This presentation is part of a larger project that explores perception and imagination in interactive media, and how they are created by sound, image, and action. It marks an early stage in extending my previous work on the virtual kinoeye in virtual worlds and video games to consider aural, haptic, and kinetic aspects of embodiment and interactivity.
In particular, the argument is that innovative sound has the potential to bridge the gaps in experiencing embodiment caused by disconnections between the perceived body in physical, representational, and imagined contexts. To explore this, I draw on Walter Murch’s spectrum of encoded and embodied sound, interviews with sound designers and composers, ways of thinking about the body from phenomenology, video game studies scholarship, and examples of the relationship between sound effects and music in video games.

Let’s start with a question philosopher Don Ihde poses when he asks his students to imagine then describe an activity they have never done. Often students choose jumping out of an airplane with a parachute, and as he works through the distinctions between the possible senses of one’s own body as the physical body, a first-person perspective of the perceiving body, which
he calls the here-body, or the objectified over-there-body, the third-person view of one’s own body, he asks, “Where does one feel the wind?” Ihde argues the full multidimensional sensory experience is in the embodied perspective.

Now we ask the same question of a video game experience, rather than something we imagine. We can’t play a game right now, so we have to watch a recording of gameplay.

**Do you feel the wind? Where? Why?**

*Grand Theft Auto IV: The Battle of Gay Tony*
Rockstar North, 2009

Figure 3: [http://www.youtube.com/watch?v=tUyDjXX0aVY](http://www.youtube.com/watch?v=tUyDjXX0aVY)

Where do you feel the wind, if you do? There is no wind, although we can imagine the feeling of wind. There is image, and depiction of motion, but it is on the 2D screen in front of you. You can imagine the parachute jump, but the only sensory input you share with the embodiment
position represented on the screen is sound. You hear the wind, in 360 degrees, and
3dimensions, in your peripersonal\textsuperscript{1} space, literally and figuratively between you and the jumper.

Let’s experience another example, a very different game from Grand Theft Auto.

\textbf{Do you feel the wind? Where? Why?}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{flower_image.png}
\caption{Flower, thatgamecompany, 2009}
\end{figure}

\textit{Flower}, a game for Sony PlayStation3 so unusual that its creators “like to think of it as a video
game version of a poem.”\textsuperscript{2} is perhaps more of a new media experience or interactive journey
than a conventional video game. It’s never clear who you are in this first person game; are you
the wind? The embodiment of a flower petal? Something else? The visual, kinetic, and aural
aspects are intertwined through interactivity and coherent meaning, with the motion-sensing
gyroscopic Sixaxis PS3 controller changing the direction and speed of movement, which causes
sound, creating particularly strong “kinesonic congruence,” to use Karen Collins’s term for hearing sounds as if we created them through action even if we didn’t. Listen for combinations of the music and the musical sounds connected to flower colors or the wind, created by composer Vincent Diamante and sound designer Steve Johnson, an intricate and reciprocally informing of the experience. thatgamecompany included composer Diamante in discussions with artists, illustrators, and level designers, so that in parts where traveling through a row of flowers causes a string of instrument sounds to intermingle with the music, Diamante contributed to deciding the location and color of the flowers because sound, image, and action are so closely connected in the game.

PLAY FLOWER CLIP

Do you feel the wind? Do you feel you are the wind? There is the sound of not only the wind, but an intricate interactive sound design of the wind, and this brings me to the point of this argument. I imagine the feeling of the wind, I feel the wind in my imagined body, the one that is cruising through the grass, because sound is always embodied, always perceived in the here-body. As Frances Dyson argues in *Sounding new media: Immersion and embodiment in the arts and culture*, “Three-dimensional, interactive, and synthesthetic, perceived in the here and now of an embodied space, sound returns to the listener the very same qualities that media mediates: that feeling of being here now, of experiencing oneself as engulfed, enveloped, absorbed, enmeshed, in short, immersed in an environment. Sound surrounds.” Moreover, Dyson contends, sound has immediate physical and psychological effects because “to hear is also to be touched, physically and emotionally”—think of a high-pitched scream, or how low sounds vibrate in the body inducing panic. How does the sound in *Flower* touch us physically and emotionally?
Walter Murch’s aural spectrum

One of the ways is by blurring the line between sound effects and music. Walter Murch describes a spectrum from “encoded sound” like language which has to be decoded and “embodied sound” like music which is experienced directly. Sound effects, to quote Murch “fall mid-way, like ‘sound centaurs,’ they are half-language, half-music”⁵. In Flower and other games, when the conventional line between music and sound effects is blurred, when sound effects become part of the music, to be listened to directly, and music conveys information, both connected to the player’s perception of interactivity, sound enables the player to create not only a series of sounds, but a sonic experience through his or her interactions, to, in the case of Flower, make music, which contributes to a sense of embodiment, because it is the experience you make.
Let’s listen to how the sound designer and composer discuss their work.

PLAY CLIP 2 OF FLOWER http://www.youtube.com/watch?v=j7xPHyMnUw

On a different spectrum, this time of interactivity, *Flower* allows the player to shape a musical experience, in a very different but analogous way that the iPad game uses a generative music system so that player’s actions and choices make music. The game doesn’t sound the same twice because different game objects have different musical elements that combine in new ways, sonic equivalents of the structures the player builds to defend their hotel.⁶

BAD HOTEL CLIP http://www.youtube.com/watch?v=zU02yx4N2Q
Bad Hotel, a game with a high degree of what some term open or complex interactivity in the player's ability to make sound has little embodiment. Of the three kinds of immersion categorized by researchers: sensory immersion, with the audiovisual (and sometimes propriocentric) input overriding actual world input, challenge-based immersion that requires motor or intellectual skills, and imaginative immersion by which people psychologically become involved in the characters, environment, and narrative of the experience, Bad Hotel works primarily through challenge-based and sensory immersion, rather than imaginative immersion and the music the player creates is part of the fun, but does not contribute to a strong experience of embodiment. In terms of a more binary psychological vs. perceptual way of discussing immersion, it has little psychological immersion, with the hotel never existing as an imagined hotel, only abstract. Teasing out the relationships between abstraction in sound, image, and action is one of the directions this study will take in the future.
What has to be overcome to experience embodiment playing a video game? Current apparatus separates what happens in the virtual environment to the virtual character or avatar from the sensory input and the physical output of the player, and psychological and perceptual immersion are experienced despite the obvious fact that we are not really doing the action in the game, nor simply pretending or imagining. Is it first-person perspective? An avatar? Sensory, or challenge-based, or imaginative immersion predominating over the others, or a magical balance of the three? Lots of interactivity? Story? A strong diegetic world? I would argue that all these aspects of video game and interactive media and more are important, and can foster embodiment, but that sound plays a determining role. Let’s turn to an example of immersive new media art, *Osmose* by Char Davies for one way to join the gaps between the physical, virtual, and imagined sites of embodiment, with the imagined body compromised of the
conscious and nonconscious perceptual, conceptual, emotional, and sensory-motor processes that occur in the mind.

**unified sound, image, action in a virtual environment**

Char Davies, *Osmose* (1994-96)

[http://www.youtube.com/watch?v=I6_J6yo0Wc](http://www.youtube.com/watch?v=I6_J6yo0Wc)

Play video of Char Davies’ *Osmose* (1994-96)

Davies explained,

The sound in *Osmose* is spatialised and interactive in real time, responding to changes in the immersant’s location, direction and speed. My goal was to have sound that was neither literal nor illustrational, musical nor sound effect-like, but which was, in equivalence to the visual aesthetic, aurally ambiguous... in *Osmose*, sound and imagery have become one, amplifying each other.⁸
Interactive art stems from a common ancestor with performance art in the Fluxus and happening movements of the 1950s and 1960s. As the categories of art work, object, artist, written play, theatrical performance, performer, audience, theater, and gallery space were opened and intermingled, the emphasis shifts from the object or performance to the experience created collaboratively in the moment by participants. Since the experimentation with cybernetic technology in art that became more widespread with the rapid ubiquity of the personal computer in the 1980s, interactive art has come to indicate involving a computer system as one of the participants in a dynamic real-time interaction. At one end of a spectrum of interactive new media art are interfaces that use the computer screen and mouse (or other GUI selecting device like a trackpad or joystick), in the middle would be haptic, vocal, and kinetic controllers including the exponential spread of touch-screen mobile and tablet devices, and at the other are virtual experiences exemplified by Davies’s “Osmose” (1994-96) (and “Ephemere” (1996-98),) in which the body—breath and balance—controlled the interface (a vest that measured expansion and contraction of the “immersant’s” chest and degree of tilt) and the output was seen in a head-mounted display through “tiny monitors over the eyes to virtually enter inside the piece” (Morse 1998, 208), by spectators viewing what the participant saw in the head mounted display projected on one wall and their silhouette on another, and both hearing real-time sound generated by the participant’s navigation (Salter 336-337; for photos and video, see: http://www.medienkunstnetz.de/works/osmose/ ). Informed by the immersive experience of scuba diving, the fluid navigation of the non-Cartesian spheres of ocean, clouds, earth, tree, forest, and leaf of “Osmose” was experienced as a profoundly embodied emotional state of being (Grau 193-199), as intended by Davies’s phenomenological hope of creating a context “in which to explore the self’s subjective experience of ‘being-in-the-world’—as embodied consciousness
in an enveloping space where boundaries between inner/outer, and mind/body dissolve” (Davies, quoted in Grau 199). Virtual experiences like “Osmose” seek to keep the body involved as a site of sensory input and embodiment, in contrast to virtual reality interfaces that sought to replace the actual body’s experience with the virtual input.⁹

Davies’s use of sound seems to me to sum up the holy grail of sound in video games and virtual environments: audio unified with the visual and kinetic, generated in real-time by the participant’s action in the environment, emotionally significant, and thematically coherent with the work as a whole.

What *Osmose* does with its head-mounted visual display is to overcome the gaps between the physical, virtual, and imagined bodies that persist in video games and much interactive media that use screens and controllers. There is no distinction between the 2D screen image and the player’s immediate visual input, and the experience is controlled with breathing, not button-pressing. The sound dances between abstract and realistic, congruent with the visual, controlled by the body.¹⁰

Video games and other virtual environments experienced on screens and with typical game interfaces cannot rely on the kind of bodily interface that *Osmose* does for immersion. Nevertheless, despite the limitations of interface and visual input, sound *could* and ideally does function in video games in the way it does in *Osmose*, because, because sound, unlike image or action, occurs not in the 2D screen space or translated through haptic or kinetic action into virtual action via the interface, but in the actual world environment of the player, and is both 3D in its volume and dimensionality as well as being part of the 2D representation of the game world. As Mark Grimshaw explains, sound “exists and operates both in reality and in virtuality;
it has a real volume and dimensionality that is a 3-dimensional representation of the 2-
dimensional representation of the 3-dimensional world of the game.”\textsuperscript{11}

Grimshaw uses first-person shooter games to show his point, explaining:

the diegetic sonic world of the FPS game extends from the screen to physically
encapsulate the player in the acoustic ecology's real resonating space. This is particularly
the case where the player is using headphones because they serve as an extension to the
player's proprioceptive auditory system greatly attenuating, and in some cases entirely
blocking out, sounds external to the game world such that, for example, the sounds of the
character breathing become the sounds of the player breathing.\textsuperscript{8} Thus, FPS game
diegetic sounds extend the game environment from a flat, 2-dimensional screen to the 3-
dimensionality of the external world. The player's proprioceptive sounds are replaced by
the character's proprioceptive sounds and all other game world sounds envelop the player
as part of the game's real resonating space. These sounds form part of not only the real
resonating space but also the virtual resonating space of the game and thus help to
immerse the player, both physically and mentally, in the FPS game acoustic ecology.
By comparing the waveforms from an inactive and active player, we can see the
difference that interaction makes. What we hear is a barrage of sound effects in the gunfight, far
more direct sensory input than the first-person visual perspective gives us, and no music. The
first-person perspective is intended to embody us, to connect the hand and gun on the screen to
our physical body as extensions to another, imagined body.

Instead of Osmose, however, we more often get something like this:

Urban Terror clip: http://www.youtube.com/watch?v=N3tMh6nrGyw
This, “Endless Ocean: Blue World”, for the Nintendo Wii, is an example of basic game audio, functional in fostering embodiment using the different types of game sound but not in a way that heightens perceived interactivity, or makes the most of the possibilities of the medium. We can hear diegetic and nondiegetic sounds: the diegetic sounds include nondynamic environmental sounds, adaptive sounds, interactive diegetic sounds, and the self-produced sounds from the regulator of the breathing (which doesn’t connect to movement but does remind the player that you are underwater and have a limited supply of oxygen); non diegetic sounds include the music, which loops repetitively and is not very interesting to me, the nondiegetic kinetic gestural audio of the wiimote, and adaptive nondiegetic sounds. The relationship between the music and the sound effects is not profound or particularly immersive.
But there are gaps, or breaches, in the experience, all the ways that the hand and gun behind the information also displayed on the screen in text and image, and my actions with keyboard and mouse control action on the screen, but my physical body is certainly not running or pointing. I am disconnected from an embodied experience, perhaps because the game is more encoded than embodied, more like language than music.

Many thinkers, in one way or another, have noted the different spaces and bodies that matter to the person engaged in virtual worlds, telematic interactions, and video games: the physical space where the computer or device and the physical body are; the screen space and whatever image-body is there, and then there is a third space, that of the imagination, which has been explored as a social “third place” in MMOs and the imaginary body. Frank Biocca, who researches human-computer interaction in virtual environments, concludes that the sense of presence oscillates around three poles of the physical (distal immediate), virtual (distal mediated), and imaginal (minimal attention to distal stimuli, like in dreams) environments. There is a body in each environment.

We can say that in almost any virtual environment system with any significant level of embodiment, there are three bodies present: the objective body, the virtual body, and the body schema. These three bodies may be present even in comparatively primitive, non-interactive virtual environments like standard television ([Meyers & Biocca, 1992]). The objective body is the physical, observable, and measurable body of the user. The virtual body is the representation of the user's body inside the virtual environment. The body schema is the user's mental or internal representation of his or her body.

Biocca uses the term body schema to include both the non-conscious system of motor-capacities that enable us to move as well as what some distinguish as the body image, the perceptions and
beliefs about one’s own body. The difference is a level of awareness. For my purposes, considering together the conscious and non-conscious aspects of the body in the mind is fine, as they are mutually informing, and we can call that the imagined body, an umbrella for the perceptual, conceptual, emotional, and sensory-motor processes that occur in the mind.

But it is not only through the sophisticated intertwining of sound effects and music, or the player’s capacity to make music through interactivity that sound bridges gaps between the imagined body, the actual body, and the virtual body. In Limbo, sound and silence are used for narrative and emotional purposes to create a powerful sense of embodiment and heightened emotional response.

**LIMBO, Playdead, 2010**
In the following clip, notice how although we are always visually distant from the boy, the sound is immediate, and changing according to circumstance as well as environmental realism. Instead of music throughout, there are times when there are only sound effects, like footsteps. *Limbo* sound designer Martin Stig Anderson used the prominence of certain sounds to indicate their importance in the game environment to the character, separating sound from visuals so that you hear “approaching obstacles and environments even before they’re revealed visually, and as you pass them they may be silenced entirely although they may still be in the frame thereby revealing new obstacles or environments to come. Besides contributing to the foreboding atmosphere of Limbo, such mixing minimizes the risk of making sounds become annoying to the player, simply because the sounds only play as long as they’re important to the actual game.”

Anderson explains his strategy: “For me it has a much bigger psychological impact when you turn a naturalistic soundscape into abstraction by making your sound effects play as “music” rather than adding some traditional background music. Moreover, making your “music” emerge from the environment is likely to make the audience more forgiving towards it since they’ll accept it as stemming, however abstractly, from the environment. This feature attains special relevance in video games where the player may get stuck from time to time and the audio elements need to be flexible in terms of duration.”

http://www.youtube.com/watch?v=rOExZpv pocY

Sound in LIMBO, therefore, moves sound effects closer to music on Murch’s spectrum, into the orange range. The abstracted sound, more orange than yellow on Murch’s spectrum, connects to the imagined body, and is affective.

“The footsteps of the avatar – the walking boy – are played with a -6dB attenuation. Once the boy walks on a different terrain surface, the footsteps change according to the new terrain but are
also played with 0 dB attenuation thus louder. The louder footsteps sounds are considered to be relevant for the player so he knows the surface has changed. Gradually, the footsteps will become softer until the -6dB level has been reached. This prevents the footsteps from becoming overly present.¹⁸

The visually-distant yet affective Limbo points out how the screen body is both the visual and kinetic representation on the screen and is also in the player’s imagination; this is why I call it the screen body and not the virtual body, because the virtual body is not only on the screen, but also in the imagination.¹⁹ The representation on the screen is not really the over-there-body, and mind’s version of the avatar is not exactly the imagined body of the player; we shift between these, and innovative uses of sound smooth the shifts, using aural input to connect the player to
the physical and emotional experience of the avatar body, as it reinforces and amplifies visual and haptic aspects of interactivity.

Ico

Like LIMBO, ICO’s sound design makes sound effects prominent over music to create the 3D environment. ICO game designer and art director Fumito Ueda explained, “The use of background music has been kept at a minimum in this game. We only use music in areas we really feel it will be effective. This helps in bringing out the great essences of the world of ICO.” The third dimensionality of the game space is represented visually as well as aurally, and there are variations in scripted camera positions during gameplay as well as in the cut scenes that guide the visual input as well as narrative. Here, sound effects position the player in the space, with emotional resonance, but it is the music, cued to the princess, that supplies the affect.
Watch and listen.

PLAY ICO CLIP

http://www.youtube.com/watch?v=JxiXQSRle5c

We hear the music with special emphasis after we are used to listening to the sound effects so intently for information about the environment, after listening to encoded sound, strengthening the meaning-making of the music.

We’ve seen examples of both first-person and third-person perspective games. Does seeing an avatar make a difference? There is more information, and sometimes more choice in camera control, from the third-person perspective, whether the follow camera position closely aligned with an avatar’s physical point of view, or from another vantage point. One of the games I’ve been following is EVE Online, a MMPORG of particular interest to me because only recently have players been able to be represented by avatars who “walk in” space “stations,” previously being represented by the ships they were in (other than in profile pictures). Although ships and other objects, especially vehicles, can be extensions of the body, the audiovisualkinentic rhetoric of the game did not seem to deliberately encourage a strong embodiment. In addition to limited experiments with avatars walking in stations, EVE seems to be adding to embodiment by strengthening the kinesonic aspects of the game, with more attention to soundscape. Most significantly for this discussion, the approach to game music shifted from an ingame jukebox to music connected to the context of place and action, some procedurally. Other changes in sound, according to language in the development blog, include “gratifying sounds” “to get that sense of accomplishment” after finishing a mission or training.²¹ It will be interesting to see how EVE Online develops, on its current platform and the new, connected PlayStation first-person shooter set on the planets the ships fly past, and the role that sound might play in its quest for greater embodiment.
EVE Online

http://www.youtube.com/watch?v=zo0G1V0Nioc
To conclude, let’s turn to Portal 2, a first-person shooter-puzzle game, with brief glimpses of Chell, the character we play. The unified sound design is characterized by sound effects and music that take on the characteristics of the other, some triggered interactively and others not, but both contributing to a sonic experience that does not recede into the background. Dan Bruno, a composer who writes about sound in video games, uses Murch to describe how a game like Portal 2 “pushes sound effects and music towards each other.” Here is an example:

PLAY AUDIO CLIP

The experience of Portal 2 plunges you into an environment where polarities are not what they might seem, in all possible ways: up or down, in or out, behind and in front, of course, but also what is rational, emotional, abstract, concrete, useful, unnecessary, mechanical, organic, are
all in flux. Chell is a center of calm, silence, action, and the sound design conveys this. Writer Geoff Keighley describes sound and music:

The first *Portal* was renowned for its musical ending, and in *Portal 2* composer Mike Morasky wanted to up the ante with interactive music that would subtly evolve as players completed a puzzle. Run along orange speed paint and the music speeds up. Successfully jump across a ledge and the music shifts to let you know you’re doing a good job. “The puzzles are thanking you for playing with them,” is how Morasky puts it. “They love you.”

This is brilliant! The puzzles love you, and the way the game indicates its emotion is the sound. It is not surprising then, that this game would invest the nonplayer characters with something out of the ordinary, and here I’d like you to go with it as I consider speech as sound.
Murch writes, “To the degree that speech has music in it, its ‘color’ will drift toward the warmer (musical) end of the spectrum,” and indeed it is the tone, the music, of the nonplayer characters that become part of the unified sonic experience. Brilliantly voiced by British actor Stephen Merchant, eyeball-robot Wheatley provides company, comedy, and exposition while the player solves the series of puzzles, but in a surprisingly natural, neurotic, and humanly-flawed way. Given that the protagonist, Chell, is silent, as so many characters in single-player games are, there is no interactive conversation between you/Chell and Wheatley, but he is so cleverly scripted, performed, and programmed that it feels like he is not only responding to your choices and outcomes, but interpellating you. Of course we decode the language that we hear, but it also
pushes the encoded sound of the dialogue closer to the embodied/encoded hybrid of sound effects.

**Portal 2, Valve, 2011**

http://www.youtube.com/watch?v=y4lYWBHAZyE

Wheatley is not the first character whose speech is not encoded sound but embodied sound; think of Mario and Luigi, or even of the princess and boy in *ICO*, who speak different made-up languages, with only Ico’s words translated in subtitles. But in Portal2, silence for Chell is more than conventional first-player silence; like the restraint in using music only, in Murch’s terms, in the red part of the spectrum in favor of sound more in the orange part, Chell’s lack of speech is meaningful in what it allows to take the prominent position in the sound design. As Portal2 head writer says about Chell, “She probably can talk, though. In our minds, she can talk. She’s just
pissed off the whole time and is refusing to dignify any of the things going by speaking about them.”

“Technology mirrors our desires; interactive technologies, in particular, reflect our desire to feel engaged.”

David Rokeby, “Transforming Mirrors: Subjectivity and Control in Interactive Media,” 1996

Artist David Rokeby commented, “Technology mirrors our desires; interactive technologies, in particular, reflect our desire to feel engaged.” Although embodiment is not the only goal of interactive media, it is often a way we measure engagement. Sound, especially when music, sound effects, and even speech are interwoven innovatively in interactive media, is perhaps the most important of the triumvirate of sound-image-action because of its impact on embodiment. I suspect that when we overlay and intersect spectrums for image and action similar to Murch’s, on which I’m working, we may have a very intriguing model for understanding embodiment and engagement in interactive media.
Another example is how the sound of the wind, connected to the wind, gives hints for game objectives that are holistic in the game world.


(Morse 1998, 208) Virtualities: Television, Media Art, and Cyberculture
By Margaret Morse
Mark Grimshaw, **SOUND AND IMMERSION IN THE FIRST-PERSON SHOOTER**

**THE ACOUSTIC ECOLOGY OF THE FIRST-PERSON SHOOTER**

---


Permalink:


17 Listen to the relationship between sound effects and music in this next clip of gameplay towards the end of the game.

PLAY CLIP 2

---

18 http://captivatingsound.com/game-developers-conference-2012-audio/

19 The imagined body of the avatar in the player’s mind is neither based on nor separate from the physical body of the player, and that is where game audio has its most impact, on the imagined experience of the avatar body.


**Interview with Fumito Ueda**

21 http://community.eveonline.com/devblog.asp?a=blog&nbid=73569

22 in his behind-the-scenes making of Portal 2 feature “The Final Hours of Portal 2”


24 “ (“Transforming Mirrors: Subjectivity and Control in Interactive Media” 1996