Rob Bridgett
Email: rbridgett@radical.ca
Web: www.sounddesign.org.uk

Dynamic Range: Subtlety and Silence in Video Game Sound

Introduction

With the increasing cinematisation of visuals, game-play and consequently audio in recent console based video games, several essential elements of game sound have been overlooked. Primarily, from an aesthetic standpoint, a competent use of subtlety and silence is distinctly missing from video game audio. Lack of dynamic range, over-compression of sound and music assets, not to mention narrative notions of tension and release are among elements that video game sound can more deeply exploit in order to achieve a more cinematic feel.

This essay will examine production practices in which sound, music and dialogue are over-compressed. It will also centre on the challenges of getting narrative tension into game-play and being able to leverage this for sound.

The focus will then shift onto how these challenges can be potentially solved with the advent of software DSP inherent in the architecture of next generation consoles such as the Playstation 3 and the Xbox 360. In particular, this closing section will concentrate on how real-time DSP effects and in-game mixing can be used to dynamically remove, reduce and prioritize sound in an interactive environment.

Dynamics Crisis

When one examines the dynamic range of the audio output from video games, one can clearly observe that they are heavily compressed, and often overloaded. There are many reasons for this, chief among which are the demands that every sound in the game needs to be audible to the player at particular moments during game-play; dialogue, music and sound effects all compete with one another on a seemingly equal footing. This reflects a similar attitude to compression and limiting that is evident in popular music, FM radio commercials/transmissions and television commercials. When one compares these heavily processed signals to waveforms from cinema soundtracks or classical music recordings, the difference in dynamic range can be easily visually recognized (compare figures 1 and 2). By physically attending orchestral music concerts one can reveal the true extent as to how much dynamic range we have lost in this age of digital reproduction.

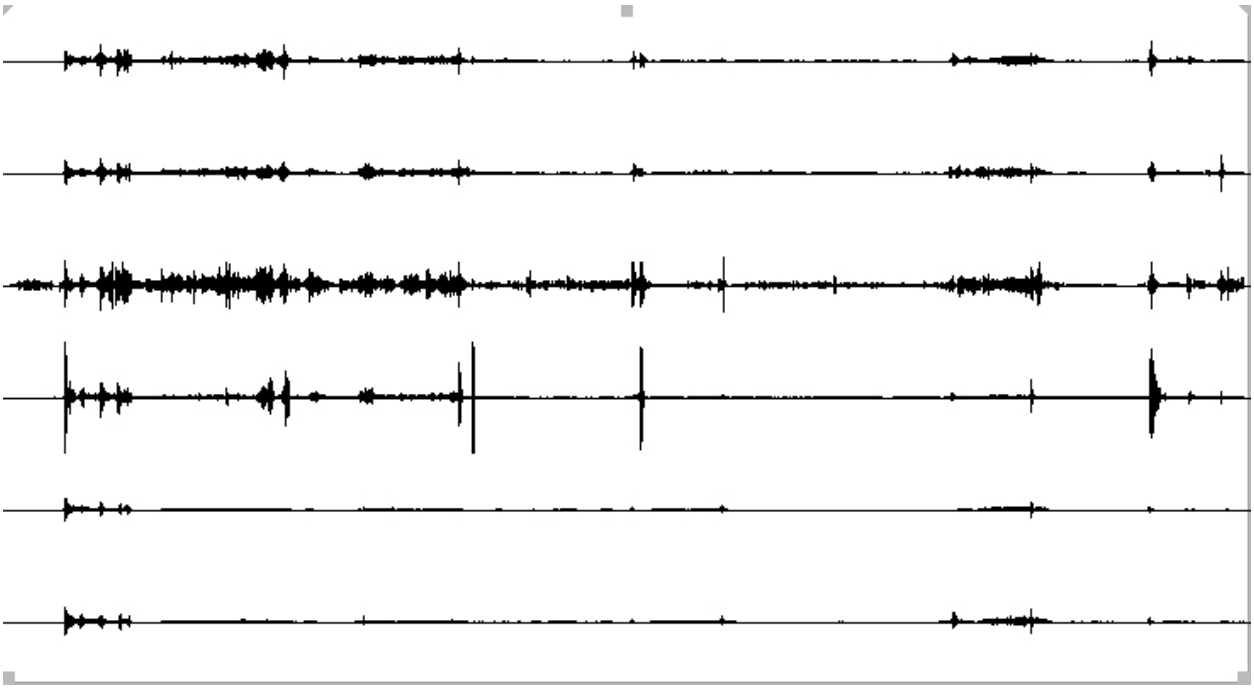

Figure 1 – The 5.1 wave form for a recent Hollywood Movie (from top to bottom L, R, C, LFE, LS, RS)

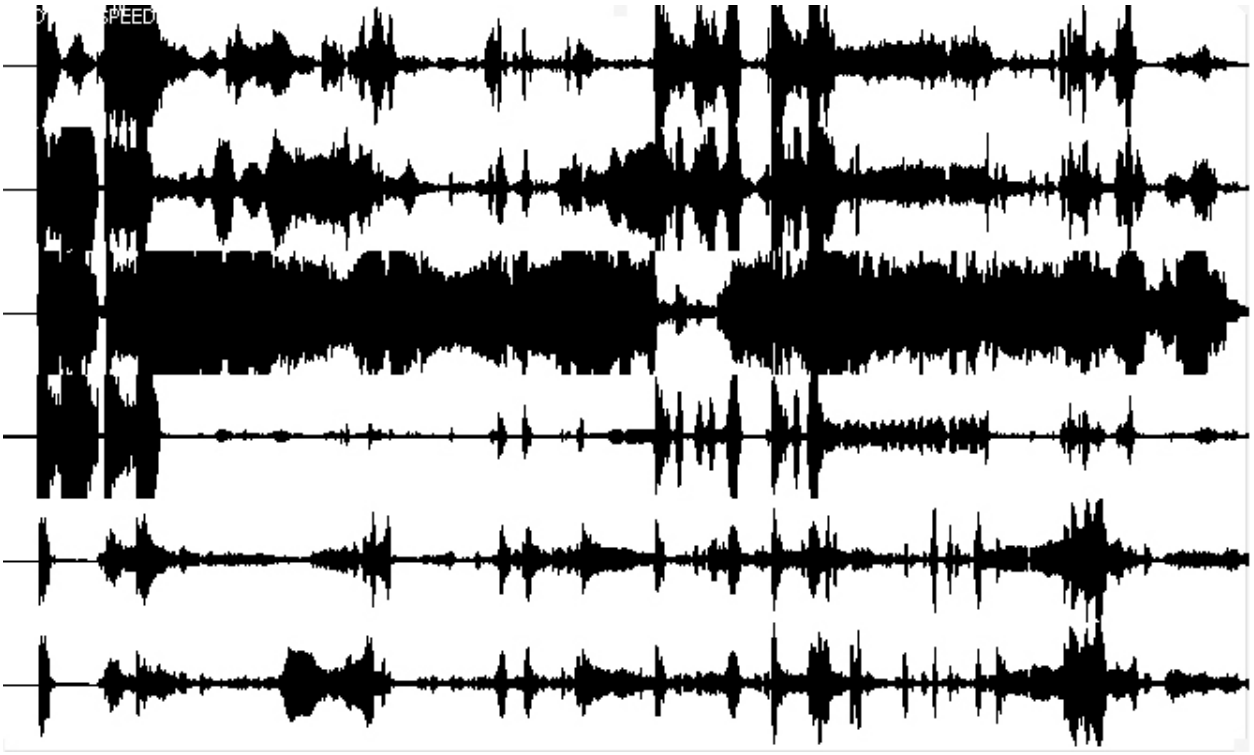

Figure 2 – The 5.1 waveform from a recent Xbox 360 title (from top to bottom, L, R, C, LFE, LS, RS)

One possible cause is that the expected playback environment for popular forms of music and entertainment (the nightclub or the home) and for video games is, or *was* relatively similar. Video games were born in arcades where they had to compete with the sounds of other nearby games consoles for the paying public's attention. At the same time as the rise of the arcade machines, various home game systems also gained popularity. The sounds of games in the home have also had to compete with many other domestic sounds in the home, much as the television does. However, cinematisation of video games is changing this cultural positioning. Classical music and film soundtracks by comparison have a cleanly and culturally established designated attention space, what one could call a 'listening etiquette'; meaning that they are designed and expected to be listened to in isolation and given the audience's undivided attention, they are not expected to compete simultaneously with other sounds – in fact any extraneous sounds to the performance or film are considered as a woefully distracting noise and irritation, no matter how tiny. By adopting a 'film sound' or 'classical music' model to game dynamics there of course needs to be the prerequisite that the game is to be enjoyed specifically as a 'cinematic' and wholly immersive experience without interruption. As stated before, there are only certain kinds of games that demand this kind of immersion and attention, God of War, Gears of War, and the Silent Hill series are good cases in point. Many games such as party games or children's games actually demand the opposite cultural environment in that they need to be heard over other sounds and potentially other participatory voices in the room, and in this case dynamic range should be treated differently. However, there is an increasing trend among console-based video games of the former category, meaning that the expectation is for these games to tell stories and to reveal narrative events to the player that rely on a filmic visual language and that immerse the player in the game world. There has of course simultaneously been a trend towards the cinematisation of video game audio. Brands like THX and Dolby and their involvement in pushing game sound quality and surround technology into the 5.1 and 7.1 realms is well documented, while the use of name film composers, movie actors and movie sound designers all point to a similar aesthetic convergence within the worlds of games and movies. However, aesthetically the convergence is still a long way from being fully realized and dynamics are a key to this.

Dynamics of Narrative

The dynamics of any game-play narrative can be easily plotted on a linear graph. Each game-play element, or mission, can be plotted showing how the narrative of that mission or event will work. Simple game-play mechanics can be drawn as a simple linear curve going from easy to difficult. A similar curve can be applied to the game structure as a whole. Each of the elements of the narrative will be matched by intensity of audio and of user interaction. The audio aesthetics of these narrative dynamics can either play with the action game-play curves or play against them. They can even begin to set up expectations in the matching of audio action to

game-play action, and then break those rules as the narrative progresses to provide even further excitement and immersion in game-play.

Simple ways to think about the dynamics of audio are by looking at survival horror games that have successfully taken the horror movie genre's narrative dynamics and mapped game-play onto these curves. Often leading with audio, the viewer's expectations are manipulated within the horror genre. Silent, or extremely quiet moments where characters (players) are listening to the world around them often precede extremely loud and violent moments in the horror movie genre. This is completely by design, and survival horror games such as Silent Hill have taken the way that sound leads and incorporated it into a variety of game-play elements. Non-diegetic sounds are used well in this type of game to play to evoke the elements of unseen horrors within the player's imagination. Their use of visual darkness and fog effects is also exemplary within video games in allowing sound to develop an often unrevealed world outside what can be seen. Moments of silence are followed by sudden and disturbing attacks from bizarre creatures. While this genre in particular lends itself to these kinds of extreme dynamics, these are certainly ideas that can be picked up within all other game genres.

Racing games, open world games, FPS games can all learn from lessons in the narrative dynamics displayed in the survival horror genre. In order to make an event seem really big, it makes sense that immediately preceding that event is a drop in action and a drop in sound levels; this will make the subsequent sonic barrage perceptually seem so much louder, even when in measured decibels it is not. One common mistake of games designers has in the past been that game-play dynamics are never really plotted out in this way until the game production has been completed. If a simple dynamics curve is applied to every game-play element, with some understanding of how to make things seem more intense by preceding them with low intensity moments, then sound, art and technology teams can use these curves in order to make critical aesthetic and technical decisions that not only match the curves and are able to deliver the intensity required *when* it is required, but also to play against these expectations, and magnify their effects on the player's senses.

It is this ability to draw in the audience, or player, with sound where dynamics begin to be fully realized in an artistic way. Without a suitably defined dynamic range game-play curve within which to do this, it is an up-hill struggle to deliver the same high and low dynamic moments within a game as there already exists in cinema or in classical music.

Games As Cultural Artifact

One way to help define a clearer position for the dynamic expectations of a game's audio is by thinking of the game as a cultural artifact – is it high-art or is it pop-art? Most games probably have elements of both, but on examination of the high-level goals of the game, examining visual style and intended audience, one can easily define the intended experience of the game relatively early in production. If a game is considered as a cultural artifact from day

one of its production (which most games are, even if not consciously), then the expectations for its reception can be easily determined. The type of cultural artifact a game aspires to be ultimately tells you about the kind of environment it is expected to be played in. Is the game a pop-culture game that encourages group play and is expected to be experienced in a noisy home environment? Is the game intended to be taken more seriously? Is the player expected to immerse themselves fully in the game with as little extraneous noise and interference as possible? Once this is determined, then a reasonable and informed approach to the dynamic range of the sound can be determined.

Interactive Mixing

One of the main reasons that so much sound, music and dialogue has been limited and compressed so harshly is that there has been, until recently, a complete lack of sophisticated mixing available in video games. On the older generation platforms such as the PS2 and the Xbox sound memory was limited enough to prevent hundreds of sounds from being played at the same time. In many cases, and there are numerous examples, there were still actually too many sounds playing at the same time on these consoles, resulting in an unintentional cacophony of sound. If a sound needed to cut through in this climate and always be reliably heard – that particular sound, or group of sounds, had to be limited and compressed to extremes in order to compete with other sounds and music– it can be argued that this process of limiting and compressing is simply because a sophisticated enough approach to ducking levels of other sounds did not exist. In more recent cases where simple mixing technically did exist, the ways of implementing it, via a text-script file, were prohibitive enough for it not to be artistically useful.

The shortfalls imposed by limited sound memory on older consoles were often a blessing in disguise for sound designers, even though they probably didn't think of it as such at the time. By limiting what sounds could be in memory at any one time, and limiting the number of voices available to play back these sounds, there were very strict limits placed upon the amount of sound that could physically be played back in any particular scene.

With the increased memory and available voices of next generation consoles and more sophisticated audio compression codecs, such as Microsoft's XNA and Sony's ATRAC allowing for a reasonable quality sample rate at roughly 10:1 compression, the amount of sounds that can now be played simultaneously has increased roughly ten-fold. This means a heavily increased role for in-game run-time mixing in order to prevent a cacophony of sound effects from playing back during run-time. Assigning a finite amount of available voices is one particularly crude way around this, but there emerge problems of really important sounds not being played back because other important sounds are also playing. Mixing in video games, as in cinema, is concerned with sound removal and subtraction rather than pushing volumes up. In mixing, it is often very subtle fader ducking that is required to allow a more important group of sounds to be made prominent above others.

When mixing a video game, the first thing that one notices is that video games have become very loud. This is due in great part to a lack of successful calibration and reference levels available to game sound designers and mixers. Movies, by contrast, have very strictly defined and calibrated output levels and much can be learned from this medium in terms of mixing and calibrating. THX Games mode certainly helps a great deal in allowing the sound designers and mixers to know that what they are hearing is a true image. In video games, it is often the case to compare the game being worked on to another finished game, in order to gauge roughly if the volume levels are similar. More often than not, the games being compared to were not mixed at reference level and were also based on other games, and there has been a tendency to make each successive game louder than the previous one, further resulting in crunching of the dynamic range. Pulling sounds back and defining maximum levels from the outset is the first step that needs to be taken for video games, once the loudest possible sound output has been defined, it is then a case of pulling everything else back from that level.

Once the maximum volume headroom is defined it is usual to be ducking extraneous sounds as and when the relevant in-game events occur within the game-play. For example, with dialogue, rather than having to be compressed and limited so that it can be audible over gunfire in the game, and thereby making the dialogue extremely loud when there is no gunfire, it makes more sense from a dynamics point of view to duck down the gunfire slightly and perhaps any ambience or other extraneous enemy dialogue in order that the player receives the right dialogue with clarity.

One major difference with video game mixing, over that of film, is that it is interactive - meaning that snapshot mixers are installed at run-time to coincide with events that are occurring in the game rather than having linear mix automation. Mixer automation, in the interactive sense, is attached to the interactive *events* themselves, rather than curves over linear time. There are a variety of attenuations aside from volume that can be performed via interactive mixers in a next-generation console game, such parameters as high or low-pass filters, LFE levels and any kind of DSP effects. This is one of the many technical advances that have been brought about by having software DSP able to function at run-time on the next generation consoles. In addition to this, having enough spare memory to allow for a great deal of multiple dynamic volumes to be adjusted at run-time has a profound effect on the aesthetics of game audio.

These advances enable sound designers the potential to work with professional mixer personnel at the end of the project when all game design, art and the majority of game-play code has been debugged and locked down, and to concentrate on a full mix for the entire game. In this sense the dynamics of the game's audio no longer need to be considered at the level of the individual sound effect, but can in fact be defined and re-considered at the mixing stage. The realm of a post-production audio mix can be used to strengthen and highlight an audio direction for the game, as well as making critical decisions about the aesthetic elements of the audio at this stage when all the components are together and functioning as the game has been designed.

Interactivity: Giving Gamers Control Over Dynamics

Ultimately, the playback environment in which a game is heard, no matter how it is posited as a cultural artifact, will be determined by only one person, the player. Because the amount of dynamic compression used by the game has such a profound affect on the overall sound, it is entirely possible that having an option of a software based run-time dynamics compressor or limiter running on the audio outputs of the game (that could be turned on or off in the sound options menu by the player) would solve any issues of adapting the dynamics of a game-play environment to the way the dynamics sound in the future. This would allow the user to set the playback situation themselves, it would also cater for an audience who doesn't have a calibrated 5.1 entertainment system so they would be able to switch on the compressor and boost all the quieter moments in the game. Many of these simple user-initiated compressors already exist on digital cable television packages. At least this method allows for an adaptable approach to the dynamics of game sound by recognizing the varied playback environments available to gamers. These choices should eventually always be made by the gamers themselves, although the default setting shipped with the game can always be the one recommended by the sound designers.