interaktivität bzw. das Gameplay des Computerspielmediums.

Spiele wie REZ und TETRIS EFFECT hingegen schaffen es durch die generativen Musikebenen die Affordanzen des Spielers als Ausgangspunkt für zahlreiche allumfassende multimodale *ludoaudiovisuelle Harmonien* zu nutzen, aus denen nicht nur eine sehr adaptive Musikuntermalung, die fast mit Musizieren gleichzusetzen ist, sondern auch eine teils überwältigende Simulation von Synästhesie hervorgeht.

Trotzdem fungiert Musik in allen Spielen auch als ordnende Entität, um akustische Informationen zuzuordnen und so den Verstand zu organisieren (ebd. 109). Durch die direkten Verbindung zur Spielstruktur dank *ludoaudiovisueller Harmonien* kann der Spieler die Musik aktiv bzw. bewusst hören, um die überwältigende Menge an Informationen während des Spielens zu ordnen. Der Spieler kann erkennen, dass die sensorischen Informationen Ursachen und Regelmäßigkeiten haben und auch Teil des regelbasierten Systems der Spiele sind, wie alle anderen Spielaspekte auch, bspw. Objekte und Charaktere.

Das Prinzip der *ludoaudiovisuellen Harmonie* wirft jedoch auch eine Frage auf, die in weiteren Arbeiten näher untersucht werden sollte: kann bei solch direkten Verbindungen von Musik mit multimodaler Darstellung, Spielstruktur und auch Spielwelten noch von einer klassischen (nicht-)diegetischen Dichotomie ausgegangen werden? Stockburger fasst jegliche Musik als *sound objects*, die nicht-diegetisch sind, da der Spieler die Musik in Optionsmenüs ausschalten kann (Stockburger 2003, 6), vergisst jedoch, dass dies erst seit den 1990er Jahren möglich ist, da Musik zuvor fest im Code programmiert werden musste (Collins 2008, 12). Collins hingegen argumentiert mit dem Beispiel der Kampfmusik in THE LEGEND OF ZELDA: OCARINA OF TIME, dass trotz interaktiver Verbindung der Musik mit der Spielwelt nicht-diegetische Musik vorliegt. Die *ludoaudiovisuellen Harmonien* in den von mir untersuchten Spielen zeigen aber, dass Musik über Rhythmus Objekte in der Spielwelt bewegen bzw. animieren kann (METAL: HELLSINGER und HI-FI RUSH),

Musik durch den Akt des Spielens erst ausgelöst werden kann (REZ und TETRIS EFFECT), die Musik Teile der Spielwelt zum Leben erweckt (TETRIS EFFECT), Musik sogar Affordanzen von Entitäten erst hervorrufen kann (BEAT HAZARD 2) und dass ganze Spielwelten sich aus der Musik generieren (AUDIOSURF 2). Insofern ist wie Huiberts und van Tol schreiben auf Grund der Interaktivität von Computerspielen eher von einem „trans-diegetischen Prozess auszugehen“ (Huiberts/van Tol 2008). Speziell mit *ludoaudiovisuellen Harmonien* verschmilzt die Musik mit dem Spiel, welches so eine musikalische Qualität erlangt: „Notions of determinably diegetic and non-diegetic sound dissolve as the whole game activity takes on a musical quality.“ (Summers 2016, 195).

## Quellenverzeichnis

### Literatur

Arsenault, Dominic (2008): Guitar Hero: „Not like playing guitar at all"? In: loading... 2 (2), S. 1–7.

Aarseth, Espen J. (1997): Cybertext – Perspectives on Ergodic Literature. Baltimore: The John Hopkins University Press.

Bakels, Jan-Hendrik (2016): Audiovisuelle Rhythmen. Filmmusik, Bewegungskomposition und die dynamische Affizierung des Zuschauers. Berlin: De Gruyter.

Berndt, Axel (2014): Ambiente Musik zur Vertonung immersiver interaktiver Medien. In: Institut für immersive Medien (ifim) (Hg.): Klänge, Musik & Soundscapes. Marburg: Schüren, S. 95–105.

Cardona-Rivera, Rogelio E.; Zagal, José P.; Debus, Michael S. (2020): Narrative Goals in Games: A Novel Nexus of Story and Gameplay. In: Georgios N. Yannakakis, Antonios Liapis, Penny Kyburz, Vanessa Volz, Foad Khosmood und Phil Lopes (Hg.): International Conference on the Foundations of Digital Games. FDG '20: International Conference on the Foundations of Digital Games. Bugibba Malta. New York: ACM, S. 1–4.

Carlson, Marvin (2004): Performance: a critical introduction, London: Routledge.

Chion, Michel (1994): Audio-vision. Sound on screen. New York: Columbia University Press.

Claussen, Jan Torge (2021): Musik als Videospiel - Guitar Games in der digitalen Musikvermittlung. Hildesheim: Universitätsverlag Hildesheim.

Collins, Karen (2008): Game sound. An introduction to the history, theory, and practice of video game music and sound design. Cambridge Mass.: MIT Press.

Csikszentmihalyi, Mihaly (1990): Flow. The Psychology of Optimal Experience. New York: Harpercollins eBooks.

Daniels, Dieter; Naumann, Sandra (2013): Shifting Aesthetics of Image-Sound Relations in the Interaction between Art, Technology, and Perception. Berlin: Peter Lang.

Debus, Michael S.; Zagal, Jose P.; Cardona-Rivera, Rogelio E. (2020): A Typology of Imperative Game Goals. In: Game Studies - the international journal of computer game research 20 (3).

Ermi, Laura; Mäyrä Frans (2005): Fundamental Components of the Gameplay Experience: Analysing Immersion. In: 2005: Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play.

Fischer, Erik; Schlüter, Bettina (2009): Klänge und Klangräume in Doom 3 – Zum Sound Design von Ego-Shootern. In: Matthias Bopp, Serjoscha Wiemer, Rolf F. Nohr (Hg.): Shooter. Eine multidisziplinäre Einführung. Münster: Lit Verlag, S. 353–371.

Fritsch, Melanie (2021): Gitarrensimulant\*innen und Daumenartist\*innen? In: Sebastian Möring, Manuela Pohl und Nathanael Riemer (Hg.): Didaktik des digitalen Spielens. DIGAREC Series 09. Potsdam: Universitätsverlag, S. 66-94.

Gibons, William (2011): Wrap Your Troubles in Dreams - popular music, narrative and dystopia in Bioshock. In: Game Studies - the international journal of computer game research 11 (3).

Harrer, Sabine (2018): Games and Bereavement. How Video Games Represent Attachment, Loss, and Grief. Bielefeld: transcript.

Herber, Norbert (2008): The composition-instrument: emergence, improvisation and interaction in games and new media. In: Karen Collins (Hg.): From Pac-Man to pop music. Interactive audio in games and new media. Realizing groundbreaking adaptive music. Aldershot Hampshire England, Burlington VT: Ashgate (Ashgate popular and folk music series), S. 103–123.

Herwig, Jana (2017): Sinnliche Immersion und haptische Medien: Utopien und Möglichkeiten. Marburg: Schüren.

Huizinga, Johan (2004). Homo Ludens: Vom Ursprung der Kultur im Spiel (H. Nachod, Übers.; 24. Aufl.). Hamburg: rororo.

Jørgensen, Kristine (2007): 'What are Those Grunts and Growls Over There?' Computer Game Audio and Player Action. Phil. Diss. Copenhagen: Copenhagen University.

Lack, Russell (1997): *Twenty Four Frames Under. A Buried History of Film Music*, London: Quartet Books.

Langer, Susanne K. (1953): *Feeling and Form. A Theory of Art Developed from Philosophy in a New Key*, New York: Charles Scribner.

Linderoth, Jonas (2011): *Beyond the digital divide: an ecological approach to gameplay*. In: *Think Design Play: Proceedings of the 2011 Digital Games*.

Manovich, Lev (2001): *The language of new media*. Cambridge, Massachusetts: MIT Press.

Medina-Gray, Elizabeth (2016): *Modularity in Video Game Music*. In: Michiel Kamp, Tim Summers, Mark Sweeney (Hg.): *Ludomusicology - Approaches to Video Game Music*. Sheffield: Equinox eBooks Publishing, S. 53-72.

Post, Jack (2006): *Phatic communication in computer games*. Marburg: Schüren.

Rupter-Kruse, Patrick (2014): *Klänge, Musik & Soundscapes. Zur Entschlüsselung auraler Präsenzen*. In: Institut für immersive Medien (ifim) (Hg.): *Klänge, Musik & Soundscapes*. Marburg: Schüren, S. 11–16.

Schwinning, Reinke (2023): *Mit klingendem Spiel. Musikalische Inszenierung von Krise und Eskalation in Videospielen*. In: *Spiel Formen 2* (1), S. 31–60.

Sites, Joshua D.; Potter, Robert F. (2018): *Everything Merges with the Game: A Generative Music System Embedded in a Videogame Increases Flow*. In: *Game Studies - the international journal of computer game research* 18 (2).

Stockburger, Axel (2003): *The Game Environment from an Auditive Perspective*. In: *Level Up, Digital Games Research Conference (2003)*, Utrecht Universiteit.

Summers, Tim (2011): Playing the Tune: Video Game Music, Gamers, and Genre. In: ACT Zeitschrift für Musik und Performance (2).

Summers, Tim (2016): Understanding video game music. Cambridge United Kingdom, New York: Cambridge University Press.

Svec, Henry Adam (2008): Becoming Machinic Virtuosos: Guitar Hero, Rez, and Multitudinous Aesthetics. In: loading... 2 (2).

van Geelen, Tim (2008): Realizing groundbreaking adaptive music. In: Karen Collins (Hg.): From Pac-Man to pop music. Interactive audio in games and new media. Realizing groundbreaking adaptive music. Aldershot Hampshire England, Burlington VT: Ashgate (Ashgate popular and folk music series), S. 93–102.

Wark, McKenzie (2007): Gamer theory. Cambridge Mass.: Harvard University Press.

Wedel, Michael (2020): Ort und Zeit. Filmische Heterotopien von Hochbaum bis Tykwer. Berlin: De Gruyter.

Wharton, Alexander; Collins, Karen (2011): Subjective Measures of the Influence of Music Customization on the Video Game Play Experience - A Pilot Study. In: Game Studies - the international journal of computer game research 11 (2).

Yannakakis, Georgios N.; Liapis, Antonios; Kyburz, Penny; Volz, Vanessa; Khosmood, Foad; Lopes, Phil (Hg.) (2020): International Conference on the Foundations of Digital Games. FDG '20: International Conference on the Foundations of Digital Games. Bugibba Malta, 15 09 2020 18 09 2020. New York: ACM.

Zagal, José P.; Fernández-Vara, Clara; Mateas, Michael (2008): Rounds, Levels, and Waves. The Early Evolution of Gameplay Segmentation. In: Games and Culture 3 (2), S. 175–198.

## **Internetquellen**

Duden.de (2024a): „Musik: Rechtschreibung, Bedeutung, Definition, Herkunft | Duden“. In: duden.de, 2024. <https://www.duden.de/rechtschreibung/Musik> (15.06.2024).

Duden.de (2024b): „Kadenz: Rechtschreibung, Bedeutung, Definition, Herkunft | Duden“. In: duden.de, 2024. <https://www.duden.de/rechtschreibung/Kadenz> (15.06.2024).

Filmlexikon (2024): „Split Screen | Was ist ein Split Screen? | Filmlexikon“. In: film-lexikon.org, 2024. <https://film-lexikon.org/split-screen/> (15.06.2024).

Genius.com (2022): „Two Feathers – Silent No More Lyrics“. In: genius.com, 2022. <https://genius.com/Two-feathers-silent-no-more-lyrics> (15.06.2024).

Hamburger-Kunsthalle.de (2024): „SEE, HEAR, PLAY KANDINSKY!“. In: hamburger-kunsthalle.de, 2024. <https://www.hamburger-kunsthalle.de/see-hear-play-kandinsky> (15.06.2024).

Hemetsberger, Paul (2024): „ludo | Übersetzung Latein-Deutsch“. In: la-de.dict.cc (2024). <https://la-de.dict.cc/?s=ludo> (15.06.2024).

Hermida, Alfred (2002): „Gamers set for sensory overload“. In: news.b-b.co.uk, 2002. <http://news.bbc.co.uk/1/hi/sci/tech/1846561.stm>. (15.06.2024).

Hjertberg, Nicklas; Björkman, Elvira (2024): „About | Two Feathers“. In: twofeathersstudio.com, 2024. <https://twofeathersstudio.com/about/> (15.06.2024).

Horetski, David (2023): „AMD tries to take on Nvidia Reflex again in latest driver update“. In: dexerto.com (2023). <https://www.dexerto.com/tech/amd-tries-to-take-on-nvidia-reflex-again-in-latest-driver-update-2281652/> (15.06.2024).

Hunt, Steve (2018): „Beat Hazard 2 - Dev Blog #3 - Game Loop“. In: youtube.com (2018). <https://youtu.be/Llecn4LnrHY?si=L2Ak30oyx3nZqNsW&t=435> (15.06.2024).

Huiberts, Sander; van Tol, Richard (2008): „IEZA - A Framework For Game Audio“. In: gamedeveloper.com (2008). <https://www.gamedeveloper.com/audio/ieza-a-framework-for-game-audio> (15.06.2024).

Iwata, Saturo (2010): „Iwata asks: Super Mario Galaxy 2 -Volume 2 -Page 5“. In: iwataasks.nintendo.com, 2010. <https://iwataasks.nintendo.com/interviews/wii/supermariogalaxy2/1/4/> (15.06.2024).

Johanas, John (2024): „Developing ‚Hi-Fi RUSH‘ Backwards and Finding Our Positive Gameplay Loop“. In: youtube.com (2024). <https://www.youtube.com/watch?v=pG4UxqRMNX0> (15.06.2024).

Kaczmarek, Ludger; zu Hünigen, James (2022): „mickey mousing [Das Lexikon der Filmbegriffe]“. In: filmlexikon.uni-kiel.de, (2022). <https://filmlexikon.uni-kiel.de/doku.php/m:mickeymousing-2065> (15.06.2024).

Macgregor, Jody (2018): „The making of Audiosurf, the synesthesia simulator“. In: pcgamer.com, (2018) <https://www.pcgamer.com/the-making-of-audiosurf-the-synesthesia-simulator/> (15.06.2024).

Musikexpress.de (2019): „The Rolling Stones haben Richard Ashcroft endlich die Urheberrechte an ‚Bitter Sweet Symphony‘ zurückgegeben“. In: musikexpress.de, 2019. <https://www.musikexpress.de/the-rolling-stones-haben-richard-ashcroft-endlich-die-urheberrechte-an-bitter-sweet-symphony-zurueckgegeben-1284379/> (15.06.2024).

Rammstein Official (2019): „Rammstein - Stripped (Official Making Of)“ In: youtube.com, 2019. <https://www.youtube.com/watch?v=mImuguOghRM> (15.06.2024).

Schlütter, Christian (2008): „Unser Interview mit dem Audiosurf-Schöpfer Dylan Fitterer“. In: pcgames.de, (2008). <https://www.pcgames.de/Audiosurf-Spiel-18404/Specials/Unser-Interview-mit-dem-Audiosurf-Schoepfer-Dylan-Fitterer-632002/> (15.06.2024).

Steam (2021): „Steamworks Development – NEU: Sehen Sie, wie viele Controller Spieler in Ihrem Spiel verwenden: Steam-News“. In: store.steampowered.com, 2021. <https://store.steampowered.com/news/group/4145017/view/3061855517864424914> (15.06.2024).

Synaesthesia.org (2024): „Was ist Synästhesie? | Deutsche Synästhesie-Gesellschaft e.V.“. In: synaesthesia.org, 2024. <http://www.synaesthesia.org/de/synaesthesia> (15.06.2024).

Wilhelm, Parker (2023): „Crack Open Hi-Fi Rush’s Newest Update with Tips from the Game Director“. In: news.xbox.com, 2023. <https://news.xbox.com/en-us/2023/07/05/hi-fi-rush-arcade-challenge-update-tips/?ran->

MID=24542&ranEAID=kXQk6\*ivFEQ&ranSiteID=kXQk6.ivFEQ-2g-rE.F5nBDypoOYA\_OrJfA&epi=kXQk6.ivFEQ-2grE.F5nBDypoOYA\_OrJfA&irg-wc=1&OCID=AIDcmm549zy227\_aff\_7593\_1243925&tduid=%28ir\_\_rqmy-hgk2kokfakilz1ixawlvxm2xdj6xpb0xtsy00%29%287593%29%281243925%29%28kXQk6.ivFEQ-2grE.F5nBDypoOYA\_OrJfA%29%28%29 (15.06.2024).

Wulf, Christopher (2022): Remediation und Narration in Computerspielen am Beispiel der Metal Gear Solid Reihe. <https://doi.org/10.13154/294-8406>.

Wulff, Hans Jürgen (2022a): „Diegese [Das Lexikon der Filmbegriffe]“. In: filmlexikon.unikiel.de, 2022. <https://filmlexikon.uni-kiel.de/doku.php/d:diegese-122> (15.06.2024).

Wulff, Hans-Jürgen (2022b): „Immersion [Das Lexikon der Filmbegriffe]“. In: filmlexikon.uni-kiel.de, 2022. <https://filmlexikon.uni-kiel.de/doku.php/i:immersion-942> (15.06.2024).

Wulff, Hans-Jürgen (2023): „vertikale Montage [Das Lexikon der Filmbegriffe]“. In: filmlexikon.uni-kiel.de, 2023. <https://filmlexikon.uni-kiel.de/doku.-php/v:vertikalemontage-4210?> (15.06.2024).

## **Film**

CRUEL INTENTIONS (USA 1999, Roger Kumble)

FULL METAL JACKET (UK/USA 1987, Stanley Kubrick)

INGLORIOUS BASTARDS (USA/Deutschland 2009, Quentin Tarantino)

PULP FICTION (USA 1994, Quentin Tarantino)

THE MATRIX (USA 1999, Laurence/Andrew Paul Wachowski)

THE ROYAL TENENBAUMS (USA 2001, Wes Anderson)

## **Fernsehen**

THE SIMPSONS (USA 1989, Matt Groening)

WER WIRD MILLIONÄR? (Deutschland 1999, Matthias Schwab)

## **Musik**

BITTERSWEET SYMPHONY (UK 1997, The Verve)

COMANCHE (USA 1961, The Revels)

HEY JUDE (UK 1968, The Beatles)

INVADERS MUST DIE! (UK 2009, The Prodigy).

THE LAST TIME (USA 1965, The Rolling Stones)

## **Musikvideos**

FREESTYLER (Finnland 1999, Miikka Lommi)

STRIPPED (Deutschland 1998, Philipp Stölzl)

## **Computerspiele**

3D TETRIS (Japan 1996, Nintendo, Virtual Boy)

ASTEROIDS (USA 1979, Atari, Arcade)

AUDIOSURF 2 (USA 2015, Dylan Fitterer, Windows PC)

BEAT HAZARD (UK 2009, Cold Beam Games, Xbox 360)

BEAT HAZARD 2 (UK 2018, Cold Beam Games, Windows PC)

BIOSHOCK (USA 2007, 2K Games, Windows PC)

DANCE DANCE REVOLUTION (Japan 1998, Konami, Arcade)

DONKEY KONGA (Japan 2003, Nintendo, Gamecube)

DOOM (USA 2016, Bethesda, Windows PC)

DOOM 3 (USA 2004, Activision, Windows PC)

FALLOUT 3 (USA 2008, Bethesda, Xbox 360)

FALLOUT 4 (USA 2015, Bethesda, Windows PC)

GTA SAN ANDREAS (USA 2005, Rockstar Games, Xbox)

GITAR HERO (USA 2005, RedOctane, Playstation 2)

JET SET RADIO FUTURE (Japan 2002, Sega, Xbox)

HI-FI RUSH (USA 2023, Bethesda, Windows PC)

METAL GEAR SOLID 4 (Japan 2008, Konami, Playstation 3)

METAL: HELLSINGER (Norwegen 2022, Funcom, Windows PC)

METROID PRIME (Japan 2002, Nintendo, Gamecube)

NEW SUPER MARIO BROS. (Japan 2006, Nintendo, Nintendo DS)

PARAPPA THE RAPPER (Japan 1996, Sony Computer Entertainment, Playstation)

PROTEUS (UK 2013, Twisted Tree, Windows PC)

QUAKE 2 (USA 1998, Acitivision, Windows PC)

REZ (Japan 2001, Sega, Dreamcast)

REZ INFINITE (Japan 2017, Enhance, Windows PC)

RIDGE RACER (Japan 1994, Namco, Playstation)

SEGA GT 2002 (Japan 2002, Sega, Xbox)

SSX3 (Kanada 2003, EA, Xbox)

STAR FOX 64 (Japan 1997, Nintendo, Nintendo 64)

STARFIELD (USA 2023, Microsoft, Windows PC)

SUPER MARIO GALAXY (Japan 2007, Nintendo, Nintendo Wii)

TETRIS (Sowjetunion 1985, Alexey Pajitnov, Elektronika 60)

TETRIS EFFECT (Japan 2018, Enhance, Windows PC)

THE ELDER SCROLLS III: MORROWIND (USA 2002, Bethesda, Windows PC)

THE LEGEND OF ZELDA: OCARINA OF TIME (Japan 1998, Nintendo, Nintendo 64)

TONY HAWK'S PRO SKATER 1 + 2 (USA 2020, Activision, Windows PC)

TWISTED METAL 4 (USA 1999, 989 Studios, Playstaion)

## **Programme**

Audacity (2000-2024, Anonym/Open Source)

DMXControl 2 (Deutschland 2024, DMXControl Projects e.V.)

GamePad Viewer (USA 2024, Christopher Rodriguez)

Open Broadcasting Software (2012-2024, Anonym/Open Source)

VLC Media Player (Frankreich 2001-2024, VideoLAN)

Winamp (USA 1997-2024, Radionomy S.A.)

## **Abbildungsverzeichnis**

### **Abbildung 1**

INGLORIOUS BASTARDS (USA/Deutschland 2009, Quentin Tarantino)

### **Abbildung 2**

THE MATRIX (USA 1999, Laurence/Andrew Paul Wachowski)

### **Abbildung 3**

FULL METAL JACKET (UK/USA 1987, Stanley Kubrick)

### **Abbildung 4**

PULP FICTION (USA 1994, Quentin Tarantino)

### **Abbildung 5**

THE ROYAL TENENBAUMS (USA 2001, Wes Anderson)

### **Abbildung 6**

CRUEL INTENTIONS (USA 1999, Roger Kumble)

### **Abbildung 7**

THE SIMPSONS (USA 1989, Matt Groening)

### **Abbildung 8**

STRIPPED (Deutschland 1998, Philipp Stölzl)

### **Abbildung 9**

FREESTYLER (Finnland 1999, Miikka Lommi)

### Abbildung 10.1

NEW SUPER MARIO BROS. (Japan 2006, Nintendo, Nintendo DS)

### Abbildung 10.2

SUPER MARIO GALAXY (Japan 2007, Nintendo, Nintendo Wii)

### Abbildung 11.1

Que Swazy: „Split Screen Local multiplayer|Tony Hawk Pro Skater 1 + 2” Youtube, 08.09.2020, [https://youtu.be/\\_nXKSpaSKGU](https://youtu.be/_nXKSpaSKGU). 06:55. (19.06.2024).

### Abbildung 11.2

SourceSpy91: „MGS 4 - Full Game Walkthrough - No Alerts - No Commentary” Youtube, 21.08.2016, <https://youtu.be/wEUtXH1cvkg>. 08:10:45. (19.06.2024).

### Abbildung 12

BEAT HAZARD 2 (UK 2018, Cold Beam Games, Windows PC)

### Abbildung 13

AUDIOSURF 2 (USA 2015, Dylan Fitterer, Windows PC)

### Abbildung 14

METAL: HELLSINGER (Norwegen 2022, Funcom, Windows PC)

### Abbildungen 15 und 16

Hi-Fi RUSH (USA 2023, Bethesda, Windows PC)

### Abbildung 17

REZ INFINITE (Japan 2017, Enhance, Windows PC)

### Abbildung 18

TETRIS EFFECT (Japan 2018, Enhance, Windows PC)

### Abbildung 19

AUDACITY (2000-2024, Anonym/Open Source)

## **Eigenständigkeitserklärung**

Hiermit versichere ich, dass ich die Arbeit selbständig angefertigt, außer den im Quellen- und Literaturverzeichnis sowie in den Anmerkungen genannten Hilfsmitteln keine weiteren benutzt und alle Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, unter Angabe der Quellen als Entlehnung kenntlich gemacht habe.

Christopher Wulf

  
-----

Unterschrift

21.06.2024

Datum