



Hearing Sounds Through Different Ears: A Video Game Case Study

Gabriel Dargains Gonzaga^(✉)

Federal University of the State of Rio de Janeiro, Rio de Janeiro, RJ 22290-255, Brazil
Gabriel.dargains@edu.unirio.br

Abstract. This article seeks to demonstrate that different listening vantage points and types of listeners are established in open world video games to create meaning and immersion. Using *Dragon Age: Inquisition* (2014) as a case study, I analyze its sound image through Smalley's concept of space-form (2007) to find these points. Even though all sound is spatialized in the sound-image, not all sounds are located in the 3D navigation space. Positioned and non-positioned sounds (in reference to the player character) lead us to understand there's a combination of both peripatetic and fixed listeners during the gameplay. In the case study, by soundwalking in one of the game's zones we conclude there is an overlap of listening vantage points that do not refer to the same space. This helps the players have a holistic view of the soundscape, making it more favorable to express meaning through the ambience's sound design.

Keywords: ludomusicology · ambient sounds · space-form

1 Introduction

Through this article I'll present different listening perspectives in open world video games, using *Dragon Age: Inquisition* as an example. While there are articles about soundscape and soundwalk in video game context¹, I couldn't find literature addressing listening perspectives that we can find while playing a navigable game. I hope to contribute with the argument that these perspectives mean something and that meaning can be manipulated when composing video game ambience. For the identification of these perspectives, that from now I shall call vantage points², I will analyze the space-form in the game's sound image and find that some of the sounds are rendered in relation to the character's location and orientation, while others are ambiguous to both.

¹ Hambleton's take on Westerkamp's Soundwalk opens the interpretation to soundwalk in virtual spaces. She uses soundwalk as a method to map out the soundscape [9]. For other articles using soundscape as a method in analyzing video game sounds, see Galloway [8] and O'Hara [13].

² This is the term used by Smalley in [15], to delineate a point from where the listener will turn his attention to the sounds in space. The reason to not choosing a term like "point of listening", analogous to 'point of view', can be a way to reinforce that the vantage point is also about perceiving or observing space and not simply hearing.

The first chapter will introduce open world video games, vantage points and the concept of sound image. In the second chapter, we will discuss space representation in object-based audio and Stockburger's categorization of sound objects [18] applied to *Dragon Age: Inquisition* (that I'll refer to as DAI for brevity). Then in the third chapter after a brief explanation of Smalley's space-form, I'll present an analysis using some of the concepts to identify the different vantage points in a zone within the game.

1.1 Open World and Vantage Points

Open world video games have players navigating 3D maps through avatars [10]. The player controls a character in that world, through a first person or third person view - in the first person the world is shown as if through the eyes of the character; in third person the player can see the controlled character as an observer, external to its body. Both first person and third person imply a point of view, one where the camera orientation is attached to the character's and another where the camera orientation is independent of the character's orientation. In this video game genre, the receiver of sounds is attached to the camera, so if the camera moves, panoramic visual and sound space also does. In both cases, the camera functions as eyes and also ears into the game world.

Since we have a point of view, we will have a listening vantage point too. Listening vantage points, as described by Smalley, can be fixed, variable or peripatetic in location and orientation [15]. The ones that interest us for this matter are the fixed and peripatetic, fixed being the position (location and orientation [11]) of the listener. Peripatetic meaning the listener has a moving position. The camera-sound receiver attached to the controlled character moves his point of view in the navigable space, so it can be considered as a peripatetic listener. We can't forget the player sitting in front of the screen is playing from a fixed vantage point. Those are the first immediate vantage points in the game experience, they both coexist as long as some sounds are directed to the character and some sounds are directed to the player.

1.2 Space in Sound

Movement and space are characteristics we can find regardless of the listening vantage point. Every sound carries some degree of spatial and causal qualities and information, we naturally imagine or 'feel' these qualities as they appear. While some qualities can be understood through spectromorphological [16] analysis, others are quite difficult to establish in an objective way. The feeling and imagination of the qualities do not belong to the auditory domain, they occur through transmodal perception of the sound³ - when we listen to a sound, we are also virtually feeling it through other senses and that happens with every sensory experience.

Our transmodal perception of the sounds might hint on vision and touch as we often describe sounds relative to its size or texture - where I come from it's not uncommon to

³ 'Transmodal' is not the only term possible, but it is the one Smalley uses. Lars Elleström [7] will refer to it by a sensorial multimodality in the first chapter in the book *Media Borders, Multimodality and Intermediality*. Chion will call it transsensorial [5]. I also recommend reading Rodolfo Caesar's writings on sound's transsensoriality [2] and [3].

hear people calling higher pitch sounds ‘fine’ or ‘tiny’ sounds, but we see transmodal links happening even with commonly accepted terms like grain, wet, dry, etc.

With audio recording (or through memory) we can attempt to inspect the image that forms in our imagination as the transmodal links occur. This image is referred to as sound-image or *i-son* as coined by François Bayle⁴. When discussing space-form, Smalley will break down this sound-image into spatial forms to be able to discuss space within acousmatic music. In video games, as with every audiovisual media, sound and image is produced separately, so all the spatial qualities in sound comes from sound recording, editing and rendering.

2 Representation of Space in Games and Music

Sound in both video games and recorded music will often be heard in similar conditions - except when in concert or performance situations - with audio output devices such as headphones, loudspeakers or television. The sound image in both cases have the same ‘depth’ and dimensions, even with different mixing methods – game engines that deal with 3D space have multichannel audio, what makes the directionality of sound more refined but the frame is still the same, a pair of headphones or loudspeakers. Although the listening conditions are similar, we find that there is a striking difference on how we perceive space through hearing. Spatialization⁵ is quite a complex topic to discuss in detail here, but there is some basic distinction we have to make before proceeding.

Nicolas Tsingos [14] can help us start the discussion by bringing up Roberta Klatzky’s [11] distinctions on spatial representation. He mentions that in order to represent space, we first need a frame of reference and explains them briefly:

[...]An egocentric frame of reference encodes object location relative to the position (location and orientation) of the observer or “self.” An allocentric frame of reference encodes object location using reference locations and directions relative to other objects in the environment. [...]

For interactive rendering, video game applications programming interfaces (APIs) generally express the positions of audio objects as allocentric world-space Cartesian coordinates. The coordinates of the objects may be converted to a listener/egocentric frame of reference at rendering time depending on the player’s position, in particular if a single perspective has to be rendered (e.g., on headphones). (pp.244–245)

By ‘player’s position’ he means the player-character (PC, the character controlled by the player). Since the player themselves doesn’t move much during gameplay, they’ll

⁴ For a more original reference, see [6] chapter 9, footnote 14 at page 259. Chion has a transcription of the definition of the term sound-image.

⁵ *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio* [14] has various articles on spatialization and you can find a more in-depth discussion there. The point is that spatialization creates a more refined space by using virtual outputs in various directions, but the output devices used are mostly the same – the same ‘frame’ or boundary is used.

look at their screens and stay still for the most part. Sounds help create the sense of causal relation between sound object and its visual object, but since that causality is artificial, the proximity of the sounds may be used to simulate distances and virtual locations. For example, an audio object can be near the PC-listener in cartesian coordinates but have qualities of a sound being heard far away.

It's important that he mentioned perspective, because orientation is mostly what tells players where objects are in the surrounding space. In first person view this is literal, but in third person even while the camera moves with the character, both might have independent orientations. In *Dragon Age: Inquisition* the camera's field of view is approximately 100° and it is independent of the character's orientation. The character's egocentric reference is important because the camera can't stray too far away from it, but the camera's reference are the eyes and ears that allows players to experience the game world.

Being able to face the character and the camera in opposite ways helps navigation and a better assessment of the surrounding space, but most of the time the player will want to see where the character is heading. The merging of character and camera's orientation reinforces roleplay, the player recognizes the intimacy between the character and themselves more readily⁶.

2.1 Positioned and Non-positioned Sounds

Ideally, in most cases the listener and player's position will be between both audio outputs and facing a direction perpendicular to the axis left and right in the stereo image. This way both music and video game listener's 'vantage points' makes the sound image an allocentric representation. To simplify the matter, I'll be referring to the audio objects positioned (located and oriented) in world-space's cartesian coordinates and those that are not positioned, as positioned and non-positioned sounds. Positioned sounds are the ones in the world-space that are rendered in reference to the player-character or camera's position (egocentric to the PC-Camera) and non-positioned sounds are the ones that have no relation to the player-character nor camera location and orientation in the game world-space (Allocentric to the player). In DAI's case, the reference to positioning is always the camera.

Positioned sounds can still be acousmatic - sounds whose sources are not seen - if their sources are not in the camera's field of view. They expand the proximity by using the PC's circumspace⁷ and hinting at objects directionality and distance. It becomes easy to spot an enemy if their sound effects are present even if the player isn't directly looking at them. Positioned sounds are the indication of a peripatetic listener and non-positioned indicates a fixed listener. Non-positioned sounds share the same orientation reference as the player, since they are facing the screen and not navigating themselves. They coexist and serve different purposes.

⁶ This intimacy can be thought of as immersion or as Modena and Parisi argues, a *process of becoming*. It is relevant to understand that camera and character's orientation can create or tear down the distance between player and avatar, but the distance is always present causing friction.

Perhaps this is a good point to further the topic.

⁷ "Space around the listener" p.48 – Smalley [15]

The main purpose of the positioned sounds is creating depth and directionality, making navigation possible (directionality includes height too). The main purpose of non-positioned sounds is presenting information directly to the player instead of making it go through the camera reference⁸. But as we'll see, non-positioned sounds often create ambiguities, not making it clear as to which listener the sound is being addressed to. When there is visual navigation but some sounds do not accompany the movement, we can't say that the camera and the player hold the only vantage points.

Camera and character position means nothing if the game world has no objects to be perceived by them. Tsingos refers to audio objects, but in video games they are often attached to entities in the game world, be it a creature, a non-player character (NPCs), representation of inanimate objects or even structures and zones. As he mentions, these objects are generally located in the game world-space and are rendered in relation to an ego (which in this case is the avatar or camera, not the player). We'll see now what kinds of sounds there are in the game.

2.2 Sound Categorization

Stockburger [18] has recognized five types of sound objects in video games according to their function: speech sound objects, effect sound objects, zone sound objects, score sound objects and interface sound objects. I shall explain the categories as we see examples of how they are used in *Dragon Age: Inquisition*.

Speech Sound Objects

The speech sound objects can be found either located as NPC's greetings and short phrases as the character walks by or in cinematic cutscenes and dialogues. The use of the speech objects in cutscenes are not located in the map's coordinates, the voice is centralized in the sound image for the sake of clarity just as in cinema [5]. In greetings, short phrases and dialogues that are not cinematic, they are subject to the PC-camera's position, so they will have more volume as the characters approach each other and will change directionality in the sound image as the player character turns in different directions. During dialogues, player's frequently have to choose how to reply, creating the feeling of choice and participation in the narrative. The answers are set and there might be additional options depending on how you created or progressed the character. Being able to perceive directionality in dialogues makes the players feel as if they are part of the conversation circle, but having to choose how to reply brings them back to the roleplay, as if they were between observing and controlling the PC.

Effect Sound Objects

Effect sound objects refer to the sounds made by objects and characters in the game world-space. They might be sound effects from actions like attacking, jumping, footsteps in different terrain, blacksmiths striking the anvil, a creature's growl and objects like

⁸ In DAI, the camera is the egocentric reference for the sound receiver, but in some other games this might not be the case. *Assassin's Creed Syndicate (2015)*, for example, utilizes the PC orientation as reference until the camera crosses a certain angle, then the camera serves as reference.

campfires, torches, forges – these are mostly positioned. In this game, during combat and roaming, you control a team of four members and even though there is a protagonist, the main PC's actions are heard from the camera's vantage point⁹. This reinforces the strategist playstyle, letting the players distance themselves and looking at each member from outside.

Object's sound sources are often used to give perspective while navigating. In most positioned sounds you can hear its approach and recession through volume. If the player only hears (non-positioned) sounds in a distance they never seem to reach, the world-space feels empty or inanimate quite fast. DAI uses mostly speech, sources of fire, non-hostile creatures and enemies to fill those gaps - the PC won't likely be walking for more than a few seconds without listening to positioned sounds.

Zone Sound Objects

Zone sound objects are often the ones that have more non-positioned sounds because they are also mostly ignored since the players have a lot of tasks and goals ahead of them. They are sounds from the environment, which is more ambiguous than it seems - a bird singing is part of the environment but if its audio object is attached to a source that is a NPC, it would make it an effect sound object. In this category we put the sounds of wind, rivers, rain and other ambient features. Some acousmatic sounds are never revealed and have ambiguous position, as I'll demonstrate during the analysis. The player hears some of them at roughly the same volume and fixed directionality, making it seem as if they are always at the same distance relative to the PC.

It happens with crowd chatter in a zone called Haven, for example, the sounds come from fixed directions in the sound image and do not adapt according to the PC's position (and this is why I haven't put this among the other speech sound objects). Wind is an exception too, because there is a non-positioned layer that keeps playing, but in open field and elevated areas you notice stronger winds shifting sides as you look around¹⁰. Of course, the player will hardly notice all of this unless they are focused on listening. Nonetheless, the zone sound objects are also the ones that evidenciate the different listening vantage points in a peculiar way since they are the only diegetic sounds that are not positioned (besides cutscene speech).

Score and Interface Sound Objects

Score and Interface sound objects are both directed to the player as they are non-positioned and not diegetic. The former sets the mood, creating narrative accompaniment and the latter consists of notifications, warnings, cues for action and indication of an action being executed. They reinforce the player's fixed vantage point but are not as interesting for the discussion since they do not create any tension over different types of listeners.

⁹ You can hear the PC's positioned bow attacks coming from the left side of the image in Loopy Longplays' video at 14:02–14:10, for example. < <https://youtu.be/hvH1Hy7HEwg?t=842> >

¹⁰ See the How Big is the Map? – Walk across DAI video at 1:38:38–1:38:45 < <https://youtu.be/3nqyCpkJV18?t=5918> >

Sound Objects Positioning

While looking at the categorization, it didn't seem fitting to group sound objects by their cause when zone and effect sound objects can easily become one another depending on their position. The player's navigation interacts with the sound image and that hints on the sound objects' positioning.

For example, critter sounds could either be environmental or tied to NPCs that are roaming around in the world-space. This has an implication on which portion of the ambient is interactive and which portion is not, rather than having sounds coming only from non-acousmatic sources. NPC sounds will change with listener's or their own positioning. Non-positioned sounds can have randomized spatialization but won't react directly to player's navigation, making the sound out of their reach and therefore indicating interactive boundaries of the world-space. We could incorporate the division between positioned and non-positioned sounds among Stockburger's categories of sound objects to differentiate the modes of spatial representation (Fig. 1).

	Positioned	Non-Positioned
Speech Objects	Short NPC lines	Cutscene/some dialogues
Effect Objects	Actions, attacks, sfx	Internal Sounds
Zone Objects	Birds, spatialized wind, water	Crowd chatter, non spatialized wind
Score Objects	Diegetic music, if any	Soundtrack
Interface Objects	Diegetic interface sounds	Regular interface sounds

Fig. 1. Stockburger's sound objects categorization divided by positioning with a few examples.

After going through this brief list, I realized Zone sound objects are the most variable since they're easily taken for Effect objects if they're positioned (or even speech in the case of a crowd chatter). Because of that, they're also more likely to convey different vantage points in ways that are not explicit – as it happens with NPC lines and cutscene speech, they can be positioned or not but their representation is clear for the player-listener. For the sake of the investigation, we will address these Zone sound objects more closely in our analysis to point out vantage points created by their positioning.

3 Game Sound Analysis

Elizabeth Hambleton's soundscape analysis [9] of the navigable narrative *Leaving Lyn-dow* (2017) does quite a good job in mapping out sound sources and it should add to what I'm proposing here, since this kind of game deals mostly with Zone sounds. It is not fit for discussing listening vantage points or orientation but some kind of mapping is important because we will always be looking out for any information on distance, depth and direction. She used schafferian soundscape analysis, which is concerned more with soundmarks (place signatures) and tonal profile. Along with Porteous and Mastin's style of mapping of SPL (sound pressure level) that is more accurate in determining distances.

What I did to investigate the positioning of DAI's sounds was a space-form analysis. By approaching the sound image, it is obvious that it only makes sense if we actively

use our other available senses - we can only perceive a positioned sound from the peripatetic listener, if we stand still in both game and life, all of the vantage points are fixed. I started the research watching videos of gameplay walking across maps and free roaming before I decided to choose this game. The videos were good for listening to sonic positioning from different zones and looking out for differences in ambient sound design, especially at the parts where they move the camera around the character to show the surroundings. For confirmation of ambiguous sounds, I had to play the game myself and try out soundwalking and static listening. I played until the PC gets access to a camp called Haven, exploring the camp and the space outside of it was the focus of this analysis.

Even though I mentioned space-form back in the introduction, I couldn't actually bring it up until now. Smalley describes it as "An approach to musical form, and its analysis, which privileges space as the primary articulator[...]" He makes a quite thorough conceptualization of space in the sound image, but it is all based on the fixed listener. Some of his categories will merge, transform and behave differently when we're talking about the game¹¹, but it is still useful since most of the times we can only access its sound image and not the project from the developers.

3.1 Perspectival Space

There's a glossary of terms at the end of Smalley's article and we won't use all of it. The terms best used were the types of space that deals with direction and distance. Regarding direction we have the ones that constitute the perspectival space: prospective, panoramic and lateral. Prospective space is formed in the frontal image, the central axis. Panoramic is the extension of the prospective space to the sides. Lateral is "the extension of panoramic space towards the rear of the listener."

Then we have the proximate and the distal space, which is a way of grouping sounds by distance from the vantage point. The distance is often understood by volume change but mainly the transformation that occur to sounds far away - difference in frequency, intelligibility of the sound, echo, response time, etc. The differences that constitute the distance in the perspective space creates a sense of scale in the sound image [15]:

With perspectival space there is a kind of contract between me, the viewer, and what is viewed. That my perceptions are rooted in me as a physical and spatial being means that all I perceive is 'sized up', located, and put into perspective in relation to the human scale - in relation to egocentric space. (p. 48)

Perspectival space concerns an egocentric reference position, so there will be the PC-camera and the player perspectives. As I mentioned before, an object's position does

¹¹ To mention the ones left out: There is no use for performed space and its components, nor the spaces that are related to cause because all of the ambience's sounds have implied causes. The movements of approach and recession is given by the moving of the PC in relation to source or vice versa and were used to understand volume and positioning but they are obvious. Overture and enclosure can be noticed by entering and leaving buildings or when hearing loud thunder noises but besides that they were not very important for the analysis either, the PC's navigation will naturally create approach, recession, overture and enclosure in the sound image.

not necessarily imply on its spatial qualities. The spatial forms and position are separate things, but they can be used together to compose the sound image in a creative way. There is no actual distance besides the one between audio output and the player's ears, every distance is simulated in the sound image, so there can be sound objects near the camera-listener that sounds like events occurring in the distal space or the other way round.

The scale in the perspectival space is managed through those simulations and differences between positioned and spatial form's distance. It will combine the positioning and the perspectival space in the sound image to achieve the scale intended for the ambient. Thus, making the positioned distance an egocentric reference to the camera-PC and the perspectival space egocentric to the player.

Vectorial Space

There is also the vectorial space, that is "the space traversed by the trajectory of a sound, whether beyond or around the listener, or crossing through egocentric space" [15] p. 56. The main tool to differentiate positioned from non-positioned sounds was moving the camera around paying attention to the vectorial space of each sound. Trying to shift a positioned sound from panoramic to prospective space and then trying to hear them crossing over to the other side of the headphones was the way to confirm that. Non-positioned sounds do not form a vectorial space that follows the camera's panoramic space, so one can hear which sounds are vectorized and which aren't. Sometimes non-positioned sounds can feel as if they are positioned when another sound with shared frequencies has a vectorial space that crosses its fixed space.

3.2 Game Context

Dragon Age: Inquisition is a CRPG (combat role playing game), which is a genre of RPG that favors strategic real time combat. One of the reasons I chose the game is that because of this strategy interest, the player controls a party of four characters, navigating and issuing commands one at a time. The genre usually features a pause function and strategic view, so the player can step back and decide how to proceed whatever time needed. For DAI, this means that the eyes and ears of the player become the camera and not the PC, even though it can't stray away too far from the party members. The player can choose, however, to play as one of the characters throughout the combat (as the difficulty scales higher, players are encouraged to micro manage each character simultaneously with the strategic view). This is in itself a clear indication of the different vantage points, the camera attached to each of the four members and the one attached to the strategic view - all of them peripatetic.

Dragon Age franchise has a long history through video games, books, tabletop rpgs, comics and a movie. It depicts a high fantasy world setting called Thedas that portrays adventures, political conflicts, romance and religious themes. Not only humans live in Thedas, there are dwarves, elves, qunari and many fictional creatures. *Dragon Age: Inquisition* starts with a nightmarish cutscene showing the character you create swarmed by monsters while a luminous entity calls for you. The PC wakes up imprisoned with a mark on his/her left hand being interrogated by the authorities as the prime suspect to a catastrophic event, the opening of a portal called The Breach. It spawns demons

and hellish hostile creatures that threaten life in Thedas. PC is then led by Cassandra Pentaghost, with the intent to use the mark to close the portal. The beginning of the gameplay consists of a tutorial area, where the player learns how to control the character, attack, loot items, and interact with objects. It ends when the PC manages to stabilize the Breach after slaying a huge monster spawned by it. The Chantry - a powerful religious organization - seeks to execute the PC for the suspicion of being responsible for these portals. Cassandra and other military characters break allegiance to the Chantry in order to promote an Inquisition against these demons and portals, using the PC's mark to restore order to Thedas.

After these events, the game's logo appears and the player is free to roam in Haven, the camp where the Inquisition was set. It is a snowy region, surrounded by mountains in the distance, it has roads leading to both east and west. There's a river that comes from the east, under a bridge and ends in a lake in the middle of the map, both river and lake are frozen. There is a big hall in a stone construction, there are wooden cottages that serve as lodging, infirmary, there's the apothecary, the forge with the blacksmith anvils, the soldier's tents where they also spar. There are some trees in the northeastern part of the map.

3.3 Field Notes

My analysis began soundwalking through and around the camp to listen to whatever sounds I could find. After having heard the sounds that are part of the zone, I had a good holistic view of the soundscape and started taking notes to organize them between positioned and non-positioned sounds. I thought there were four non-positioned sounds: birds, wind, dogs and crowd chatter. Then I chose a spot to try to pick them apart, away from positioned speech since they capture our attention and space in the sound image.



Fig. 2. In front of the apothecary's lodge, facing north.

The first spot was in front of the apothecary's lodge, right between two other wooden houses and facing the wooden palisade (Fig. 2). In the proximate space I could hear crackling fire of the torches in front of the houses, birds (lateral space while facing north) and wind (as a permanent texture, centralized in prospective space). In the distal panoramic space, there was crowd chatter¹² and dogs barking. When turning the camera around to force vectorial spaces, I noticed the birds were actually positioned sounds, probably attached to the middle or top part of the trees. They are misleading because they can be heard often at the same volume, making it seem as if they were really close in the proximate space. Here the birds and the dogs are acousmatic (at least I couldn't find any). The chatter is indeed non-positioned, during the search for vectorial traces of single vocal sounds I noticed they were fixed in the sound image, in the prospective but mostly on the edges of panoramic space. Wind is always centralized in the image, but depending on where the PC is, you can hear another layer of positioned gales. This happens usually out in the open and at elevated places like top of hills and mountains, as I found out walking outside of the camp site.

The dogs were left for last because they were the most elusive. I went on another walk to try to pinpoint anything I could find regarding their direction and got to a small pier north of the frozen lake (Fig. 3). To the south you could see the camp gates, where the dogs should be coming from. And to the north, behind my PC was a few trees that had birds chirping. From time to time there was a sound of something big made of wood creaking. Not many sounds compared to the first spot, so I could test out the position of the dogs in a more precise way. I stood by the end of the pier facing a direction until I heard the dogs bark, after I exhausted the four directions, I noticed their direction was fixed. I could hear dogs barking from somewhere to the south and somewhere from the east. This sound resists positioning to some degree as it is always in the distal space, never in the proximity. If you follow the trail leading to the direction from where you heard the bark, you end up hearing the same dogs barking from other directions.

I deduced the dogs' barking sounds are part of the zone's texture, always playing within the map but with fixed directions when played. It is understood through abstraction that there are dogs somewhere, but they are never found. They are always in the distal space, following the PC around the zone at roughly the same perceived volume level. The opposite happens with the birds, we often hear them as if more proximate than they really are.

If the dogs are part of the zone the PC is walking through, positioned but always at an unreachable relative distance, it addresses a vantage point that is not the PC's - maybe a point fixed in space or at a different scale. If the chatter is non-positioned, the addressed vantage point is not clear either. It is fixed as from a listener that doesn't move along the navigation of the PC - could it be an imagination of what would be heard in the area? Even though those sounds are clearly diegetic, non-positioning could refer to the presence of a listener that is neither the PC-camera nor the player, that has a fixed listening vantage point to be able to hear them.

¹² Like the wind, can also be understood as a permanent texture, but the chatter is confined to the camp's limits while the wind's texture can be heard throughout the whole map zone.



Fig. 3. Map distance between the apothecary's lodge and the spot in the frozen lake. (Straight line between the two circles).

We can also see the distance of these specific vantage points as a distal interpolation that doesn't rely on closure of the proximate space¹³. Some vantage points could be representing listeners that are listening from elsewhere, like a stethoscope [17], while others are positioned in a clearly represented space. This external listener adds up to the other vantage points, creating a virtual space that convey different distances being heard.

On a different key, there is a mood set by the crowd chatter in Haven, indicating that even though in a cold place with people recovering from wounds, they are engaging in social activity, the place is filled with movement. For the context of the start of the game, Haven is a boiling pot for social and political activities. The sonic atmosphere sets the feeling of having a lot of work ahead, while the PC travels around the continents in political missions. Its purpose is figurative and does not need to abide to causal relations the same way as other positioned sounds, making use of an external listener to deliver the mood to the player efficiently.

4 Conclusion

The tension between space in world-space positioning and space in the sound image create multiple vantage points and types of listeners. These vantage points are not always coherent in terms of where the sounds are being heard from, but they have an effect on how the players understand the zones. Even though most players won't hear in a detailed manner, the details are meaningful and can be used to help compose ambient sounds that describes the zone creatively.

¹³ Smalley defines it as “a temporary rupture in ongoing proximate space thereby permitting access to a distal view”.

The combination of different listeners with their own vantage points creates a virtual result of the listening experience, assembling an entity with many ears, listening to sounds from different locations at the same time. There is the peripatetic camera, the fixed player, the crowd listener, the non-oriented wind listener, the dog's listener – even though they all express different vantage points, they are seen as one since the sound image mediates them all.¹⁴

Stockburger's categorization could be expanded upon by including positioned and non-positioned sounds as a way to separate sounds that are represented in the world-space and those that aren't. That could help sound designers think about sound and spatial representation within their projects, contributing to the structure design and helping shape the sound identity of the game.

Also, Smalley's space-form analysis has proved to be a useful method for investigating ambient sound design processes in video games and could be further developed by condensing the core concepts and techniques in a way that's easier to apply without in-depth understanding of terms that might not be relevant when talking about a video game's sound image. Being able to draw information from sound image might assist case studies since most game projects are not shared publicly (what makes audio implementation studies somehow obscure and reliant on hearing and personal knowledge on the audio engines used in the process).

Besides that, sound takes an important role in the way we understand our surroundings, both real and virtual space [16], so investigating spatial representation and cognition is important in order to create environments that can express feelings and narrative through either matching or expressively rejecting expectations of causality. The crowd sound loop and the barking dogs are nowhere to be seen and still they are largely responsible for how the mentioned zone feels active and lively for the player. If these sounds were positioned, the player would certainly lose the effect of the lingering texture throughout the zone – making it more realistic, but also imparting the absence of activity at more spots around the map. On one hand, the world-space feels empty if the player only hears non-positioned sounds since there isn't perspective being formed in the sound image. On the other hand, it feels equally empty if the player only hears positioned sounds and there is not enough activity to fill the perceived space.

This discussion is relevant to any kind of open world game and it would be interesting to elaborate and apply it on navigable narratives (also called 'walking simulators'¹⁵) as one of the facets of environmental storytelling [4]. Ambience can create or reinforce meanings and gameplay features through causal and figurative uses of sound. The mood set in Haven is one of the many possible examples and it's only because we have different vantage points at the same time. There is a lot of room for combining different vantage

¹⁴ It's possible to add and remove listening vantage points while dealing with the non-positioned and acousmatic ones, but the overlapping of virtual non-positioned listeners makes them sound as if they are one since they belong to the same sound image and share the same reference. There is also the possibility of adding vantage points through using extra audio output devices, offering distinguishable fixed listeners (as it occurs with the inbuilt speaker of PS5's controller).

¹⁵ There is a debate on the pejorative use of the term to refer to this genre. Elizabeth Hambleton [6] proposes 'navigable narratives' as an alternative and after reading other sources on the subject, this is the one I preferred.

points and listeners in order to create meaning in the holistic view of the sounds. The virtual multi-eared player gets to manage all of the vantage points, making the auditory experience of the game unique among other media.

References

1. Bioware.: *Dragon Age: Inquisition*. EA, [Windows/PlayStation 3/ PlayStation 4/Xbox 360/Xbox One/Xbox Cloud Gaming] (2014)
2. Caesar, R.: The bandwidth of soundness: between sound and image. XXIII Congresso da Associação Nacional de Pesquisa e Pós-Graduação em Música. Natal (2013)
3. Caesar, R.: Sound isn't a thing in itself, but a carrier for leaking things. *Dossiê Sonologia* 297–308 (2020)
4. Chang, A.: Games as environmental texts. *Qui Parle* **19**(2), 56–84 (2011)
5. Chion, M.: *Audio-Vision: Sound on Screen*. Columbia University Press, New York Chichester, West Sussex (1994)
6. Chion, M.: *Sound: an Acoulogical Treatise*. Duke University Press, Durham (2016)
7. Elleström, L.: *Media Borders*. Palgrave Macmillan, Multimodality and Intermediality (2010)
8. Galloway, K., edit. Gibbons, W., Reale, S.: *Music in the Role-Playing Game : Heroes & Harmonies* (chapter 10) *Soundwalking and the Aurality of Stardew Valley: An Ethnography of Listening to and Interacting with Environmental Game Audio*. Routledge, New York (2020)
9. Hambleton, E.: Gray areas analyzing navigable narratives in the not-so-uncanny valley between soundwalks, video games, and literary computer games. *J. Sound Music Games* **1**(1), 20–43 (2020)
10. Jørgensen, F., edit. Svensson, D., Saltzman, K., Sörlin, S.: *Pathways: Exploring the Routes of a Movement Heritage* (chapter 9) *Walking and Worlding: Trails as Storylines in Video Games*. White Horse Press (2022)
11. Klatzky, R.: *Allocentric and Egocentric Spatial Representations: Definitions, Distinctions, and Interconnections*. Springer, *Spatial Cognition An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* (1998)
12. Modena, E., Parisi, F.: Exploring stories, reading environments: flow, immersion, and presence as processes of becoming *Elisabetta modena*. *Cinergie – Il cinema e le altre arti*. **19** (2021)
13. O'Hara, W.: Mapping sound play, performance, and analysis in proteus. *J. Sound Music Games* **1**(3), 35–66 (2020)
14. Roginska, A., Geluso, P.: *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*. Routledge, New York (2018)
15. Smalley, D.: Space-form and the acousmatic image. *Organised Sound* **12**(1), 35–58 (2007)
16. Smalley, D.: Spectromorphology: explaining sound-shapes. *Organised Sound* **2**(2), 107–126 (1997)
17. Stankievecch, C.: From stethoscopes to headphones: an acoustic spatialization of subjectivity. *Leonardo Music J.* **17**, 55–59 (2007)
18. Stockburger, A.: The game environment from an auditive perspective. In: *Proceedings of the Level Up, Digital Games Research Conference*. NL (2003). http://www.stockburger.at/files/2010/04/gameenvironment_stockburger1.pdf