

IEZA: A Framework For Game Audio

by Richard van Tol, Sander Huiberts

Surprisingly little has been written in the field of ludology about the structure and composition of game audio. The available literature mainly focuses on production issues (such as recording and mixing) and technological aspects (for example hardware, programming and implementation). Typologies for game audio are scarce and a coherent framework for game audio does not yet exist.

This article describes our search for a usable and coherent framework for game audio, in order to contribute to a critical discourse that can help designers and developers of different disciplines communicate and expand the borders of this emerging field.

Based on a review of existing literature and repertoire we have formulated a framework for game audio. It describes the dimensions of game audio and introduces design properties for each dimension.

Introduction

Over the last 35 years, game audio has evolved drastically -- from analogue bleeps, beeps and clicks and crude, simplistic melodies to three-dimensional sound effects and epic orchestral soundtracks. Sound has established itself as an indispensable constituent in current computer games, dynamizing¹ as well as optimizing² gameplay.

It is striking that in this emerging field, theory on game audio is still rather scarce. While most literature focuses on the production and implementation of game audio, like recording techniques and programming of sound engines, surprisingly little has been written in the field of ludology about the structure and composition of game audio.

Many fundamental questions, such as what game audio consists of and how (and why) it functions in games, still remain unanswered. At the moment, the field of game studies

¹ Making the gameplay experience more intense and thrilling.

² Helping the player play the game by providing necessary gameplay information.

lacks a usable and coherent framework for game audio. A critical discourse for game audio can help designers and developers of different disciplines communicate and expand the borders of the field. It can serve as a tool for research, design and education, its structure providing new insights in our understanding of game audio and revealing design possibilities that may eventually lead to new conventions in game audio.

This article describes our search for a usable and coherent framework for game audio. We will review a number of existing typologies for game audio and discuss their usability for both the field of ludology, as well as their value for game audio designers. We will then propose an alternative framework for game audio. Although we are convinced frameworks and models can contribute to a critical discourse, we acknowledge the fact that one definition of game audio might contradict other definitions, which, in the words of Katie Salen and Eric Zimmerman "might not be necessarily wrong and which could be useful too" (2004, p.3). We agree with their statement that a definition is not a closed or scientific representation of "reality".

We initially focus on a useful categorization of game audio within the context of interactive computer game play only. The term "game audio" also applies to sound during certain non-interactive parts of the game -- for instance the introduction movie and cutscenes. It concerns parts of the game that do feature sound and interactivity as well, but do not include gameplay, like the main menu. It even includes applications of game audio completely outside the context of the game, such as game music that invades the international music charts and sound for game trailers. We intentionally leave out the use of audio in these contexts for the moment, as there might be other, more suitable, frameworks or models to analyze audio in each of these contexts -- for example, film sound theory for an analysis of sound in a cutscene.

Typologies for game audio

Several typologies and classifications for game audio exist in the field. The most common classification is based on the three types of sound: speech, sound and music which seems derived from the workflow of game audio production, each of these three types having its own specific production process. Award-winning game music composer Troels

Follman (2004) extends this classification by distinguishing vocalization, sound-FX, ambient-FX and music and even divides each category into multiple subcategories.

Although these three terms are widely used by many designers in the game industry, a classification based on the three types of sound does not specifically provide an insight in the organization of game audio and says very little about the functionality of audio in games.

A field of knowledge that is closely related to game audio is that of film sound. A commonly known film sound categorization comes from Walter Murch in Weis and Belton, (1985: 357). Sound is divided into foreground, mid-ground and background, each describing a different level of attention intended by the designer. Foreground is meant to be listened to, while mid-ground and background are more or less to be simply heard. Mid-ground provides a context to foreground and has a direct bearing on the subject in hand, while background sets the scene of it all. Others, such as film sound theoretician Michael Chion (1994), have introduced similar "three-stage" taxonomies.

We foresee that this classification can play an important role in the recently emerged area of real time adaptive mixing in games, which revolves around dynamically focusing the attention of the player on specific parts of the auditory game environment. However, these three levels of attention provide no insight in the structure and composition of game audio.

Friberg and Gardenfors (2004, p.4) suggest another approach, namely a categorization system according to the implementation of audio in three games developed within the TiM project.³ In their approach, audio is divided according to the organization of sound assets within the game code. Their typology consists of avatar sounds, object sounds, (non-player) character sounds, ornamental sounds and instructions.

Besides the considerable overlap between the categories of this categorization (for instance, the distinction between object sounds and non-player character sounds can be rather ambiguous), this approach is very specific to only specific game designs. It says very little about the structure of sound in games.

³ A project that researched the adaptation of mainstream games for blind children: <http://inova.snv.jussieu.fr/tim/>

Axel Stockburger (2003) combines both the approach of sound types and how sound is organized in the game code, but also looks at where in the game environment sound is originating from. Based on his observation of sound in the game Metal Gear Solid 2, Stockburger differentiates five categories of "sound objects": score, effect, interface, zone and speech.



Although Stockburger is not consistent when describing categories of sound on one hand (zone, effect, and interface) and types of sound on the other (score [or music] and speech), the approach of looking at where in the game environment sound is emitted can help distinguish an underlying structure of game audio. The three categories of sound (effect, zone, interface) are very close to a framework and therefore a good starting point. But in order to develop a coherent framework, a clear distinction between categories of sound and types of sound is needed.

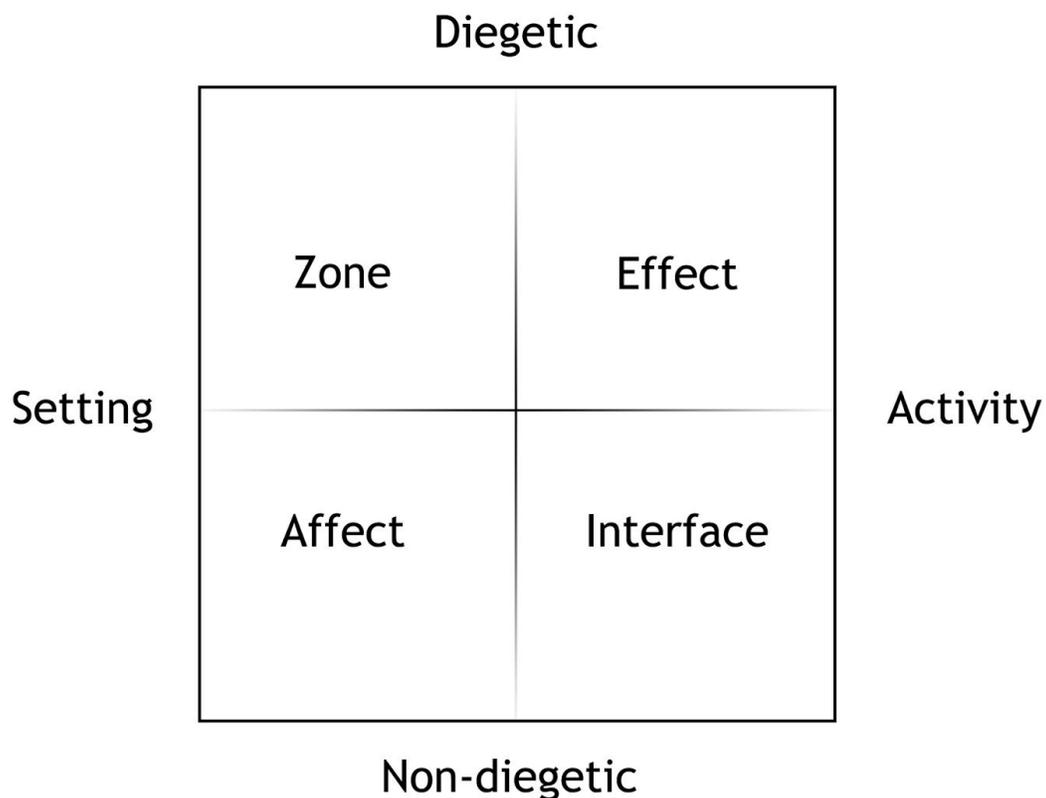
We may conclude that the field of game theory does not yet provide a coherent framework for game audio. Current typologies say little about the structure of game. Designers and researchers have not yet arrived at a definition of sound in games that is

complete, usable and more than only a typology. In the following paragraph we will present an alternative framework for game audio.

The IEZA framework for audio in games

Based on our review of literature and repertoire we have formulated a framework that uses an alternate approach to classify game audio: the IEZA framework. The primary purpose is to refine insight in game audio by providing a coherent organization of categories and by exposing the various properties of and relations between these categories.

The categories and dimensions of the IEZA framework will be described in the following paragraphs and are represented in the following illustration:



The first dimension

On one hand, the game environment provides sound that represents separate sound sources from within the fictional game world, for example the footsteps of a game character in a first-person shooter, the sounds of colliding billiard balls in a snooker game, the rain and thunder of a thunderstorm in a survival horror game and the chatter and clatter of a busy restaurant setting in an adventure game.

On the other hand, there is sound that seemingly emanates from sound sources outside of the fictional game world, such as a background music track, the clicks and bleeps when pressing buttons in the Heads Up Display (HUD), as well as sound related to HUD-elements such as progress bars, health bars and events such as score updates. In other words, sound originating from a part of the game environment that is on a different ontological level as the fictional game world.

Stockburger (2003) was the first to describe this distinction in the game environment and uses the terms diegetic and non-diegetic. These two terms originate from literary theory, but are used in film sound theory as well (for instance by Chion (1994, p.73)). When they are applied to game environments, one has to consider the fact that games often contain non-diegetic elements like buttons, menus and health bars that are visible on screen⁴. Film rarely features non-diegetic visuals and even if it does, these visuals are not often accompanied by sound.

The diegetic side of the framework

Effect

The diegetic side of the IEZA framework consists of two categories. In the first category, named Effect, audio is found that is cognitively linked to specific sound sources belonging to the diegetic part of the game. This part of game audio is perceived as being produced by or is attributed to sources, either on-screen or off-screen, that exist within the game world. Common examples of the Effect category in current games are the sounds of the avatar (i.e. footsteps, breathing), characters (dialog), weapons (gunshots, swords), vehicles (engines, car horns, skidding tires) and colliding objects.

Of course, there are many games that do not feature such realistic, real-world elements and therefore no realistic sound sources. Examples are games such as Tetris, Rez and New Super Mario Bros. The latter features only a few samples of speech (that of the characters Mario and Luigi) while the rest of the audio consists of synthesized beeps, beeps and plings. These non-iconic signs refer to activity of the avatar Mario and events and sound sources within the diegetic part of the game and we therefore consider these part of the Effect category. Sound of the Effect category generally provides immediate response of player activity in the diegetic part of the game environment, as well as immediate notification of events and occurs, triggered by the game, in the diegetic part of the game environment.

Sound of the Effect category often mimics the realistic behavior of sound in the real world. In many games it is the part of game audio that is dynamically processed using techniques such as real-time volume changes, panning, filtering and acoustics.⁴

When the terms diegetic and non-diegetic are used in the context of games, one has to acknowledge the fact that non-diegetic information can influence the diegesis, because of interactivity. For example, a player controlling an avatar can decide to take caution when noticing a change in the non-diegetic musical score of the game, resulting in a change of behavior of the avatar in the diegetic part of the game. In some cases, this trans-diegetic process needs to be taken into account when using the terms diegetic and non-diegetic. Yet, diegetic and non-diegetic have more or less become the established terms within the field of game studies to describe this particular distinction in the game environment.

Zone

The second category, Zone, consists of sound sources that originate from the diegetic part of the game and which are linked to the environment in which the game is played. In many games of today, like Grand Theft Auto: San Andreas and FIFA 07, such

⁴ When the terms diegetic and non-diegetic are used in the context of games, one has to acknowledge the fact that non-diegetic information can influence the diegesis, because of interactivity. For example, a player controlling an avatar can decide to take caution when noticing a change in the non-diegetic musical score of the game, resulting in a change of behavior of the avatar in the diegetic part of the game. In some cases, this trans-diegetic process needs to be taken into account when using the terms diegetic and non-diegetic. Yet, diegetic and non-diegetic have more or less become the established terms within the field of game studies to describe this particular distinction in the game environment.

environments are a virtual representation of environments found in the real world. A zone can be understood as a different spatial setting that contains a finite number of visual and sound objects in the game environment (Stockburger, 2003, p. 6). It might be a whole level in a given game, or part of a set of zones constituting the level.

Sound designers in the field often refer to Zone as ambient, environmental or background sound. Auditory examples include weather sounds of wind and rain, city noise, industrial noise or jungle sounds. The main difference between the Effect and Zone category is that the Zone category consists chiefly of one cognitive layer of sound instead of separate specific sound sources. Also, in many of today's games, the Effect category is directly synced to player activity and game events in the diegetic part of the game environment.

Sound design of the Zone category is generally linked to how environments sound in our real world. Zone also often offers "set noise", minimal feedback of the game world, to prevent complete silence in the game when no other sound is heard. The attention (and therefore immersion in the game) of the player can benefit from this functionality.

The non-diegetic side of the framework

Interface

The first category of the non-diegetic side of the IEZA framework, Interface, consists of sound that represents sound sources outside of the fictional game world. Sound of the Interface category expresses activity in the non-diegetic part of the game environment, such as player activity and game events. In many games Interface contains sounds related to the HUD (Heads Up Display) such as sounds synced to health and status bars, pop-up menus and the score display.

Sound of the Interface category often distinguishes itself from sound belonging to the diegetic part of the game (Effect and Zone) because of interface sound design conventions: ICT-like sound design using iconic and non-iconic signs. This is because many elements of this part of the game environment have no equivalent sound source in real life. Many games intentionally blur the boundaries of Interface and Effect by mimicking the diegetic concept. In Tony Hawk's Pro Skater 4, Interface sound instances

consist of the skidding, grinding and sliding sounds of skateboards. Designers choose to project properties of the game world onto the sound design of Interface, but there is no real (functional) relation with the game world.

Affect

The second category of the non-diegetic side of the framework, Affect, consists of sound that is linked to the non-diegetic part of the game environment and specifically that part that expresses the non-diegetic setting of the game. Examples include orchestral music in an adventure game and horror sound effects in a survival horror game. The main difference between Interface and Affect is that the Interface category provides information of player activity and events triggered by the game in the non-diegetic part of the game environment, while the Affect category expresses the setting of the non-diegetic part of the game environment.

The Affect category is a very powerful tool for designers to add or enlarge social, cultural and emotional references to a game. For instance, the music in Tony Hawk's Pro Skater 4 clearly refers to a specific subculture and is meant to appeal to the target audience of this game. The Affect category often features affects of sub-cultures found in modern popular music, but the affects of other media are also found in many games. Because most players are familiar with media such as film and popular music it is a very effective way to include the intrinsic value of the affects.



The second dimension of the framework

As we have seen, the first dimension distinguishes categories belonging to the game world (diegetic) and those who are not belonging to the game world (non-diegetic). But there also is a second dimension. The right side of the IEZA framework (Interface and Effect) contains categories that convey information about the activity of the game, while the left side (Zone and Affect) contains categories that convey information about the setting of the game.

Many games are designed in such a way that the setting is somehow related to the activity, for example, by gradually changing the contents of Zone and Affect according to parameters such as level of threat and success rate, which are controlled by the game activity. We also gain an insight concerning the responsiveness of game audio: only the right side of the framework contains sound that can be directly triggered by the players themselves.

Summary

- The IEZA framework defines the structure of game audio as consisting of two dimensions. The first dimension describes a division in the origin of game audio. The second dimension describes a division in the expression of game audio.
- The IEZA framework divides the game environment (and the sound it emits) into diegetic (Effect and Zone) and non-diegetic (Interface and Affect).
- The IEZA framework divides the expression of game audio into activity (Interface and Effect) and setting (Zone and Affect) of the game.
- The Interface category expresses the activity in the non-diegetic part of the game environment. In many games of today this is sound that is synced with activity in the HUD, either as a response to player activity or as a response to game activity.
- The Effect category expresses the activity in the diegetic part of the game. Sound is often synced to events in the game world, either triggered by the player or by the game itself. However, activity in the diegetic part of the game can also include sound streams, such as the sound of a continuously burning fire.
- The Zone category expresses the setting (for example the geographical or topological setting) of the diegetic part of the game environment. In many games of today, Zone is often designed in such a way (using real time adaptation) that it reflects the consequences of game play on a game's world.
- The Affect category expresses the setting (for example the emotional, social and/or cultural setting) of the non-diegetic part of the game environment. Affect is often designed in such a way (using real time adaptation) that it reflects the emotional status of the game or that it anticipates upcoming events in the game.

4. Discussions and future work

In this article we have described the fundamentals of the IEZA framework, which we developed between 2005 and 2007. The framework has been used at the Utrecht School of the Arts (in the Netherlands) for three consecutive years as an alternative tool to teach

game audio to game design students and audio design students. For two successive years we gave our students the assignment to design a simple audio game.⁵

The framework was only presented to the students of the second year as a design method. We found that the audio games developed in the second year featured richer sound design (more worlds and diversity), better understandable sounds (for instance, the students made a clear separation between Interface and Effect) and more innovative game design (games based on audio instead of game concepts based on visual game design). The students indicated that the framework offered them a better understanding of the structure of game audio and that this helped them conceptualize their audio game designs.

The framework offers many avenues for further exploration. For instance, it is interesting to look at the properties of and the relationships between the different categories. An example of this is the observation that both Effect and Zone in essence share an acoustic space (with similar properties and behavior), as opposed to Interface and Affect, which share a different (often non-) acoustic space.⁶

In many multiplayer games, it is only the acoustic space of Effect and Zone that is shared in real time by players. Such observations can not only be valuable for a game sound designer, but also for a developer of a game audio engine. It is also relevant for designers incorporating player sound in games, because whether or not the player sound is processed with diegetic properties, defines how players perceive the origin of the sound.

An insight we discovered while designing with the framework is that the right side (Interface and Effect) is more suited to convey specific game information such as data and statistics, whereas the left side (Zone and Affect) is more suited to convey game information such as the feel of the game.

The IEZA framework is intended as a vocabulary and a tool for game audio design. By distinguishing different categories, each with specific properties or characteristics, insight

⁵ Games consisting only of sound. For more information please visit: www.audiogames.net.

⁶ Interface is the part of game audio that is the most likely to remind the player that he or she is playing a game. Therefore designers often add diegetic properties such as acoustics or similar sound design to make the interface category less intrusive.

is gained in the mechanics of game audio. We believe the IEZA framework provides a useful typology for game audio from which future research and discussions can benefit.

References

Chion, M. (1994). *Audiovision, Sound on Screen*. New York: Columbia University Press.

Collins, K. (2007). "An Introduction to the Participatory and Non-Linear Aspects of Video Games Audio." Eds. Stan Hawkins and John Richardson. *Essays on Sound and Vision*. Helsinki : Helsinki University Press. (Forthcoming)

Folmann, T. (2004). *Dimensions of Game Audio*. SHHHHHHHHHHHHH audio blog. Retrieved November 5, 2004 from:

<http://www.itu.dk/people/folmann/2004/11/dimensions-of-game-audio.html>

Friberg, J. and Gardenfors, D. (2004). *Audio games - New perspectives on game audio*. Paper presented at the ACE conference in Singapore, June 2004. Retrieved August 2004 from:

http://www.sitrec.kth.se/bildbank/pdf/G%E4rdenfors_Friberg_ACE2004.pdf

Leeuwen, T. van. (1999, October). *Speech, Music, Sound*. London: Macpress.

Salen, K. and Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge, Massachusetts: MIT Press.

Stockburger, A. (2003). *The game environment from an auditive perspective*. In: Copier, M. and Raessens, J. *Level Up, Digital Games Research Conference (PDF on CD-ROM)*. Utrecht, The Netherlands: Faculty of Arts, Utrecht University.

Weis, E and Belton J. (1985). *Film Sound: theory and practice*. New York: Columbia University Press.

Games:

Grand Theft Auto III (PC). Rockstar Games. Released May 21, 2002.

FIFA 07 (PC). EA Canada, EA Sports. Released September 2006.

Tony Hawk's Pro Skater 4 (PC). Aspyr, Beenox. Released Aug 27, 2003.

New Super Mario Bros (Nintendo DS). Nintendo. Released May 15, 2006.

Tetris (Nintendo Game Boy). Nintendo. Released June 14, 1989.

Rez (PlayStation 2). United Game Artists, Sony Computer Entertainment. Released November 21, 2001.

Huiberts, S. and Tol, R. van, (2008). IEZA: a framework for game audio. Retrieved December 1, 2008, from:

http://www.gamasutra.com/view/feature/131915/ieza_a_framework_for_game_audio.php