Sound and Music in Games

by Peter Peerdeman

http://www.few.vu.nl/~ppeerde

Abstract

Sound and music play an ever growing role in the development of video games. Because of technological advancements and the broadening target audience game experiences with hollywood movie quality scripts and audio are uprising. This trend also opens up a whole new genre of musical games that let gamers experience music in a whole new interactive manner. By discussing the different aspects, technologies and trends of video game audio I give an introduction to the increasing importance of audio in the video game industry.

Introduction

In the past decade, sound and music in games have evolved rapidly. While sounds from the first arcade machines were merely designed to attract attention from gamers walking around the arcade hall, modern day games feature complex orchestrated dynamic soundtracks to accompany the movie-like experience along with context supporting sound effects that contribute to the immersion, or “emotional involvement” of the player.

For as long as video games have existed, audio has always offered challenges for game developers. Some of the challenges have been solved by technology, but other fundamental challenges still exist, for example: “how does a game with a dynamic soundtrack tell a consistent and interesting story?” On the other hand, audio also offers even more opportunities, for instance the ability to evoke emotional responses, immerse players, and create a sense of location within a game.

Unlike hollywood movies, that are carefully directed and inherently linear, games have the exciting feature of being non-linear and interactive. Contrary to movies, games should be able to respond adequately to a player’s actions within the game with both visuals and audio. This brings up the need for dynamic audio, which is a solution to incorporate that cinematic realism in games that people enjoy in movie theaters.

When game developers saw the potential of incorporating music in games as a strong selling point, some other genres of games emerged: the “rhythm based action games” and
later the “free form music games”. These games offer the player a sense of interaction with or even creation of music, which creates a whole new breed of musical talents that enjoy music in a more artistic way than before. This genre has proven to be a very successful one, attracting not only a game interested audience but also the music interested audience.

The interesting part of the integration of music and games is that it is really a symbiotic relation instead of a mutual addition. The rise of games like Rez[19], SingStar[6] and Guitar Hero [9] have not only caused music to become a part of games, but also that games have become a part of music, with the music industry finding a new media format to publish songs and albums in.

In comparison to movies and music, games are another medium to provide an experience of an art form, in a more non-linear way than in for instance a movie theater. Though the linearity of these media is different and there are a lot of different goals such as education, propaganda, entertainment, they at least share the goal of immersing an audience with a carefully designed experience, where sound and music are of great value.

1 Interactive audio: opportunities and challenges

Audio has proved to play a significant role in consumer preference in product selection and is viewed as one of the most important element of games [24]. This is mainly due to the degree of storytelling. As Steven Spielberg once said “Sound and music make up more than half of communicating a story, greater even than what you’re seeing”. In addition to this, sound is a great sensory stimulus to the player’s consciousness and even to the subconsciousness, affecting the mental processes without the player even noticing.

![Figure 1: Metal Gear Solid 2 for Sony’s Playstation 2](image_url)

There are many possibilities for sound effects to affect the gamer’s experience while playing a game. A great example of a game that incorporates a lot of these effects is the Metal Gear Solid (MGS) series, directed by Hideo Kojima and published by Konami[11]. The games subtitle “tactical espionage action” explains that this game is a sneaking game with strong action elements. The main goal of creator Hideo Kojima has always been to make a game with hollywood sound and visuals, feeling like an action movie[7]. This results in a very epic, moody and atmospherical game that has set a standard in the gaming industry for interactive storytelling.

1.1 Sound effect uses

One of the main uses for sound effects is to attract attention. In the MGS games this is used in a multitude of ways. For instance, when a player picks up a new item in the game it is accompanied by a sound effect that sounds like the cocking of a gun, attending the player on the
fact that he/she has acquired a new item. In another case, for example at the intro sequence of MGS1, the attention of the player is attracted with the epic soundtrack immediately setting the mood for the game.

Another audio property that is commonly used in games is that sound can **evoke emotional response** such as the feeling of success, failure, tension, relief etc. A great example in the MGS game is the sound that is played when the player is killed: a high pitched screech like synthesizer sound accompanied by the support character screaming the main characters name through the radio that really evokes a strong sense of failure.

Audio can also be used to create a certain **ambience or suggestion**. By picking the right music for a certain scene, it is more apparent what the goal of that piece of gameplay might be. If the music is very aggressive and loud, a player may consider to shoot an enemy, while the player would consider sneaking past the enemy if there is ambient and sneaky-feeling music playing. In a likewise manner, audio can be used to make the player **anticipate on what is going to happen**, for instance silencing out all the other sounds and playing an eerie ambient track before the player is scared by a final boss fighting sequence.

This brings up another feature of audio. Combined with visuals and other sensory stimuli audio can help to **resonate a memorable moment**. Because of the strong emotional feelings, or even synesthesia that is evoked in people when listening to music and sound effects in general, certain moments can be remembered a lot better when these moments are amplified with a memorable piece of music or sound.

The last feature in this list is the ability of audio to **enhance the structure and adding a sense of location** in game environment. By linking certain areas with certain soundtracks or sound effects, players will get a better understanding of the worlds map, and could for instance immediately detect when they are going somewhere they have been before. If there is a certain task that has to be performed in part of the game before the game can continue, for instance eliminating all the enemies in a room, the player will know what to do next time the same music is used.

### 1.2 Challenges in non-linear audio

Creating music is inherently a linear process. The quality of music, as perceived in general, is greatly depending on the “appropriate structure on many levels”\(^1\). Music is carefully composed by an artist to capture a certain feeling or mood that is transferred to the listener through the medium. In a linear form, the composer has full control over the timeline of a certain composition and can for instance build up to a certain climax, creating a dramatic curve within the continuity of the motion.

However, in games time is controlled by the player. While playing a game, the player is “an active agent of change” \(^3\) that has a certain control over the timeline of the game. This poses a lot of challenges to the composer, because of the non-deterministic origin of the player’s actions. In an interview\(^7\) the composer of the MGS soundtracks, Harry Gregson-Williams explains that the music written for the gameplay is not melodically based, but even more so on rhythm, for it is far easier to cut to another tempo in mid beat than it is to interrupt a melody line.

A common workflow between composer and sound director in a game studio is usually a feedback loop in which the composer will write some music according to the description the sound director delivers. When the music is finished, the composed music is then cut up in pieces that can be looped or dynamically changed according to the actions taken in the game. On one moment, the player may be sneaking around unnoticed, but when he/she gets sighted by the enemy, the music should instantly adapt to the “alerted” situation.

Another problem in adaptive non-linear sound is the vast amount of resources needed to create a truly adaptive virtual environment. When certain games are played for months, years even, players will experience the bounds of the adaptive freedom a game can take. Limited to a certain disk size or access time, there is only so much content a developer can
deliver with its game. An example of a system that achieves non-linear dynamic audio or so-called “intelligent control” over the in game audio track is discussed in appendix A.2.

2 Sound production: from Hollywood to the game industry

With the possibilities of game development expanding each day, game developers can now produce movie-like games with dramatic story lines and impressive action sequences. As said before, it is the sound that an audience hears which makes up for more than half of the story that is being told. Therefore, movie-like games need movie-like sound and music.

When looking at the production process of sound and music for games[3], a lot has been learned from hollywood. Sound effects that are used in games are often recorded in a “foley studio”. This is basically a room full of extraordinary objects that allow a foley artist to create sounds such as clothing sounds and footsteps. These sounds are recorded in the studio and can then be used in the game to create a realistic soundscape. As sound directors from movies have witnessed, the original recorded “real” sound effects often do not have as much impact on the viewer as a “fake” sound effect. This leaves the sound director with a whole lot of aesthetic choices of metaphors that suit the visuals best.

2.1 Alternative content sources

For the sounds that can’t be produced in the foley studio, field recording is used to record real life sounds such as rifle sounds or just outdoor location sounds such as forests and city environments. Because games are not “filmed” on location and the created visuals have no environmental sound of their own, all of the ambient sound has to be recorded either in foley studio or in field recording. As a last refuge, sometimes stock sound CD’s are used as source for sound effects.

For movie franchise games, such as the Harry Potter game series, game developers have access to a large amount of content that is used in the movie, such as the soundtracks but also the foley studio sound effects and ambient set recordings[1]. This is very important for franchise games because the goal of the game is to feel as much as the original movie as possible, increasing the impact the franchise has on the target audience.

An example where this cross medial storytelling is taken to an extreme is the Matrix series. The game “Enter the Matrix” tells a parallel story that takes place during the second movie of the Matrix trilogy. During the production of the second movie, small movie clips were recorded at the original sets that would not get featured in the movie, but were able to be viewed by players between game sections in the Matrix video game. Not only did the game developers have access to the sounds used in the movie but also to exclusive not seen before footage that was intended to be used in the games from the beginning. This gives the symbiotic link between game and movie a whole new angle, which makes the movie interesting for both the hardcore gamer and the game interesting for someone who watched the movie.

2.2 Playable movies: cinematic gameplay

The cinematic aspect of games posed a lot of interesting opportunities for writers who wanted to tell an intricate story. In 2005, the game developers of Quantic Dream published their game Fahrenheit, a “unique combination of gaming and cinematography” which was argued to be “the savior for interactive movies and adventure games” by several game critics.[15] Although Quantic Dream didn’t officially used the label, Fahrenheit could very well be called an interactive movie.

The big difference between Fahrenheit and other games is that Fahrenheit’s gameplay is more time based as opposed to the more common player initiated gameplay. This makes “playing” Fahrenheit more like watching a movie that has subtle interactive capabilities.
Moreover, most of the full motion video (fmv) featured in the game is motion captured from live actors, which gives the cinematography a more authentic feel.

The narrative structure that is inherent to the interactiveness of the story that is being told is being handled in a bulging tree approach. This means that within one scene, the player could for instance make some choices which have effect on the next event (the branching tree “bulges” or expands) after which the scene is ended and the next scene is initiated bringing the possible outcomes of the previous scene together. This way of branching reduces the amount of possible routes through the story, effectively reducing the amount of content that has to be available for the game.

Next generation games, such as Quantum Dream’s: “Heavy Rain” for Playstation 3 have access to a lot more content space and processing power, allowing the story to branch further and really create a unique experience for each individual player. With the extra processing power, new techniques such as real time digital sound processing (DSP) are used more often to increase the adaptiveness of sound effects to non-linear games, making way for new and even more realistic interactive movies. More information on DSP can be found in appendix A.3.

2.3 Narrative Structure

In both ordinary and movie-like games music is a great medium to empower the narrative structure. While listening to a voiceover that explains a part of the story or by hearing a soundtrack that corresponds to the visuals you see on the screen, audio proves to be a very valuable asset in transferring a message from the media to the observer.

An important part of a game’s narrative power is the mapping of the gameplay narrative events in time to the intensity of these narrative events. When creating a game, the game designer has to carefully think about the dynamic range\[18\] of the game experience, that has to be consistent in the gameplay, visuals and audio. The dynamic range of a game is the amount of difference between the least intensive moments and the most intensive moments in the game experience. To avoid listener fatigue and boredom, the intensities of the different game events should vary between each event, creating a dramatic curve that enhances the narrative structure and is interesting to experience.

![Dynamic Map: Linear Shooter: Level 01](image)

Figure 2: Dynamic map for a sequence of events in a linear shooter game

The graph seen in figure 2 is a visualization of a dynamic map\[1\] that shows a sequence of events in a linear shooter game, quite like a dynamics map for a hollywood movie. The graph represents the intensity of all the events within one game level, showing a build up

\[1\] Dynamic map graphs made by Rob Bridgett, see [18]
in intensity with 2 peaks of combat to a big climax at the end of the scene. Three lines in
the graph represent different layers of audio and one black line represents the general game
intensity. Each of these layers play a distinct role in the storytelling while following the
general cadence of the game intensity line.

The most gradual line in this graph is the music, slowly swelling up towards the climax
with some peaks of intensity during the combat events. The line with the most variety in
intensity are the sound effects, that play a very important role in immersing the player in
the combat sequences while stepping down during the cinematic. The line with the greatest
intensity is the dialogue, which plays a big part in the identification of the player with the
game characters and form the structure in which the assignments are given for the tasks the
player has to complete.

Figure 3: Dynamics map for an open world game detailing the escalation and reduction of
intensity

However, not all games are as linear as the first example. In other games such as in open
world games in which the gameplay is more dependant on actions of the player, the dynamics
can’t be sequentially scripted beforehand for there is no knowledge of what the player might
do next. For these cases, other structures for dynamic mapping have to be created, as seen
in figure 3. In this figure, the game state is described in 4 different “stages” between which
the game can step down or step up. The stepping process can occur at any point in time,
effectively creating a unique dynamics map for each player experience.

Figure 4: Overall dynamics map for an open world game
Though there is some added control over the intensity of certain audible layers, game designers are still offered with a challenge to build up to a satisfying climax when offering open ended gameplay. This can be achieved by for instance chaining multiple pieces of open world gameplay together with some scripted sequences in between, as shown in figure 4. In this example the biggest part of dialog and story is conveyed in the scripted events between open world gameplay, in addition to the increasing intensity of the game’s music. In this way a natural build up is made to the climax, which is also open world gameplay, but with the most intense music and sound effects.

Carefully crafting the dynamics map for the available gameplay in a game gives game designers a lot of control over the impact of the game on the player. In order to create space for very intensive combat events there have to be some low intensity events to avoid listener fatigue. In comparison with movies that only last for 1 to 2 hours, games have to entertain a player for at least 5 to 10 hours creating an even bigger need for well balanced dynamics maps. Adding a freedom of choice for the player creates extra difficulties in achieving an interesting and structured dynamics map, effectively limiting the strength of the narrative structure. To get the best of both worlds, a compromise has to be made between offering an interesting, compelling story and immersive, dynamic gameplay.

3 Music game genres

For movie-like gaming experiences, usually bombastic and epic orchestral compositions are used quite like in cinema. However, in game genres such as racing and sports games, licensed music is used more often. The big difference between licensed music and specifically composed music is the replaceability of the music and the inability to adapt this music to the onscreen actions.

Because the licensed soundtracks are purely linear it is very hard for developers to have the music support the narrative. Often the soundtrack is chosen randomly from a selection of the same kind of music, all punk or all electronic songs for example. The advantage that sports and racing games have is that the length of the gameplay parts is usually known beforehand, making it easier to fill in the linear soundtracks into the game.

The big difference in the aforementioned genres are that they are reality simulations more than fictional story games as discussed before. The advantage of this realism is that a song from a popular band does not stand out in games like these and acts more as a background music than really supporting the visuals and actions performed in game.

On the other hand, there is another genre of games that does rely heavily on the background music tracks and in game sounds, taking audio in games one step further. These games can be described as music games, subdivided in several categories such as generative-form music games, rhythm-based games and free form music games.

3.1 Generative form music game: Rez

A very popular and innovative Playstation title called Rez is an example of a generative form music game. The game is a so called “rail shooter” that takes place in an abstract wireframe model environment. The player is moved through the virtual world automatically on a fixed path (rail) and has the ability to look around and shoot the enemies down.

Though rail shooters were quite common, Rez innovated by fully integrating the sound effects and soundtrack to the gameplay. The electronic soundtrack is in style with the digital environments, and reacts on the players actions. Because every sound effect is quantized to the beat, which means that each sound effect is always played on the beat of the soundtrack, the player experiences the audio as if he or she is creating it.

The stated goal of Rez was to “experience synesthesia”[13], in other words reaching a condition in which one type of stimulation evokes the sensation of another, in this case manipulating the controller leading to the sensation of sound and visuals. Because of this tight
integration the player feels as if he or she is actually creating a musical composition, hearing the hi-hat chime and part of a generated melody played when the enemy is targeted. Another good example of real musical interactivity is when the player is selecting all the enemies in one sweep instead of targeting them one by one resulting in one big arpeggio or short staccato notes respectively, creating a sense of compositional freedom while playing the game.

3.2 Rhythm-based music games

The oldest approach to music based games are the rhythm-based games. In these games the player has to press one of the buttons on the controller in synch with a certain audio and/or visual queue. In the last couple of years the Guitar Hero series have reintroduced this genre to the gaming industry.

Quite like older rhythm-based games such as Parappa the Rapper, Dance Dance Revolution and Vib Ribbon, players have to listen to the soundtrack carefully and press the right buttons on the beat of the music. The games are often bundled with a gaming peripheral such as a dance mat which allows the player to use his feet to press the buttons or a guitar shaped controller in the case of Guitar Hero that allows the player to get the feeling of virtually playing guitar while actually just pressing buttons on the plastic guitar shaped controller.

Contrary to the earlier described generative form music games, the rhythm action based games have no compositional freedom whatsoever, for all gameplay and audio is based on the predetermined soundtrack of the game. There are clear visual and sometimes audible signs when a player fails to press the right button, such as a “wrong note” sound effect in Guitar Hero or the character shrinking in size in Vib Ribbon, but the player is always “performing” the same song over and over again.

To avoid this so called “listener fatigue” and to extend the content of the game with user content, the game designers of Vib Ribbon implemented an interesting feature in their game. Because the game is loaded in RAM memory, the game disc can be swapped while selecting the song to choose, offering the player the ability to insert his or her favorite audio cd. The game then analyzes the beat in the song and generates the level content accordingly. This literally creates an endless supply of game content, though the gameplay stays limited to only pushing the buttons on the right time. The advantage of this feature is that the game is
not limited to the licensed music that is shipped with the game.

3.3 Free Form Music Games

The last genre mentioned in the introduction of this section is the free form music game genre. These “games” often lack a real goal or narrative, but have the creation of music as a primary driving motive. Therefore, the line between “game” and “music production software” is quite vague in these games.

However, “playing a game” might not be that different from “playing an instrument”. By manipulating some controls, be it a controller or for instance a piano, we get some feedback in return, be it visuals and sound from the screen or acoustic sound coming from the piano. In this way, these “games” can even be seen as digital music instruments, rather than a video game.

![Figure 6: Electroplankton for Nintendo DS](image)

A good example of such a “digital music instrument video game” is Electroplankton[16], released for the Nintendo DS portable console. It was one of the few launch titles for the DS which increased the attention it got from early adopters. The game consists of several mini games which are little interactive environments that create a certain sound or musical loop when interacted with. For instance, drawing a line with the stylus on the touch screen makes an “electroplankton” character follow that line and produce sounds along the way, according to the shape of the line and the speed at which the line was drawn.

Because of the playful character and accessibility of these sorts of games, a big audience is reached that might never have produced music before or thought of ever producing music. Though Electroplankton is one of the most game-like free form music games, it has opened the way for more abstract free form music games that offer some more control and features, turning the gaming platform into a multitude of musical instruments.

Some other examples of games that try to stimulate gaming audience to create and com- pose music are Korg DS-10 and Jamsession for Nintendo DS and Beaterator for Playstation-Portable. Unlike Electroplankton, these games are more like digital instruments that allow you to create music without knowing how to play an ordinary instrument, making it a lot more accessible for people to pick up and try.

Korg DS-10 is a detailed emulation of Korg’s MS-10 synthesizer and sequencer[12]. Including all the functionality of the original synthesizer it is really a portable version of a fully functional synthesizer, with a huge variety in soundscapes. Because of the amount of functions it isn’t the easiest game because it requires quite some synthesizer knowledge before it can be used to its full extend. However, in the hands of a skilled player the game console is really transformed into a functional instrument.

Jamsession on the other hand is a lot more accessible and could almost be considered as a tutoring or instructional game for guitarist[17]. By dragging the stylus across the touchscreen
the strings of the guitar are strummed, the chord that is played is controllable by pressing down one of the function buttons. In the first game mode the game acts as an interactive music sheet and shows the chords that have to be strummed to play one of the supplied songs. The second free play mode allows the player to select their own chords and lets the user play whatever they want. By assigning chords to the buttons players can play any song they like, effectively turning the game console into a very accessible guitar simulation.

The final example is Beaterator, a beat producing game for the Playstation Portable[8]. This game is really a mix between an accessible song producing game and a fully functional sequencer. Though not as advanced as the Korg DS-10, the Beaterator is much easier to use, broadening the target audience. As shown in their promotional campaign, even children from the age of 10 and up are able to create music using the game. By allowing the user to record their own samples and playing them back within the game, the available content for the game is infinite. The Beaterator game is really an excellent example of providing a large audience with a way to express themselves in a musical fashion without investing time and money in lessons and equipment, but offering that experience on a game console primarily used for gaming purposes.

4 Conclusion

Sound and music in games is an ever evolving field of expertise that plays a very significant role in the immersion and narrative structure of video games. With the increasing processing power of the video game consoles, more and more complexity and quality can be built into game audio engines that produce even more realistic and compelling audible accompaniment to the gaming experience.

When comparing games to hollywood movies, a lot of similarities can be found. Not only the box-office figures that are generated by movies but also the techniques of creating a memorable experience are found in the field of game development. As the audio designer of the Metal Gear Solid 4 game mentioned in an interview, the game business strives to create games that can compete with blockbuster movies: “The quality of the script and the length of the scripts are starting to feel a lot more like feature films ... years ago, doing a video game was just short choppy lines that involved the gameplay”[5].

New streaming capabilities, increasing storage size and processing power result in a video games arms race that drives next generation video games toward realism, immersion and entertainment. The audio used in games is even conforming to a certain “listening etiquette” instead of just trying to draw attention like the early arcade machines. Realtime sound rendering for dynamic environments, positioning sound and music in 3D space and realtime priority and mixing capabilities even enhance the realism beyond realism to offer the most pleasing and entertaining experience to the gamers.

This ever growing group of gamers growing not only in amount but in range as well, is the core of an even more lucrative game business. Even record companies can’t deny the fact that a big part of their target group is playing video games and inevitably has to adept their strategy to this business, which is providing a new media format to convey content, and popularize certain music by offering them in immersive music games.

For the sound designers in the game development world, the focus of creating the soundscape for a video game has changed quite a bit, along with the available technology to incorporate their vision into the final video game. However, sound and music in games is still an aesthetic choice of metaphors, an arrangement of recorded content that not only creates an interactive experience for the end user but also completes the video game as a totally integrated work of art.
References


This is a very nice small documentary on the sound effects and music in the Harry Potter franchise games. In the genre of games that are based on a movie the audio design is approached differently for the game has to feel as if you were playing the movie. This means that the same sounds and effects have to be used that were present in the movie, effectively restricting the game designers in their creativity.


This paper discusses the sound capabilities of video game consoles of the 8-bit era (c. 1975-85), the technical approaches and the emotional responses the audience had to these distinctive sounds. One of the interesting remarks is that because of the technical limitations, 8-bit composers are forced to be creative with the limited set of musical resources. The biggest advantage of 8-bit audio is the adaptability of the music, that is able to very directly respond to the games context.


This book contains a general overview to the history, theory and practice of video game music and sound design. The book also has a lot of technical details about the inner workings of certain audio formats and technology that enables game designers to effectively use sound and music in their games. It also discusses immersion and general game experience when involving music and adaptive audio. Finally there is a section on compositional approaches to dynamic game music, discussing the ability of games to produce variable sound tracks.


In this short documentary two game designers that developed the sound engine of the game Battlefield Bad Company tell about the influence of mixing sound effects on the users experience while playing a game. By prioritizing sounds in a dynamic way, players always hear the sound that is most important for that situation the loudest.


This documentary on the extra’s disc that comes with Metal Gear Solid 2: Sons of Liberty for Sony’s PlayStation 2 gives an in-depth look into the different aspects of the game studio that produced the hollywood like action game Metal Gear Solid. In this documentary the lead sound design explains the philosophy of the production team about immersion and the cinematographic nature of the game which is driven by an epic adaptive soundtrack.
In this paper, several methods of algorithmic music generation are discussed, accompanied by the rationale that describes what makes music “interesting” or pleasant to hear. Some conclusion are posed that indicate that the technology for generating music is “sufficient” enough to avoid boredom and that these compositions are crafted well enough to let the human consciousness find “meaning” in the meaningless generated music.

This paper discusses the concept of interactive movies and how this concept has been criticized through time by the media. It takes the game Fahrenheit as an example that interactive movies aren’t as dead as the media thought and that there is a worthwhile future for these type of games.

This article on narrative dynamics gives a very brief introduction to some mechanisms used to create an interesting experience, using the difference in intensity to build up to certain climaxes and to relax players between big events. By visualizing the intensity over time in dynamics maps the structure of events can be overviewed in a structured and abstract way.

This article gives a very brief introduction to some mechanisms used to create an interesting experience, using the difference in intensity to build up to certain climaxes and to relax players between big events. By visualizing the intensity over time in dynamics maps the structure of events can be overviewed in a structured and abstract way.

This article gives a very brief introduction to some mechanisms used to create an interesting experience, using the difference in intensity to build up to certain climaxes and to relax players between big events. By visualizing the intensity over time in dynamics maps the structure of events can be overviewed in a structured and abstract way.

This article gives a very brief introduction to some mechanisms used to create an interesting experience, using the difference in intensity to build up to certain climaxes and to relax players between big events. By visualizing the intensity over time in dynamics maps the structure of events can be overviewed in a structured and abstract way.


This presentation gives a well argued view on the importance of audio in games using some quotes and convincing demographics. The talk is meant to motivate game designers to put even more effort and budget into audio design because sound it is very important when communicating a story.
Appendices

A Trends in Dynamic Audio

The evolvement of games through history have brought up some needs in terms of dynamic audio. Dynamic audio can be split up in two different forms. When games want to react on a player’s actions, not only the visuals but also the audio has to react and change to provide a real immersive experience. This form of dynamic audio is called interactive audio, triggered by the player’s actions. In addition to this, the game’s sound should respond to the current game state, which is called adaptive audio.

A great example of these two forms of dynamic audio can be seen in the game Fable 2 by Lionhead Studios. This game is an action role playing game in which the player leads his or her avatar through a digital world that is constantly changing to the player’s actions and the game’s state. For instance, the music and sound are adaptive to the purity or corruption of the player and interactive to the movement of the player, playing all sorts of sound effects such as footsteps, clothing and weapon sounds when moving through the game.

When the music composition is fully automated to change according to the game state we can speak of intelligent control. There have been several technological developments in time that have tried to achieve this kind of control, from which some are worthwhile to discuss in the upcoming sections. The biggest change in technology we have seen in the history of music in games is the change from Musical Instrument Device Interface (MIDI) to sampled audio recordings (Redbook).

A.1 MIDI vs Redbook audio

In early game development, the space available for music and sound for games was very limited. Sound directors had the task to get as much audible content out of as little bytes as possible. In pc games, the first soundcards, equipped with the YM3812 Yamaha FM chip, produced the first sounds in games using the MIDI sequencing language. The big advantage of MIDI was that music could be composed and recorded without programming the synthesizer chips, in a format that was understandable for musicians as well. Because the MIDI format consists only of commands, e.g. note C on and note C off messages, MIDI files do not contain any waveform data like Redbook audio does. This greatly reduces the valuable disk space needed for music files, as well as the amount of time needed to load the files in memory for playback.

Another big advantage of MIDI is that the sequenced passages are inherently loopable and transformable in comparison to the static waveform files in Redbook audio. This allowed sound directors to create real adaptive music in games without using more disk space containing the same music altered in a certain way. Because the MIDI commands initiate the instrument (pc soundcard) sounds, transitions between passages can also be implemented very fluently. In comparison, Redbook audio doesn’t transition between parts that easily, for each transition has to be recorded separately. Instead, games usually use techniques such as hard cuts and hard fades to transition between music parts, where the first part of the track is just stopped or faded out, before the next one is started or faded in.

The big disadvantage of MIDI or other sequenced digital music is that the sound quality of the songs is determined by the sound chips incorporated in the sound cards. These sounds sound more digital than actual sampled music recordings. These digital sounds, sometimes referred to as the 8-bit era, have built up a whole cult of followers that are enthusiastic about the distinct and recognizable sound aesthetic. However, the creation of a soundtrack for a game which can compete with soundtracks created for modern hollywood movies need a higher level of sound fidelity, for it has to be suited for a much broader audience than just 8-bit game fanatics.
This is where Redbook audio for games comes into play. Because of technological advancements in streaming speeds, available disk space and compression techniques, Redbook audio has become the new standard for game music, for it allows very high quality music that is indistinguishable from hollywood movie soundtracks. The big challenge that exists within non-linear media is that the static Redbook audio has to get as dynamic as was possible with MIDI sounds.

A.2 Non-linear audio and Intelligent control

A good example of one of the first “intelligent control” systems can be found in the iMuse software by LucasArts developed in 1991.[3] iMuse uses songs indexed with markers and hooks to create adaptive audio. For instance, when the player walks from one area to another, the hook to the transition between the two areas is queued for the next beat. After this the new area’s music is queued, creating one consistent soundtrack without any hiccups, fades or hard cuts in the song.

iMuse provided a way for sound designers to create decision points within songs, effectively creating music that could be branched based on the state of the game. Not only could the music jump from sequence to sequence, it could also alter the content by e.g. transposing the song in a different key or changing the volume of the sequence. This truly distinguished the music in iMuse enabled games from linear musical media.

Another form of intelligent control is found in open form music enabled games. In addition to creating static markers for specific transitions, the sequence of the musical parts is randomly generated by the game. A good example of this kind of free form music is found in the game The Legend of Zelda: Ocarina of Time. In this game, the main theme consists of several sequences that were played back in a random order. This way, each time the game is played the song plays differently. The main drawbacks of open form music in games is that it is hard to create an “interesting” music piece which has some form of suspense and dramatic curve. If the music has to be consistent with the story that is being told, it is very hard to create generic sequences that are appropriate for all the parts of the story.

A.3 Digital signal processing

Unlike MIDI music, Redbook audio is only malleable in the studio until it is frozen into a recording. After recording the audio into this frozen state, it has to be processed and altered in some way to get a dynamic effect from the music instead of just playing back the recording. In game development this kind of post production processing is called Digital Signal Processing (DSP).

The first step in making the music more dynamic is by splitting the composition into different sections that can be looped within the game to cover the variable time that has to be spanned within a game sequence. However, technology in dynamic audio in games is constantly evolving, squeezing more and more dynamic content from frozen high quality recordings.

One of the tricks used in the Metal Gear Solid series to create dynamic music is to create one soundtrack for a certain part in the game, with several parts in it that reflect the different states within the game. For instance, a soundtrack for a certain area might be a sneaky ambient like soundtrack, which transforms into a more higher paced excited piece of music when the player is discovered by the enemy. By splitting up the soundtrack in several parts for different game states the music can quickly react on the players actions. Techniques like these have been extended more in later versions of the game where even the dialog of the players is recorded in different emotional states to make sure no playthrough of the game plays the same combination of dialog.

With hardware evolving and offering more processing power for audio, realtime DSP can be implemented to “render” music in a more realistic fashion. In the Halo game, all
sound effects and dialog are positioned in 3d space, and are subject to occlusion, obstruction and Doppler by using filters that manage the gain and different frequencies. This produces a realistic effect in situations where there is a piece of geometry between the source of the sound and the player. Another function in the Halo game called Head Related Transfer Function (HRTF) even keeps the direction the player is facing in mind when positioning the sound.

A.4 Ducking systems

Prioritization of sounds plays an important role in the way the games audio is perceived by the player. In a big action scene with a lot of loud sound effects, it is important to mix all these sounds in such a way that the most important sounds such as the instructive dialogs are not overwhelmed by less important sound effects such as the engine sound of a car. This “sound cluttering” in games are called logjams. To prevent logjams from occurring, so called ducking systems are used in games to get control over the volume of sounds.

The audio designers of the game Battlefield Bad Company lift ducking systems to an even higher level. Instead of just assigning static rules that prioritize the sounds, they have developed a “sophisticated run-time system that deploys the available dynamic range intelligently moment by moment, logically selecting and balancing the loudest sounds for each player”[4]. By constantly monitoring the complete sound range it is possible to dynamically fit in important sounds without stepping down other sounds to a fixed level, creating a constant and full sound experience whether the player is in a crowded situation or in a regular situation.

A.5 Software and Tools

Throughout the years, tools and software for (game) audio design have evolved and expanded rapidly. Game designers were limited to very restricted resources in early years of game development; audio chip functionality in early video game consoles was very limited and music had to be programmed by hand which was a time consuming task. This section will elaborate on some audio tools from the past, as well as some modern software packages that are used in modern audio production.

A.5.1 Conversion tools

In early video game consoles music had to be programmed in assembly to directly address the audio chip that produced the sounds[3]. This was a very cumbersome task and took programmers a lot of time. Because music could only be programmed, this task wasn’t done by composers. Programmers would often take their favorite music and simply program it for the sound chip.

As discussed before, MIDI made a great change in music production for this universal sequence language was far more accessible for composers. By the late 1980’s tools were written to convert MIDI language compositions into sound chip machine code[3]. Because MIDI had preset midi instruments, conversion tool writers had to map the MIDI instruments to sounds the audio chip could produce and create an audio representation for the various MIDI parameters.

A.5.2 MOD Manipulation

The 1987 game console “Amiga” sported a sound chip called Paula, featuring four-channel 8-bit stereo sound. Because there weren’t a lot of utilities initially available, most game developers created in-house programs to sequence music. One of these programs was called “Ultimate SoundTracker”, created by Karsten Obarski[3]. This program stored files in the MOD format, which was easily adaptable for games, incorporating volume changes, jumps to sequences, starting and stopping instruments and so on.
The biggest advantage of the MOD format over MIDI was that digital samples could be used. These samples were always played back as the sound designers designed the sound. This ability offered far more realistic sound effects and greatly expanded the possible sounds that the chips could produce in addition to the preprogrammed MIDI sounds.

A.5.3 MusyX

Composing music for Nintendo 64 consoles was done with several available tools, one of which was MusyX[3]. MusyX offered its own programming language called SMaL which could manipulate the playback of the stored samples using MIDI controller values, quite like the MOD format. MusyX had a lot possibilities regarding markers and sequences, like the iMuse program. These markers and sequences could be used to dynamically jump to certain sections of the soundtrack. Also, MusyX allowed multiple sequences to be played at the same time to allow cross-fading and layering of sounds.

A.5.4 ISACT

A more modern toolset is the Interactive Spatial Audio Composition Tool (ISACT) from Creative[10]. ISACT is a run-time mixing environment that focuses on multichannel speaker playback systems. The ISACT was one of the first toolsets that could be used in conjunction with a 3D audio engine. By using a visual interface, audio designers could make decisions on the way the sounds would be played back. ISACT is proprietary software that can be acquired through developer affiliation with Creative.

A.5.5 Wwise

Wwise (Wave Works Interactive Sound Engine), introduced in 2006 by Audiokinetic[10], offers developers to rapidly prototype audio integration into games through simulation outside the game engine. A lot of sound engine features such as realtime environmental effects, sound occlusion, sound obstruction and sound prioritization for real-time mixing are built in to the engine. Advanced features such audio profile validation for all current generation video game platforms help developers keep track of CPU and memory usage.

A.5.6 fmod

The interactive audio middleware “fmod” by Firelight Technologies is a library and toolkit for the creation and playback of interactive audio[21]. Because of its support for all next generation video game console platforms and easy to use software it has quickly gained a strong worldwide reputation. By providing a ready-to-use sound engine, game developers have instant access to DSP effects, advanced compression algorithms and other advanced mixing capabilities such as HRTF (see section A.3). Thousands of simultaneous sounds can be placed in 3D space while running on minimal hardware, relieving game developers from mixing and ducking the audio manually.

A.5.7 Pro Tools

Pro Tools 8 by Avid is one of the many popular audio sequencing and production programs[23]. By combining MIDI sequencing, virtual instruments, audio loops and advanced recording possibilities, programs such as Pro Tools offer a complete package to record foley, dialog and soundtracks and create large and complex mixes which can be used in any movie, sound or game production.
A.5.8 Nuendo

The software program Nuendo by Steinberg[22] is an audio post production tool that can be used to further process audio productions with vocal correction, surround mixing features, creating different mix versions and so on.

Another well known technique to localize movies and games is called Automated Dialog Replacement (ADR). Using the Nuendo ADR workflow, original dialogue can be easily re-recorded in another language or more intelligible dialogue track, guided by an accompanying video or other visual queue.