

On the Functional Aspects of Computer Game Audio

Kristine Jørgensen, Ph.D. student,
Section of Film & Media Studies, Copenhagen University,
kristinej@hum.ku.dk

Abstract: What is the role of computer game audio? What formal functions does game audio have? These are central questions in this paper which seeks to outline an overview of the functionalities of sound in games. Based on a concluding chapter in my Ph.D. dissertation in which formal functions are identified and discussed, this paper will be a sum-up of some of the most crucial points in my current Ph.D. research on the functionality of sound and music in computer games. The research shows that game audio has important functions related to actions and events in the game world, and also related to the definition and delimitation of spaces in computer games.

1 Introduction

This paper concerns the role of computer game audio, and seeks to outline an overview of important functionalities that can be identified in modern games. The present discussions are based on findings in my Ph.D. research and demonstrate a range of different, but related functions of game audio. These are connected to usability, mood and atmosphere, orientation, control and identification. An important prerequisite for understanding the functions that computer game audio has is seeing computer games as dual in the sense that they are game systems as well as fictional worlds [1]. This means that game audio has the overarching role of supporting a user system while also supporting the sense of presence in a fictional world.

The identification of these functions are based on my current Ph.D. research that studies computer game sound and music with focus on the relationship between audio and player action as well as events in games. The study is based on theories about film sound and music [2, 3], auditory display studies [4, 5, 6, 7], and qualitative studies of game audio designers and computer game players. The theoretical and empirical perspectives have together provided the understanding of game audio functionality presented in this paper. However, since my project has focussed on two specific games, namely Io Interactive's stealth-based action game *Hitman Contracts* (2004), and Blizzard's real-time strategy game *Warcraft III* (2002), it is likely that additional functions may be discovered when studying games within other genres. Still the results presented in this paper are diverse, because of the great difference in genre and audio use in the two games in question. However, this paper will also draw on examples from other games.

2 Theoretical Background

As noted above, understanding the functionality of game audio is connected to understanding the dual origin of computer games as 1) game systems that focus on usability, and 2) fictional worlds that focus on the sense of presence in the game environment. When talking about usability in relation to the game system, I want to emphasise that sound has the role of easing the use of the system by providing specific information to the player about states of the system. This idea is supported by auditory display-related theories.

When talking about the sense of presence in a fictional world, I want to point out that most modern computer games are set in virtual environments that depict fictional, virtual worlds. In this context, *fictional world* should be understood as an imaginary, hypothetical world separate from our own which the players are

asked to believe in when playing a computer game. The player must also accept that the fictional world is the frame of reference for what happens in the game. A fictional world may depict a setting that has no real world counterpart and in which non-existent features are present, or it may depict a setting which has a real world counterpart but presents hypothetical events and features. An example of the first is *Warcraft III*'s fantasy world Azeroth that features the existence of dragons, orcs and magic, and an example of the latter is *Hitman Contracts*' world that is very similar to our own by featuring settings called Amsterdam and Belgrade, but in which the main character and his enemies never were existing persons. In both contexts, sound is used to emphasise the fictional world by being connected to sound-producing sources in a similar manner to real world sounds and by contributing to the atmosphere and the dramatic developments in this world. This point is supported by theories of film sound and music.

Film theory traditionally separates between *diegetic* and *extradiegetic* sound. *Diegetic* sound is that which has a perceived source in the film universe, and which the fictional characters consequently are able to hear. *Extradiegetic* sound, on the other hand, are sounds that are part of the film, but which do not seem to have a physical source within the film universe. Thus, extradiegetic sounds cannot be heard by the fictional characters and communicate to the audience by contributing to the mood or drama within the film [2, 10].

However, in computer games, extradiegetic sound often has a different informative role since the player may use information available in extradiegetic sound when evaluating his choice of actions in the game world. In effect, this means that extradiegetic sound has the power to influence what happens in a game, while it does not have this power in a film. An example of this is the use of adaptive music in games: when a certain piece of extradiegetic music starts playing when the avatar is riding in the forest in *The Elder Scrolls IV: Oblivion* (Bethesda 2006), the player knows that a hostile creature is on its way to attack, and s/he may either try to evade the creature, or stop to kill it. In comparison, when the special shark theme appears when someone is swimming in the thriller film *Jaws* (Spielberg 1975), the spectator can only watch as the character knows nothing of the approaching danger.

Diegetic sounds in computer games may also have a different role than that diegetic sounds in films. When the avatar produces the line "I cannot attack that" when the player uses the attack command in *World of Warcraft* (Blizzard 2004), this is of course a system message, but it also seems that the avatar itself is speaking directly to the player. In this sense, the illusion of the fictional universe is broken because a fictional character is

addressing an entity situated outside the game universe. However, when the traditional concepts of diegetic and extradiegetic spaces seem to break down in games, I call the sounds *trans-diegetic* [11, 12]. It should be noted that transdiegetic sounds are consciously utilized in computer games, where they have a clear functional, usability-oriented role. This will be demonstrated in the following.

Auditory display studies are concerned with the use of sound as a communication system in physical and virtual interfaces, and the field derives from human-computer interaction studies and ecological psychoacoustics. This field utilizes sound as a semi-otic system in which a sound is used to represent a specific message or event.

Auditory display studies often separate between two kinds of signals called *auditory icons* and *earcons*. *Auditory icons* are characteristic sounds based on a principle of similarity or direct physical correspondence and which can be recognized as sounds connected to corresponding real world events; while *earcons* are symbolic and arbitrary sounds such as artificial noises and music which may be seen as abstract in the sense that they cannot immediately be recognized [5, 6, 8, 9]. This separation between two types of signals also applies to computer game audio. When using sound as an information system, computer games utilize both auditory icons and earcons. Broadly speaking, auditory icons are used in connection with all kinds of communicative and source-oriented diegetic sounds, while earcons are used in connection with extradiegetic music and interface-related sounds. In general, these terms are used for non-verbal audio, but in the case of computer games there is an exception to this. When voices are used in order to identify a human source and not for its semantic qualities [3], the voice does not present detailed linguistic information and may be used in a similar manner to other object-related sounds. Examples of auditory icons in games are the sound of a gun shot, the sound of enemies shouting, and the sound of footsteps on the ground, while examples of earcons are the use of music to signal hostile presence, a jingle playing when the avatar reaches a new level in an MMORPG, and the sound playing when Super Mario is jumping.

Concerning the purpose of auditory signals, studies of auditory display often speak of two central functions. These may be described as *urgency* and *response* functions. *Urgency* signals are proactive in the sense that they provide information that the user needs to respond to or evaluate shortly. Urgency signals are often alarms and other alerts pointing towards emergency situations, and may be separated into different priority levels based on whether they demand immediate action or evaluation only [7]. *Response* signals, on the other hand, are reactive, and work to inform the user that a certain action or command has been registered by the system. In order to be experienced as responses, the sound must appear immediately after a the player has executed a command or an action, and it must be clearly connected to a specific event [4, 5]. In a game, an urgency message may be the voiceover message “our forces are under attack” in *Warcraft III*, while a response message may be the sound of a mouseclick when selecting a certain ability from the interface menu in the same game.

Together these concepts form a fruitful framework for understanding why computer game audio is realized the way it is, and it also provides an understanding of different functions that game audio may be said to have. The response and urgency functions explain game audio in terms of the usability of a computer system. In addition, the concept of auditory icon explains

how sounds that seem natural to the game universe also have strong informative value, while the concept of earcon explains provides an understanding of why game music and artificial noises make meaning without disturbance in computer games. More importantly, these ideas also help explain why there are transdiegetic sounds in computer games. When a game developer wants to utilize sound for urgency and response purposes, while also maintaining a direct link to the game universe, it becomes necessary to break the border between real world space and virtual space in order to enable communication between the player and the game world.

3 Different Functions

Related to the above theoretical assumptions, this part of the paper will discuss five different overarching functions that have been disclosed during my research. As noted above, the identification of these functions are based on analyses, interviews and observations related to two specific games, and it is likely that the study of more games will reveal additional functions.

3.1 Action-Oriented Functions

This research has identified uses of game audio which relate to events and player actions in the game world, and which corresponds to auditory display studies’ urgency and response functions. Most modern games utilize sound for these purposes to an extensive degree, although it is not always evident that this is the formal and intended function of the sound. It seems to depend on how auditory icons and earcons are used.

Hitman Contracts integrates auditory icons as naturally occurring sounds from events in the environment. In this sense, the communicative role of the sounds becomes transparent by giving the impression that sounds are present for a realistic purpose instead of a functional purpose. For instance, when the avatar is in a knife fight, sound will be a good indicator of whether he hits or not. When the avatar hits, the slashing sound of a knife against flesh will be heard, accompanied by screams or moans from the enemy, and when the avatar misses, the sound of a knife whooshing through the air is heard. These are of course examples of a *confirmation* and a *rejection response* to player actions, and work as a usability feature although they also seem natural to the setting and the situation.

However, it is also possible to use auditory icons in a less transparent manner, in which the auditory icons more clearly stand out as auditory signals intended for communicating specific messages. In *Warcraft III*, objects produce specific sounds when manipulated. For instance, when the player selects the lumber mill, the sound of a saw is heard. Also, when the barracks is selected, the player hears the sound of marching feet. Although these responses have diegetic sources, the sounds do not seem natural to the game world in the same manner as the knife sounds in *Hitman Contracts*. The reason for this is that they are produced only when the player selects the specific building, and in the case of the barracks, this is not the exact sound one expects to hear at a real-world barracks. We see that the sound is suitable for the specific object, although not in this precise format. According to Keller & Stevens [6], this demonstrates non-iconic use of auditory icons, while the example from *Hitman Contracts* demonstrates iconic use of auditory icons. This difference also emphasises the fact that sounds with a seemingly naturalistic motivation do have usability functions.

Concerning the use of earcons for response purposes, *Hitman Contracts* has music that informs the player whether his/her

current activities are going well or badly. The music changes into a combat theme which will play a particular piece of music if the player is doing well, and another if the player is doing badly. However, although this follows the idea of earcons, this use of music is also adopted from the use of dramatic music in films. This makes the use of musical earcons feel familiar and suitable even though it does not feel natural to a specific setting.

Both earcons and auditory icons are used for urgency purposes in computer games. Although a range of different priority levels may be identified in games, I will limit myself to the two most common. Games often separate between urgency signals that work as *notifications* that do not demand immediate player action; and urgency signals that work as *warnings* that demand some kind of action. Notifications provide information about events in the environment that the player needs to know about, but which s/he does not have to react to. S/he may, however, need to evaluate the situation. An example from *Warcraft III* is the message “work complete” which is played when a worker has finished its task. Warnings, on the other hand, provide information about immediate threats or dangers to the player. These will always need an immediate evaluation, and possibly action, but dependent on the situation, the player may choose to not take any action if he regards the situation under control. An example is the message “our forces are under attack” which is played in *Warcraft III* when the player’s units are being attacked by the enemy.

3.2 Atmospheric Functions

Working in a more subtle manner, the atmospheric functions of game audio may still be regarded as one of the most central. The use of music in films for emotionally engaging the audience is well known [12], and games try to adopt a similar way of using music. Most mainstream games utilize music to emphasise certain areas, locations and situations.

An example is a game such as *World of Warcraft*, where the large cities have distinct music. When entering the orcish capital of Orgrimmar, the player hears that a certain piece of music starts, dominated by wardrums. This music is distinct from the music heard when entering the human capital of Stormwind, which has a more Wagnerian epic style. In both cases, the music is there as a mood enhancer that emphasises classical fantasy conventions of the noble humans and the savage orcs. In this context, it is important to point out that atmospheric function of music is guided by genre conventions. In survival horror games such as the *Silent Hill* series (Konami 1999-2004), atmospheric sound and music are used to emphasise a very specific mood of anxiety and horror. However, it should be noted that this mood also has the power to influence the player’s behaviour in the game. When the player becomes anxious he may act more carefully in order to avoid any dangerous enemies and unpleasant situations. In this sense, atmospheric sound may thus work indirectly to influence player action.

Atmospheric sounds may also influence player behaviour in more direct manners. When music is used for responsive and urgency purposes, it will also have atmospheric properties. In the example from *Hitman Contracts* above, we see that different pieces of music provide different kinds of information to the player. The music does not only work as pure information, it also emphasises mood. For instance, when the player is in a combat situation, the music becomes more aggressive by an increased tempo and a more vivacious melody.

Although music may be the more persuasive kind of atmospheric sound, environmental and object sounds as well as dia-

logue also contribute to the specific mood of a game or a situation. The overall soundscape contributes to a sense of presence or even immersion in a game by creating an illusion of the game world as an actual space. Sound may thus give the impression of a realistic space by presenting virtual offscreen sources. In this context, ambient environmental sound is of interest. Ambience should be understood as environmental background sounds added to the game for the purpose of adding the sense of presence and a specific mood to the game. Thus, these sounds are not present in order to influence player action by giving the player specific information about objects, events or situations, and they are often not connected to specific sources in the game. Instead they may be connected to virtual sources, or be collected into a separate soundtrack. The first technique is found in *Lineage II* (NC Soft 2004), where for instance insects can be heard in each bush. When looking for the actual sources, however, these cannot be found as visual objects. The second technique is found in *Sacred* (Ascaron 2004), where the ambient background noise for each setting is stored as a separate mp3-file. Thus, when the player is exploring dungeons, a specific soundtrack consisting of reverberated wind and running water is played, while when the player visits villages, the sounds of children laughing and dogs barking are heard.

Observations and conversations with players reveal that the engagement in the game may decrease when the sound is removed from the game. Players notice that the immersion decreases, and that the fictional world seems to disappear and that the game is reduced to rules and game mechanics when sound is removed.

3.3 Orienting Functions

The orienting functions of game audio are related to action-oriented functions in the sense that both provide information about events and objects in the game environment although in different ways. While the action-oriented functions are reactive and proactive, the orienting functions inform about the presence and relative location of objects and events. The functions described in this section were identified in my qualitative research where player performance was studied in the absence and presence of game audio.

In connection with the orienting functions of game audio, it is important to note that sound seems to extend the player’s visual perception beyond what is possible without sound. In the presence of sound, the player receives information that the visual system cannot process, such as for instance events and objects situated outside the line of sight. It also enables the player to know what is going on in locations not in the immediate vicinity of the player.

The perhaps most obvious orienting function of sound is that it provides information about the presence of objects as well as the direction of sound sources. This is especially important in the context of offscreen sources. Sound may thus reveal the appearance and presence of an object before the player has actually seen it, and provides therefore information that the visual system could not provide on its own. A good example is the shouting voices of offscreen guards in *Hitman Contracts*. Today’s computer games utilize the stereo channels to inform the player about the relative direction of a sound source. However, although a stereo sound system does reveal the relative direction of a certain source, it is not able to provide information on whether the source is located in front of or behind the player. True surround systems demonstrate significant possibilities for providing detailed information about the location of an offscreen

source, and prove to be interesting for the further development of game audio functionality. These orienting functions are also demonstrated in the research on audio-only games for the blind and visually impaired. This research demonstrates the use of characteristic sounds that identify objects and events, their presence, and their relative location [4, 5, 13].

3.4 Control-Related Functions

Tightly connected to the orienting functions are control-related functions. These are related to the idea that sound extends visual perception, and point to what sound directly contributes to concerning the player's control over the game environment.

Since game audio extends visual perception, it enables the player to be in control over unseen areas. Strategy games often provide good examples of this. In the real-time strategy game *Warcraft III*, the player receives auditory information about events happening on remote areas of the map. When the player is busy leading his/her army to battle, s/he still receives voice-over messages about status of the base, such as "upgrade complete" and "ready to work". These messages contribute to increased control over all activities in the game. The same game also utilizes sound to provide the player with more detailed information than what visuals can provide. Combat situations in this game tend to be chaotic due to the fact that there is a huge number of military units fighting on each side. It is therefore difficult for the player to see exactly what happens in combat. The sounds of bowstrings and metal against metal inform the player what units are fighting, and screams tell the player that units are dying. In this example we see that sound contributes to ease the player's management of the game by providing information that is difficult to provide by visuals only.

These examples are also related to the idea that the presence of sound eases or increases the player's attention and perception. This was suggested by the informants of my study, who emphasised the idea that channel redundancy, or presenting the same information through different perceptual channels, increased the ability to register certain messages [14]. When sound was absent, *Warcraft III* players had difficulties noticing written messages appearing on the screen. This is probably due to the high tempo of the game, and the fact that the player's visual perception is focussed on specific tasks in the game.

3.5 Identifying Functions

Another interesting function connected to sound is its ability to identify objects and to imply an objects value. The fact that sound identifies may not seem surprising, since sound in general indicates its producing source. However, this is utilized in games, not only in the format of auditory icons that automatically are recognized, but also in the format of earcons that needs to be learned before they can be recognized as belonging to a specific source. We have already discussed the example from *Hitman Contracts* where music is used to identify certain situations.

Warcraft III connects identifying sounds to units and buildings. From the player's top-down view on the environment it may be difficult to distinguish objects from each other. However, as noted above, when the player selects the lumber mill, s/he will hear the sound of a saw, and when s/he selects the barracks, s/he hears the sound of marching feet. This enables the player to easily recognize the building without having a clear view of it. In the case of units, each of them presents an utterance of recognition when produced and when manipulated. This means that a worker says things such as "ready to work", while a knight says

"ready for action". However, it is interesting to see that these utterances not only identify the unit; they also signal the relative value of it. This means that the more powerful a unit is, the more distinct its sound of recognition is. Within *Warcraft III*'s orc team, the workers utter sentences that suggest obedience and humbleness such as "work, work", "ready to work" and "be happy too". The named warchief which represent the most powerful units in the game, on the other hand, utter sentences such as "I have an axe to grind", "for my ancestors", and "an excellent plan", which emphasise aggressiveness, honour, and strategic insight. In addition, its voice is deeper than the voices of other units, as well as the fact that the footsteps of the unit sound heavily. Thus, we see that the quality and content of the sound are used in order to ease recognition of certain objects in the game as well as to signal the value of different units.

4 General Discussion

The functions identified above are closely related to each other although they seem to stem from different aspects of games. Most of the functions seem to be motivated by usability, although the atmospheric function seems to go against this by emphasising presence and immersion into the fictional game world. These two seemingly different purposes of game audio are connected to the fact that computer games are user systems at the same time as they are set in fictional worlds. However, it is important to note that computer games also bridge these two domains, something which also becomes evident through their use and implementation of audio.

How, then, does this fusion of user system and fictional world happen? To say it bluntly, it happens through giving many sounds a double function where they belong to in-game sources and are accepted as fiction at the same time as they provide specific information to the player. We can identify three central techniques that ensure that this merge seems transparent and intuitive; namely the use of auditory icons, earcons, and transdiegetic sounds.

Since auditory icons have an immediately recognizable relation to its source, these are very well suitable for combining the usability function with the fictional world. The sounds seem natural to the game environment, at the same time as they provide the player with information relevant for improved usability of the system. This is what hinders the sound from the buildings in *Warcraft III* to seem misplaced.

Earcons may be said to work the other way around, since they illustrate an artificially constructed relation between sound and source. The use of artificial noises may contribute to a certain auditory message becoming very noticeable or even disturbing because of its unexpected relation to a certain source, such as is the case with the squeaking negative response produced when the player tries to make an illegal action in *Warcraft III*. On the other hand, the use of game music does not seem disturbing because it utilizes accepted conventions from film music and adds mood to the game. This is why the player accepts music which changes according to the situation in a game such as *Hitman Contracts*, and which plays in major when the player is doing well and in minor when the player is doing badly.

The third technique that makes the fusion between usability and presence in a fictional world transparent is transdiegetic sounds. Transdiegetic sounds break the conventional division between diegetic and extradiegetic sounds by either having diegetic sources that communicate directly to the player, or by being extradiegetic sounds that game characters virtually can hear.

When sound in films breaks this common separation between diegesis and extradiegesis, it is understood as a stylistic, artistic and uncommon way of using sound, but games utilize this functionally to bind together usability and fictional space. This means that it does not feel disturbing when a unit in *Warcraft III* says “What do you want?” with direct address to the player, although the unit is regarded a fictional character and the player who has no avatar in the game is situated in real world space. Neither does it seem strange that the avatar as a fictional character in *The Elder Scrolls IV: Oblivion* (Bethesda 2006) reacts by drawing its sword when the musical theme that suggests nearby danger starts playing – although a film character would not react in this way, a game character can due to the link between avatar and player.

In this sense, computer game audio aims to combine usability with presence and immersion in the fictional game world, and by doing this the realization and functionality of game audio becomes in different ways similar to both film audio and auditory displays and interfaces. This creates a very unique way of utilizing audio which is especially designed to emphasise how modern computer games work.

5 Summary

As a summary of the concluding chapter of my upcoming Ph.D. thesis on the functionality of game audio in relation to actions and events, this paper has concerned computer game audio functionality. The paper identifies and describes the most important functions of computer game audio and provided an explanation of why these functions are central to computer game audio. The main argument is that modern computer games are set in fictional, virtual worlds at the same time as they are user systems, and in order to combine this in the most transparent way, they break the common concept of diegesis by utilizing auditory icons and earcons for informative purposes.

6 References

- [1] Juul, Jesper, *Half-Real. Video Games Between Real Rules and Fictional Worlds*. Copenhagen: IT University of Copenhagen, (2003).
- [2] Branigan, Edward, *Narrative Comprehension and Film*. London, New York: Routledge, (1992).
- [3] Chion, Michel, *Audio-Vision. Sound on Screen*. New York: Columbia University Press, (1994).
- [4] Drewes, Thomas M. & Elizabeth D. Mynatt, “Sleuth: An Audio Experience”, *Proceedings from ICAD 2000*. Available: <http://www.cc.gatech.edu/~everyday-computing/publications/sleuth-icad2000.pdf> [03.06.2005], (2000).
- [5] Friberg, Johnny & Dan Gärdenfors, “Audio Games: New Perspectives on Game Audio”, *Proceedings from ACE conference 2004*. Available: www.cms.livjm.ac.uk/library/AAA-GAMES-Conferences/ACM-ACE/ACE2004/FP-18friberg.johnny.audiogames.pdf [02.08.06], (2000).
- [6] Keller, Peter & Catherine Stevens (2004): “Meaning From Environmental Sounds: Types of Signal-Referent Relations and Their Effect on Recognizing Auditory Icons”, in *Journal of Experimental Psychology: Applied*. Vol. 10, No. 1. American Psychological Association Inc., 3-12, (2004).
- [7] Sorkin, Robert D., “Design of Auditory and Tactile Displays”, in Salvendy, Gavriel (ed.): *Handbook of Human Factors*. New York, Chichester, Brisbane, Toronto, Singapore: John Wiley & Sons, 549-576, (1987).
- [8] McKeown, Denis, “Candidates for Within-Vehicle Auditory Displays”, *Proceedings of ICAD 05*. Available: <http://www.idc.ul.ie/icad2005/downloads/fl18.pdf> [10.04.06], (2005).
- [9] Suied, Clara, Patrick Susini, Nicolas Misdariis, Sabine Langlois, Bennett K. Smith, & Stephen McAdams (2005): “Toward a Sound Design Methodology: Application to Electronic Automotive Sounds”, *Proceedings of ICAD 05*. Available: <http://www.idc.ul.ie/icad2005/downloads/f93.pdf> [10.04.06], (2005).
- [10] Bordwell, David & Kristin Thompson, *Film Art: An Introduction*. New York: Mc-Graw Hill, (1997).
- [11] Jørgensen, Kristine, “On Transdiegetic Sounds in Computer Games”, *Northern Lights 2006*, Copenhagen: Museum Tusulanums Forlag, (2006).
- [12] Gorbman, Claudia, *Unheard Melodies? Narrative Film Music*, Indiana University Press, (1987).
- [13] Röber, Niklas & Maic Masuch, “Leaving the Screen. New Perspectives in Audio-Only Gaming”, *Proceedings of ICAD-05*. Available: <http://www.idc.ul.ie/icad2005/downloads/fl109.pdf> [02.08.06], 2005.
- [14] Heeter, Carrie & Pericles Gomes, “It’s Time for Hypermedia to Move to Talking Pictures”, *Journal of Educational Multimedia and Hypermedia, winter, 1992*. Available: <http://commtechlab.msu.edu/publications/files/talking.html> [03.08.06], 1992.