Check for updates

Article

animation: an interdisciplinary journal

Animation:
An Interdisciplinary Journal
8(3) 283–299
© The Author(s) 2013
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1746847713507164
anm.sagepub.com
SAGE

Sonic Subjectivity and Auditory Perspective in *Ratatouille*

Karen Collins
University of Waterloo, Canada

Abstract
Using examples from *Ratatouille*, this article illustrates how sound can be used to create identification with characters through an auditory perspective. This auditory perspective is created through reinforcing or contradicting the on-screen image with microphone placement to create distance, loudspeaker positioning to create location, digital signal processing effects to create environments, and subjective perspectives that position us as an insider or outsider, and which illustrate the internal subjectivity of characters.

Keywords
animation, auditory perspective, identification, point of view, *Ratatouille*, sound, spectatorship, subjectivity

Sound design in animation differs from most live action film in significant ways. Summarized by *Wall-E* director Andrew Stanton:

> The big thing that's unique about sound design and animation is [that] you get nothing for free. You don't get on a set and hear the way the environment sounds naturally or the way someone walks across a room or just a voice. (*Animation Sound Design*, 2008)

Indeed, while most film today relies significantly on post-production sound, the sound designers are usually given a base from which to start: real-world sounds created by the actions of actors, and the environments recorded on set. Animation sound designers, like video game sound designers, must construct an entire sonic world from their imagination. In fact, famed sound designer Walter

Corresponding author:
Karen Collins, Associate Professor, Department of Drama and Speech Communication, University of Waterloo, 200 University Ave West, Waterloo, ON N2L 3G1, Canada.
Email: collinsk@uwaterloo.ca

Murch suggests that it was animation that historically paved the way for more creative use of sound in live action films:

> In the beginning of the sound era, it was so astonishing to hear people speak and move and sing and shoot one another in sync that almost any sound was more than acceptable. But with animated characters this did not work: they are two dimensional creatures who make no sound at all unless the illusion is created through sound from one reality transposed onto another. (cited in Oldenbourg, 2005)

It is this construction of a believable fictional world that is a significant role of sound in animation, where real-world sound adds 'credibility', 'to sell to the audience the reality of what's really just a very fantastic world' (Ben Burtt in *Animation Sound Design* 2008).

The differences between some types of live action film (particularly science fiction) and animation sound may be disintegrating in the face of increasing use of computer generated graphics, as described by sound designer Randy Thom:

> There isn't as much of a difference as you would think (between live action and animated movies). A lot of the big live action movies these days have so much computer graphics in them that they might as well be animated, in the sense that you have to invent a whole world, a whole sonic environment or a whole sonic world, most of whose imagery is coming out of a computer. (in 'The Sound of *Ratatouille*', Sound Works Collection, nd)

Nevertheless, animation sound still contains some important differences, most notably in aiding the construction of a credible three-dimensional space in which the action takes place, and in helping the audience to anthropomorphize the 'two-dimensional creatures', thus aiding identification and empathy with those characters.

Identification is a complex concept with a long and loaded history that cannot be reasonably explored here (for a summary of this history as it relates to media theory, see for instance McQuail, 1997, or Cohen, 2001). I use the term to refer to feelings of affinity, empathy, similarity and liking of a character by the audience (Cohen, 2001: 249). Jonathan Cohen argues that we should also add to this definition a sense of identification as an experience; that is to say, that we may mentally adopt the goals and identity of the character. In other words, we assume the perspective of the character with whom we identify, and experience the narrative as if through that character – as if events that are happening to the character are happening to us. Through this psychological roleplay, we therefore feel *with* the character, rather than *about* the character. Moreover, this adoption of the character's persona means that we as an audience lose our own self-awareness as we temporarily become the other character. Cohen suggests that most texts are intentionally set up to provoke identification, through techniques such as camera angle and perspective, and through creating characters with whom the audience can sympathize; thus, 'Unlike the more distanced mode of reception – that of spectatorship – identification is a mechanism through which audience members experience reception and interpretation of the text from the inside, as if the events were happening to them' (Cohen, 2001: 249).

A variety of theories have arisen about our identification with on-screen characters. Focusing on video games, James Paul Gee (2004: 55–56) employs the concept of the 'projective identity', which is one of three simultaneous identities that occur as one plays a game: the player (the real world), the character (the virtual world), and the projective identity, which is 'the interface between – the interactions between – the real-world person and the virtual character'. This projective identity, argues Gee, is a combination of the character and the player's belief (projection) about the character's personality. Such an approach can easily be extended to animated film: although we do

not physically control the character in the ways that we can in a video game, there are various means by which we are led to identify with the character, and project our own thoughts, feelings and emotions onto them. Discussing television, Cassandra Amesley (1989) suggests that there is a 'double viewing', in which the characters are both real and constructed – in a sense, fitting into this space of the projective identity, since to recognize the characters as constructed is to acknowledge the artifice of the narrative, and to develop the sense that characters are real is to project one's own beliefs and personality onto them. Thus, the character becomes us, as much as we become them, in a blurring of real and imagined personality traits. Robynn Stilwell (2005: 51) suggests that:

> Experiencing a strong identification with a character in the film places us in another's subject position, creating an emotionally empathetic response. Film has many ways of coaxing the audience into that position, from character development, narrative discourse and events, to the more 'visceral' point-of-view shot compositions and sound design.

Most scholarly writing on identification with screen characters deals with the visual, more specifically the camera angles, where camera angle equates roughly to the point of view of the character, 'where telling is attributed to a character in the narrative and received by us as if we were in the situation of a character' (Branigan, 1984: 73; see also Johnson, 1993; Flitterman-Lewis, 1987). The audience's viewpoint may be switched periodically from a third-person observer to a first-person angle from which they view through the character's eyes; thus, we are physically put in the position of the character temporarily to facilitate identification. Lev Manovich (2001: 108) explains that the

> classical cinema positions the spectator in terms of the best viewpoint of each shot, inside the virtual space. This situation is usually conceptualized in terms of the spectator's identification with the camera eye. The body of the spectator remains in her seat while her eye is coupled with a mobile camera.

Both Manovich and Anne Friedberg (1993) conceive of the frame of the cinema screen as a window to 'the existence of another virtual space', a 'space of representation', in which 'the viewer simultaneously experiences two absolutely different spaces that somehow coexist' (Manovich, 2001: 95). The visual on-screen objects cannot be directly, meaningfully interacted with, and thus 'The metaphors of the frame and the window both suggest a fundamental barrier between the viewer and the representational objects seen in the image-screen' (Cleland, 2008: 171). This dual space maps onto the 'double viewing' or dual identity experienced by the audience. But suggestions of a barrier formed by the screen fail to recognize the role that sound can play in our phenomenological experience of the construction of space in cinema: film does not only take place on the screen, but also in the auditory space around us.

Using examples from *Ratatouille*, this article illustrates how, in addition to vision, sound can be used to create identification through an auditory perspective. This auditory perspective is created through reinforcing or contradicting the on-screen image with microphone placement, loudspeaker positioning, digital signal processing effects, subjective perspectives and through illustrating the internal subjectivity of characters. While such techniques are often useful in most film, including live action, they are highlighted by animation film, which uses these techniques to construct a believable world.

Ratatouille is a computer-animated feature film produced by Pixar in 2007, directed by Brad Bird. The film has two protagonists: Remy, a rat with a remarkable sense of smell who fancies himself as a chef, and his human chef partner, a young man named Linguini. Remy wants to become a chef, but because he is a rat he must work through Linguini by hiding in the man's hat

and helping him to select ingredients for the meals Linguini is cooking. The film was selected for analysis for its particularly notable use of sound: indeed, it was nominated for Academy Awards in the categories of Best Original Score (by Michael Giacchino), Best Sound Editing and Best Sound Mixing (by Randy Thom).

Auditory perspective and the point-of-audition

As with camera angle, sound can have a significant impact on our feelings of immersion in a narrative, and in our ability to identify with characters. The auditory perspective in cinema has been referred to as the 'point of audition', defined as a reference point from which the auditory perspective is created, roughly analogous to the point of view. Rick Altman (1992: 60) suggests that

> point-of-audition sound always has the effect of luring the listener into the diegesis not at the point of enunciation of the sound, but at the point of its audition. Point-of-audition sound thus relates us to the narrative not as external auditors identified with the camera and its position … nor as participant in the dialogue (the standard situation of the 'intelligible' approach), but as an internal auditor.

In this way, argues Altman,

> We are asked not to hear, but to identify with someone who will hear for us. Instead of giving us the freedom to move about the film's space at will, this technique locates us in a very specific place – the body of the character who hears for us. Point-of-audition sound thus constitutes the perfect interpellation, for it inserts us into the narrative at the very intersection of two spaces which the image alone is incapable of linking. (pp. 60–61)

As with camera point of view, then, the point-of-audition allows us to hear the fictional space as if we were one of the characters.

Film theorist Michel Chion (1994) argues that the image always creates the point of audition, in that the image dominates our subjective experience, and indeed there is much research to support the idea of a visual dominance over the auditory sense. The ventriloquism effect (e.g. Chloe et al., 1975), for instance, has shown that we hear a sound as coming from an image if we associate the sound with that image, even if the sound is arriving from another direction. If we see a car moving on the left-hand side of the screen, for example, then even if a car sound emanates from a right-hand side loudspeaker we will often hear it as coming from the left, from the car. In such cases, even when the auditory mix and visual perspective are not aligned, the visual image may dominate. However, recent experiments have shown that rather than visuals always dominating sound, the relationship is much more complex, and there are now examples of sound dominating what we see (see Collins, 2013: 26–32 for an overview).

Rather than always strictly following image and existing on a single point, then, sound more accurately exists in a *zone* of audition, since many types of sound are difficult to localize (Chion, 1994: 89–92). This zone of audition, according to Chion, provides both a *spatial sense* (the point from which the space is represented auditorially) and a subjective sense (the character from whose perspective we are hearing). The spatial sense asks, 'from where do I hear, from what point in the space represented on the screen or on the soundtrack?', whereas the subjective sense asks, 'which character, at a given moment of the story, is (apparently) hearing what I hear?' (p. 90). Both the spatial sense and the subjective sense of sound presentation in cinema serve to create the auditory perspective and to facilitate identification with a character. Below, I separate the techniques of sound production that create the spatial sense and the subjective sense, illustrating how each

technique contributes to identification in animated films such as Ratatouille. However, as will be shown, the spatial and subjective sense cannot easily be separated. Indeed, the spatial sense serves to reinforce and create the subjective sense in some instances.

I The spatial sense

Mike Jones (2005) describes that:

> There is an unvoiced acceptance on the part of viewers that all that is important in a scene will take place within the screen's frame. But in the 21st century, many of the key aesthetics of audience acceptance and visual understanding of a broader cinematic space derive not from cinema but from computer gaming. A larger, more complex imaginary world composed by an auteur in-space rather than in-frame.

The visual experience of cinema can only exist on-screen, and while camera positions can give the illusion of depth and of continued space beyond the screen, it is sound that extends the fictional world into our own. The spatial sense of auditory perspective is constructed by a variety of techniques that create or reinforce the physical sense of space for the listener through the use of spatialized sound. These techniques combine physical acoustics with psychoacoustics, the perceptual aspects of our response to sound. For example, the perceived location of a sound can appear to emanate from between two loudspeakers, in what is referred to as a 'phantom image'. The techniques commonly used to create and reinforce a sense of acoustic space for the listener, including microphone placement, loudspeaker placement, and digital signal processing effects.

A number of acoustic properties and effects determine the ways in which sound propagates in space. These are too complex to cover in any detail here, but a basic understanding will benefit the discussion that follows. Many sound sources have a particular direction in which they propagate (such as loudspeakers), which can be referred to as the facing angle. When a sound wave meets a new surface boundary of a material, some energy is absorbed and some is reflected. The amount absorbed depends on the frequency of the sound, the angle with which the sound meets the material, and the material itself. The absorption properties of the material impact the amplitude and the phase of the reflection, and the material's surface structure impacts the direction or angle of the reflection (as related to the angle of incidence). The more reflective the surface walls of that room, the more reverberant the space will be, as the reflections bounce off the wall until their energy is all absorbed. Sound will reflect off the surfaces, and with each reflection change timbre and amplitude slightly (as some frequencies are absorbed and others are reflected). Evidence suggests that we only need a few seconds to 'calibrate' a room to develop a sense of its space, which is determined primarily by the character of the reflections of sound in that room (Plenge, 1974: 44). Indeed, it is possible for our brains to determine the approximate size and shape of a space by listening to the reverberation patterns (see, e.g., Schenkman, 2010). In other words, sound can play a significant role in the construction of the illusion of a material, tactile, three-dimensional space.

The ways in which we perceive directionality and distance of sounds depends on a number of characteristics of both the listener and the environmental space. The perceived location of the sound can depend on 'the experience and expectations of the listener and on the type of attention he is paying to the sound' (Gerzon, 1974). Our perception of source direction in most circumstances depends on the differences in time and direction between the signal reaching our left and right ears. Our ability to perceive distance is less accurate than our ability to perceive direction, and often depends on our prior experience and knowledge of the sound (for instance, we know a car to be an approximate size and can judge distance based on our experience of cars at a distance from us). While the use of loudness is often pointed to as the primary means to distinguish distance,

there are many other factors that come into play. Timbre significantly influences localization, since timbral definition diminishes over distance.

Rumsey (2006: 652) argues that in cases where spaces are entirely constructed – as in animated film – without a

> 'natural' reference or perceptual anchor', the space is 'a form of 'acoustic fiction' or 'acoustic art'… It also brings with it the challenge to evaluate highly complex and changing spatial audio scenes, containing elements that may not have a direct parallel in natural listening or which may be mutually contradictory (dry and reverberant sources could be combined within a single mix, for example).

The creation or re-creation of auditory spaces is, in other words, part art and part science. Although a realistic auditory space is developed with a 'naturalness … mediated by different understandings of perception' (Lastra, 2000: 191), in many cases the auditory space (in terms of real-world acoustics) is fabricated and inaccurate. Spatial sound techniques, in other words, are critical to aiding the construction of a believable space, and of giving that space materiality. This is particularly the case in animation, where we are sometimes as an audience to take solid shapes and lines for realistic walls and barriers. By using spatial sound techniques, those shapes, lines and objects are given tactile, haptic credibility.

Microphone placement

The sense of distance and space can be created in part through the use of microphone placement. Von Békésy (1960) has shown that a radio broadcast listener can judge with some accuracy the proximity between the microphone and the speaker, due primarily to the amount of reverberation on the signal, which is to say, the amount of direct versus reflected sound. Direct sounds are sounds that arrive at a listener's ears directly, without any reflections off surfaces, whereas reflected sounds are the reverberations of that sound off objects in the physical space, which creates a short delay and colours the sound through attenuating some of the frequencies. In cinema, the auditory perspective is often created in part by a careful positioning of the microphone to blend direct and reflected sound, to duplicate the real space in which the acting takes place, in such a way that the sound perspective matches the visual perspective. The degree of loudness, or sound attenuation (loss of intensity), gives the illusion of proximity from the sound emitter to the listener (perceptually located at the microphone). Simply put, microphone distance affects not just the loudness, but also the tone colour of the sound source – the reverberation patterns.

When the microphone is at a distance of up to about one foot from the emitter, it is considered 'close miking', which emphasizes the direct signal over the reflected signals. 'Distant miking', on the other hand, places microphones at several feet from the sound emitter, usually capturing as much of the sound of the reflections as of the direct sound. Even farther from the source, 'ambient miking' allows the room signal to dominate over the direct signal, for instance capturing the sounds of a crowd at a concert. In this way, microphone placement plays a role in our perception of the location of sounds in relation to ourselves, and of other objects that may obscure that sound in the environment.

For example, imagine a listener in a cathedral, with a speaker at the front, at point *x* (see Figure 1). The listener at point *y* would be able to hear the speaker quite clearly, primarily receiving the direct signal, but would still capture some of the reverberation. The listener at point *z*, however, would hear much more of the reverberation over the direct signal. Anyone who has stood at the rear of an old cathedral would have experienced the difficulty with the clarity of the speech being obscured by the reverberations.

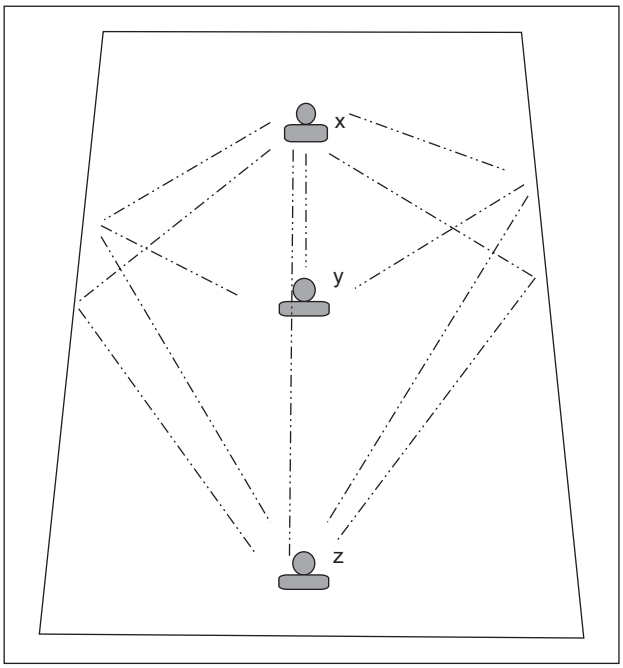


Figure 1. Microphone distance creating a sense of space.

In a live action film, the microphone mix (overall loudness, and the mix of direct versus reflected sound through distance) typically mimics the angle and distance of the camera to the sound source (Belton, 1985: 68). For the most part, the camera angle and sonic spatial point of audition reinforce each other and create a sense of distance between the audience and the objects or characters on-screen. The approximate visual distance of the audience from the screen objects usually matches the approximate sonic distance. However, there are also times when the auditory and visual perspective do not align, such as with long shots paired with close-up sound.

In animation, of course, the distance is entirely constructed through the use of visual perspective illusions, which is typically reinforced through the use of sound, although as with live action, the use of microphone distancing does not always reinforce the visual distance. For example, in an early scene in *Ratatouille*, the rat Remy is walking with his brother through a field. Most shots show the rats up close, and microphone distance reinforces this visual. However, the camera pulls back at about 00:05:41 and we now see the rats from a distance (Figure 2), without a change in the microphone distance. Here, then, the microphone distance contradicts the visual distance that we experience. This creates a subjective distance of us as being close to or walking with the rats, while simultaneously providing us with a sense of the entire environment in which this part of the story takes place. Distance miking in this shot would not only have placed us physically distant from the rats, but also at a psychological or subjective distance.

Particularly important to the sense of identification with characters is the use of auditory *proxemics*. Proxemics is the study of the distance between people as they interact, including the culturally-specific personal space that we all maintain (Hall, 1963). According to Edward T Hall, we maintain social distances that can be divided into roughly four zones. The closest space to our bodies, at a range of less than 18 inches, is our intimate zone, where we touch, embrace, or whisper. Beyond this is our personal space, up to a distance of about 4 feet, generally reserved for family



Figure 2. *Ratatouille* (2007, dir. Brad Bird, Pixar/Disney, screen capture). The camera angle changes from close-up to distance shots without change in mic.

and close friends. The social distance for acquaintances extends as far as 12 feet, and finally we have a public distance beyond the social zone. These distances are culturally specific, but for the most part are accurate for Western film audiences.

Auditory proxemics thus relates to the distance between a recorded presence and the listener (see Moore et al., 2011). Hall outlines seven sub-categories of vocal effect: silent, very soft, soft, normal, normal +, loud, and very loud. The distance of the microphone can be analogous to our proxemic zones. In other words, the loudness of a voice can convey a degree of intimacy to us (as the character being spoken to in a film). Close-miking, for instance, can give the illusion that we are more intimate and closer to the speaker than if the microphone is at a distance from the speaker's mouth when recorded. This close-miking technique was used in the era of the 'crooners' in popular music, enabled by the invention of the electric microphone, allowing for a genre of intimate and personal music to emerge (see, e.g., Lockheart, 2003). Feelings of intimacy can thus be created by whispering and close-miking, whereas a social or psychological distancing effect can be created by distance miking. Thus, the hearing of sighs or soft breathing is often used to facilitate a poignant emotional connection.

For example, in *Ratatouille*, when Remy discovers that his hero Chef Gusteau has died, he is shortly thereafter chased with the rest of his clan, and escapes downstream on Gusteau's cookbook. Remy gets separated from his family, and, lost and alone in the sewers, he flips through a cookbook, and with a wave of a hand, sighs deeply and turns away, unable to look at his mentor's image. The closeness of the microphone distance, allowing us to hear Remy as if we are in his intimate zone, leads to a strong sense of empathy and identification, but also, since the character is a fictional animated character, it lends life to the rat, by allowing us to hear his breath.

Microphone placement, in other words, creates a sense of physical space and distance from objects to the protagonist, but also creates a sense of the emotional distance and relationships between the protagonist and the other characters. The illusion of proximity and distance between the characters and the audience is thus reflected in the ways in which we are to identify with the characters, creating feelings of distance or intimacy.

Positioning in the loudspeaker

The spatial positioning of sound effects around us as the audience using loudspeaker (or stereo/headphone) positioning (mixing), also helps to represent the sonic environment of the visualized

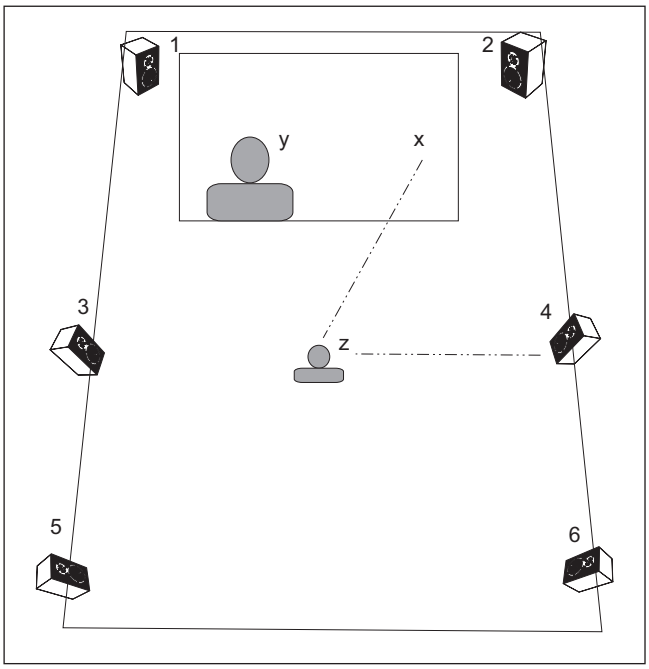


Figure 3. Speaker set-up and sound location.

space, and can extend beyond the screen into the off-screen (acousmatic) space. Sounds can appear to emanate from a physical place around us using the physical positioning of loudspeakers or panning and phantom imaging techniques. The effect can be so significant as to have us physically turn our gaze towards a sound, in what is commonly referred to as the 'exit sign effect', where a discrete sound located in a speaker position to the side or rear will have us turn our eyes towards the exit signs in the theatre (see, e.g., Holman, 2008: the term is often ascribed to *Star Wars* sound designer Ben Burtt).

As with microphone techniques, the use of loudspeaker mixing techniques can create both a subjective as well as a physical spatial position. For example, given a fairly standard 7.1 surround sound theatre set-up (see Figure 3: the centre channel and subwoofer are not pictured), imagine that a character on-screen is located at point *y*, with a sound emitter at point *x* on the screen, and a listener at the 'sweet spot' of point *z* in the audience. If I am intended to be an audience–observer – that is, in third-person perspective and external to the scene – the sound should appear to emit from somewhere around or point x (primarily panned to speaker #2). If, on the other hand, I am hearing from character *y*'s ears, in first-person perspective, the sound should appear to come from my right, closer to speaker #4. The selection of how to mix sounds in the speakers, in other words, can help the audience to identify with the character.

In a dialogue between Remy and Linguini, where Linguini is situated on the left-hand side of the screen, and Remy is situated to the right (Figure 4, at about 00:32:22), depending on which character whose ears we are to hear with, the mixing would be different. If we 'are' Remy, in the sense that we are to hear through Remy's ears, then Linguini's voice should appear in front of us, rather than to our left. However, as with microphone positioning, the speaker positioning will typically reinforce the visual image. Here we are clearly visually in third-person perspective, external to the characters. As such, we hear Linguini mid-left of centre in this particular shot, placing us in



Figure 4. Third-person loudspeaker positioning. *Ratatouille* (2007, dir. Brad Bird, Pixar/Disney).

a third-person, external auditory position. Indeed, it would be strange in either third-person or first-person point of view, to have a contradictory point of audition when it comes to the sound's speaker positioning. Nearly always, when it comes to loudspeaker positioning, the sound reinforces the point of view (there may be exceptions in more abstract, artistic productions).

Digital signal processing effects and the mix

Another means to create the sense of environmental space – and thus our location within that space – is through digital signal processing effects, including reverberation (as discussed above), and various filters. In this way, the techniques spatialize the sound to 'endow it with 'presence' [that] guarantees the singularity and stability of a point of audition, thus holding at bay the potential trauma of dispersal, dismemberment, difference' (Doane, 1980: 45). For instance, when a sound emitter is completely behind an obstruction such as a wall that separates the emitter (at point *x*) from listener (at point *y*), the sound is *occluded* (see Figure 5 top). A simple analogy is to imagine closing a window to neighbour noise – the sound can still be heard, although the overall volume is attenuated and certain frequencies are significantly reduced. To create this effect sonically, then, typically, the overall volume will be reduced and a low-pass filter will be applied. A low-pass filter sets a cut-off frequency above which the sound signal is attenuated or removed, allowing the lower frequencies to 'pass through', while cutting off the higher frequencies. The degree of low-pass filter applied (the settings controlling the intensity of volume attenuation and the amount of frequencies cut off) depends on the type and thickness of the material or object between the emitter and receiver. Both the direct path of the sound and any reverberations in the space are therefore 'muffled' by attenuation and the filter.

If the direct path between the emitter at *x* and the listener at *y* is blocked, but the reverberation angles are *not* blocked, the sound is said to be obstructed (Figure 5 middle). Here, the direct signal is muffled through attenuation and filters, but the reverberations/reflections of the signal arrive clearly at the listener. Finally, if the direct path of the sound signal is clear but there is an obstruction between the reverberations from the sonic emitter at point *x* and the listener at point *y*, the sound is said to be *excluded* (Figure 5 bottom). The direct signal is clear, but the reflections are attenuated and treated with a low-pass filter.

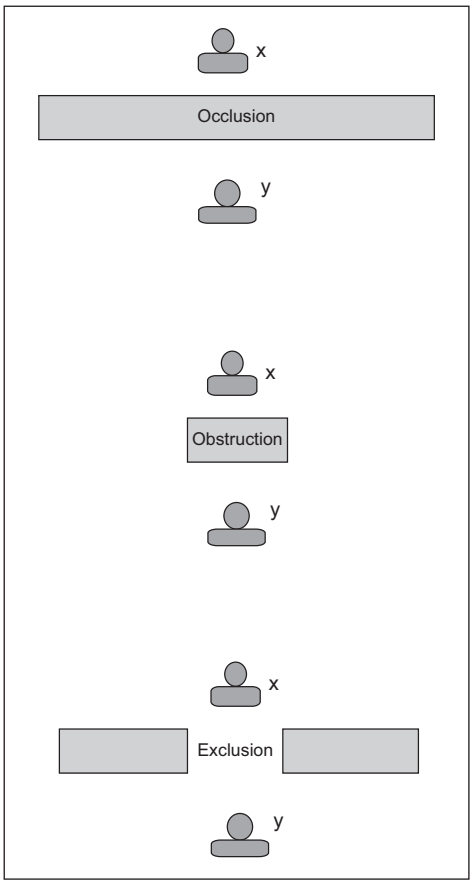


Figure 5. Top: occlusion between emitter *x* and listener *y*; middle: obstruction between emitter *x* and listener *y*; bottom: exclusion between emitter *x* and listener.

In other words, the use of filters to mimic sound propagation effects can help to create a sense of the physical, environmental space in which sounds take place. Such filters can also, of course, be used to create subjective distances, such as in a scene, where, for instance, a character is stuck inside a vehicle sinking in a lake. The character's banging and shouting on the glass not only illustrates their physical position in space, but reinforces the sense that they are separated from us (and their escape).

In *Ratatouille*, filter effects are used to create space and subjective positions. The first scene I'd like to focus on here is a back-and-forth between the human and the rat perspective. Remy is in the kitchen of Gusteau's restaurant, watching the chefs cook (at approximately 00:26:40). Remy hides inside an overturned metal colander as he watches an argument take place amongst the chefs over the soup to which he has recently added ingredients (which gets blamed on Linguini). As we cut back and forth between the rat's eye view and the kitchen, we switch between the natural ambiance sound of the kitchen, and the heavily reverberated sound experienced by the rat in the steel colander. The rapid cutting back and forth between the reverberated sounds from inside the colander and the relatively dry (untreated) sounds outside emphasizes the perspective of the rat, placing us in Remy's perspective during the times we see through the colander from Remy's point of view. This



Figure 6. Inside and outside the jar. *Ratatouille* (2007, dir. Brad Bird, Pixar/Disney, screen capture).

spatial distance reinforces the subjective distance of Remy, who wants desperately to be a part of the human world, but who must hide from it.

In addition to using such filter techniques to reinforce the physical and subjective space, it is also possible that a *lack* of expected filters can emphasize perspective. In a scene shortly after the one just described, Remy is caught inside a large jar by Linguini, who is ordered to dispose of the rat. Linguini heads out down to the riverside on his bicycle, jar in hand. He stops at the edge of the river and sits down on a wall. Once again, we hear a strong juxtaposition between the rat-world and the human-world. Remy's breathing and footsteps inside the jar are greatly exaggerated and treated with a significant amount of reverberation (see Figure 6, 00:29:30). After being afforded this brief rat perspective, the following discussion that takes place between Remy and Linguini – regardless of camera point of view – takes place from *outside* the jar, without any vocal treatment. Despite sometimes being in the visual perspective of the rat, we don't hear Linguini's voice as occluded. The lack of digital signal processing effects here places us outside the jar, on the side of Linguini, although this was most likely done for clarity rather than any intended effect.

II The subjective sense

As described above, the subjective sense is not segregated from the spatial sense. Spatial techniques that are used to create a sense of distance or physical environment also simultaneously can create a subjective auditory perspective. But other sound design techniques are also employed that are particularly effective at creating or reinforcing the subjective sense.

The subjective perspective

Auditory perspective can be created by emphasizing or focusing on specific sounds, through which character's subjective viewpoint we are hearing. Focal points can be created through volume and effects, but also through their presence or absence, and through their difference in relation to other auditory perspectives. For example, we might imagine a chaotic disaster scene, where despite extremely loud noises we can still hear someone whimper or a baby cry, as a focal point is created to lead us over to that person. Physically, we should not be able to clearly hear those sounds, but by emphasizing their presence in the mix, we are shown what we as the audience are supposed to focus on.



Figure 7. The subjective perspective. *Ratatouille* (2007, dir. Brad Bird, Pixar/Disney, screen capture).

Subjective perspective can also be used to give us a particular auditory perspective in a scene, allowing us to hear through a character's ears and thus identify with that character. In the first scenes of *Ratatouille*, we hear the rats from their auditory perspective: the rats speak in English, and in crowd scenes the 'walla' (background murmurs of the rats) are all in English. The rats break into a farmhouse kitchen, where the television news reports that the famous chef Gusteau has died. It is here that the auditory perspective first shifts (at about 00:09:40): Remy, stunned, asks aloud, 'Gusteau … is dead?' An old lady watching the television shuts the TV off and sees the rat, who has now sonically morphed into being a rat – squealing like a rat, rather than speaking English (see Figure 7). As Remy and his brother Emile try to escape from the woman, who wields a shotgun in an attempt to rid her house of the rats, the auditory perspective switches back and forth between that of the rats (speaking English) and that of the old lady, who hears only rat squeals. We are treated to the subjective perspective of the rats at first – they speak English so that we may understand what they are saying to each other, and to humanize them for us – but we also meet the rats from the auditory perspective of the old lady – as unwanted rodents. The subjective perspective is switched through the vocalization of the rats.

In the next scene, the rats escape onto rafts down a river, and the auditory perspective shifts back to the rats, with the tiny sounds of the rats paddling and the loud sound of the large raindrops on water, as we are transported into the tiny world of the rats. We hear the rats worrying about their babies and scrambling to safety as if they are tiny humans. Briefly, the subjective perspective switches back to the old lady, now wearing a gas mask and breathing like Darth Vader. We hear her breathing from her own perspective inside the mask, loud and deep. We then switch back to the perspective of the rats. Raindrops become giant sonar-like booms as the rats enter a sewer tunnel (at about 00:12:50). The sonar-like sounds, echoes of raindrops inside the tunnel, perhaps also hint at the super-human sensory abilities of the rat. We are then treated to a throw-back to the opening sequence of *Saving Private Ryan* (Steven Spielberg, 1993) as Remy is plunged underwater and resurfaces several times.[1] Here, we sonically follow the rat above and below, as the auditory perspective shifts from being above and below water with low-pass filter effects and deep bass rumbles (00:13:50). If we were not in the subjective perspective of Remy at this stage, we would not hear the underwater experience of the rat, we would continue to hear the tunnel. The shifting of our auditory perspective to that of the rat helps us to empathize with the character and feel like we are 'there' too. Moreover, the hearkening to live-action film (and through extension, 'real life') also contributes to the believability and credibility of the animated scene.



Figure 8. Internal subjectivity. *Ratatouille* (2007, dir. Brad Bird, Pixar/Disney, screen capture).

Internal subjectivity

Finally, we also experience the auditory perspective of various characters through hearing their internal dialogues, internal sounds and imaginary sounds. Unlike books, which often tell us what a character is thinking, we must often guess at a character's inner world in the movies. We are sometimes treated to glimpses of this inner world through various narrative and auditory techniques. I would suggest that these interior glimpses happen more often in animated film than in live action, since we are missing the nuances of human facial expression that can inform the audience of interior thought. There are many sequences in *Ratatouille*, for instance, which are narrated, allowing us to fill in the narrative gaps that cannot be shown. Likewise, Remy frequently speaks to an imaginary Gusteau who comes to life off the pages of the cookbook. 'You are an illustration. Why am I talking to you?' Remy demands of Gusteau (at about 00:14:40). These dialogues allow us into the inner thoughts of the rat, facilitating a greater sense of empathy.

In another scene, Remy tries to teach his brother Emile about food. As Emile is given tastes of different foods, a sense of elation is emphasized by a psychedelic series of colourful, musical flourishes (at about 00:55:10, see Figure 8). The bright lights are timed with the notes playing in the soft music, from a bass to finger snaps. This sonic and visual display illustrates the inner emotions of the rat as Emile experiences these tastes for the first time. These are internal, imaginary images, and the sound is thus used to emphasize feelings to which we otherwise would not be afforded access.

In a later scene, young chef Linguini has to serve a meal to the food critic Anton Ego (at about 01:23:32). In his dreams of how the evening will go, Linguini nervously asks Ego, 'Do you know what you would like this evening, sir?' to which Ego responds, 'yes, I would like your heart roasted on a spit.' Ego then laughs with a great echo, and we hear Linguini's heart beating loudly, reminding us of the interiority of this dialogue and, again, allowing us access to the inner feelings of the characters. Such use of reverberation effects helps to create an interior spatial sense. These 'internal subjective' sounds, then, primarily enable us to experience the emotions that the characters are feeling to which we otherwise would not be privy. While facial expression can go some way to facilitating empathy, the sounds are employed to create an immediate sense of the interior lives of the characters, adding depth and emotion that we might otherwise miss.

Conclusions

The auditory perspective is often lacking in discussions of identification in cinema. The use of sound to create this sense of identification is particularly important in animation, where we are often asked to identify with and empathize with non-human characters, and where the fictional world lacks the detail and nuances of the physical world. I have outlined here a variety of means of creating the auditory perspective that combine the spatial sense and the subjective sense, including microphone placement, positioning in the loudspeaker, digital signal processing effects, and subjective perspectives. These spatial and subjective senses are not neatly segregated: as shown, when it comes to the auditory perspective, the spatial sense often contributes to a subjective sense, and vice versa.

As with camera angle, the auditory perspective can allow us to experience the fictional world through the body of a character, thus enabling identification with that character. The techniques I described transpose and recreate the fictional space into our own physical space, helping to extend the world on the screen. The 'window' described by spectatorship theory is thus not an auditory barrier between the fictional and the real, but only a visual barrier. The shift in animation towards 3D computer graphics from cel shading, the renewed interest in stereoscopic 3D, and recent techniques that have been adapted from video games have helped to perceptually extend the visual screen (see Elsaesser, 2013; Jones, 2005). The role of the visual's influence on three-dimensional space in combination with these auditory techniques is outside the scope of this article, but would be an interesting area for future exploration.

There is still much work to be done to understand how the auditory perspective works in terms of our identification with characters. There are ways that sound is used in creating a subjective perspective that are not discussed here, such as some of the metaphoric uses of sound, the use of silence, and foreshadowing of emotional turmoil, for instance. Moreover, the auditory means of enabling identification and empathy may work differently in different types of movies and different visuals or camera angles. A greater understanding of the auditory perspective can thus expand our language of audiovisual media, and rectify some of the ocularcentric approaches to screen media that have dominated the scholarly literature.

Acknowledgements

This article was supported by the Social Sciences and Humanities Research Council of Canada in the form of a Standard Research Grant [410-2011-0997].

Note

1. It may be worth mentioning here that Gary Rydstrom won an Academy Award for Best Sound Editing in 1999 for *Saving Private Ryan*, and that the Omaha Beach scene has been repeatedly called the most 'realistic' and most 'memorable' battle scene ever, in a large part due to the sound (see, e.g., http://www.starpulse.com/news/index.php/2010/11/05/saving_private_ryan_has_the_most_memor).

References

Altman R (1992) Sound space. In Altman R (ed.) *Sound Theory/Sound Practice*. New York: Routledge, 46–64.

Amesley C (1989) How to watch Star Trek. *Cultural Studies* 3(3): 323–339.

Belton J (1985) Technology and aesthetics of film sound. In: Weis E, Belton J (eds) *Film Sound: Theory and Practice*. New York: Columbia University Press, 63–72.

Branigan E (1984) *Point of View in Cinema: A Theory of Narration and Subjectivity in Classical Film*. New York: Mouton.

Chion M (1994). *Audio-Vision: Sound on Screen*. New York: Columbia University Press.

Choe CS et al. (1975) The 'ventriloquist effect': Visual dominance or response bias? *Attention, Perception and Psychophysics* 18(1): 55–60.

Cleland K (2008) Image avatars: Self–other encounters in a mediated world. PhD dissertation, University of Technology, Sydney.
Cohen J (2001) Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society* 4(3): 245–264.
Collins K (2013) *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge, MA: MIT Press.
Doane MA (1980) The voice in cinema: The articulation of body and space. *Yale French Studies* 60: 35–50.
Elsaesser T (2013) The 'return' of 3-D: On some of the logics and genealogies of the image in the twenty-first century. *Critical Inquiry* 39(2): 217–246.
Flitterman-Lewis S (1987) Psychoanalysis, film and television. In: Allen RC (ed.) *Channels of Discourse: Television and Contemporary Criticism*. Chapel Hill: University of North Carolina Press, 172–210.
Friedberg A (1993) *Window Shopping: Cinema and the Postmodern*. Berkeley: University of California Press.
Gee JP (2004) *What Video Games Have to Teach Us about Learning and Literacy*. Basingstoke: Palgrave Macmillan.
Gerzon M (1974) Surround-sound psychoacoustics criteria for the design of matrix and discrete surround-sound systems. *Wireless World*, December.
Hall ET (1963) A system for the notation of proxemic behaviour. *American Anthropologist* 65(5): 1003–1026.
Holman T (2008) *Surround Sound: Up and Running*. Burlington, MA: Focal Press.
Johnson K (1993) The point of view of the wandering camera. *Cinema Journal* 32(2): 49–56.
Jones M (2005) Composing space: Cinema and computer gaming – The macro-mise en scene and spatial composition. Paper presented at the Imaginary Worlds Symposium, Sydney, December. Available at: http://www.dab.uts.edu.au/research/conferences/imaginary-worlds/composing_space.pdf (accessed 26 September 2013).
Lastra J (2000). *Sound Technology and the American Cinema: Perception, Representation, Modernity*. New York: Columbia University Press.
Lockheart P (2003) A history of early microphone singing, 1925–1939: American mainstream popular singing at the advent of electronic microphone amplification. *Popular Music and Society* 26(3): 367–385.
Manovich L (2001) *The Language of New Media*. Cambridge, MA: MIT Press.
McQuail D (1997) *Audience Analysis*. Thousand Oaks, CA: Sage.
Moore AF, Schmidt P and Dockwray R (2011) A hermeneutics of spatialization for recorded song. *Twentieth-Century Music* 6(1): 83–114.
Oldenbourg ND (2005) A conversation with Walter Murch. *Transom Review* 5(1). Available at: http://www.transom.org/guests/review/200504.review.murch3.html (accessed 26 September 2013).
Plenge G (1974) On the differences between localisation and lateralisation. *Journal of the Acoustical Society of America* 56(3): 944–951.
Rumsey F (2006) Spatial audio and sensory evaluation techniques – context, history and aims. *Proceedings of Spatial Audio and Sensory Evaluation Techniques*, 6–7 April, Guildford, UK. Available at: http://epubs.surrey.ac.uk/530/1/fulltext.pdf (accessed 26 September 2013).
Schenkman B (2010) Human echolocation: The detection of objects by the blind. PhD dissertation, Uppsala University.
Sound Works Collection (nd) Randy Thom: *The Sound of Ratatouille*. Available at: http://soundworkscollection.com/videos/ratatouille (accessed 26 September 2013).
Stillwell RJ (2005) Sound and empathy: Subjectivity, gender and the cinematic soundscape. In: Furby J, Randell K (eds) *Screen Methods: Comparative Readings in Film Studies*. London: Wallflower, 48–58.
Von Békésy G (1960) *Experiments in Hearing*, trans. EG Wever. New York: McGraw-Hill.

Filmography

Animation Sound Design: Building Worlds from the Sound Up, *Wall-E* DVD Extra (2008, dir. Andrew Stanton).
Ratatouille (2007, dir. Brad Bird).
Saving Private Ryan (1993, dir. Steven Spielberg).

Author biography

Karen Collins is Canada Research Chair in Interactive Audio at the Games Institute, University of Waterloo. Her work focuses primarily on sound and music in games, and explores how new technologies impact our audio experience. This research has included studies of sound on smart tables (horizontal computers), mobile phones, and video games, and how sound is used in slot machines to manipulate players. She has published three books on game audio: *Game Sound* (MIT Press, 2008), *From Pac-Man to Pop Music* (Ashgate, 2008) and *Playing with Sound* (MIT Press, 2013). She is currently co-editing a volume on interactive sound for Oxford University Press. Collins is also active as a sound designer, working on games and animated short films such as *Small Sacrifices.*