



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

[Delbridge, Matthew](#) (2013) The ecological approach to visual perception and the actor performance captured in the gaming landscape. *Animation Practice, Process & Production*, 2(1). (In Press)

This file was downloaded from: <http://eprints.qut.edu.au/41189/>

© Copyright 2013 Intellect Ltd.

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

The Ecological approach to Visual Perception and the Actor Performance Captured in the Gaming Landscape

Matt Delbridge, Queensland University of Technology (QUT)

matthew.delbridge@qut.edu.au

Abstract:

The use of Performance Capture techniques in the creation of games that involve Motion Capture is a relatively new phenomenon. To date there is no prescribed methodology that prepares actors for the rigors of this new industry and as such there are many questions to be answered around how actors navigate these environments successfully when all available training and theoretical material is focused on performance for theatre and film. This article proposes that through a deployment of an Ecological Approach to Visual Perception we may begin to chart this territory for actors and begin to contend with the demands of performing for the motion captured gaming scenario.

Keywords:

Motion Capture, Gaming, Performance Capture, Visual Perception, Actor Training

1 Introduction

“the earth before animals evolved was not an environment properly speaking, it was a potential environment”

(Gibson, *An Ecological Approach to Visual Perception* 1976 p3)

Performance Capture is a term first employed by the Director/Producer Robert Zemeckis during the Warner Bros production of *The Polar Express* (Zemeckis, 2004) used to describe the total recording of a performance without cuts using a Motion Capture system. Performance Capture is inherently theatrical in that it allows for a performance in its entirety to be captured in one take allowing for all traditional framing questions and dramatic

devices to be employed after the performance has been recorded. Essentially this captures the entirety of an actor's performance allowing the subsequent filmic notions of selecting frame and edit points, (now essential components of all modern gaming environments) to be completed after the act. It deploys a much more theatrical approach to performance that allows the actor to explore the entirety of a scene in real time unhindered by the encumbrance of the focused device (the camera). Now commonly used in the high end video game industry, this allows for a freedom of performance for actors that is not hindered by the constant hurdles encountered in film production where actors are continually repeating small sections of dramatic storylines or waiting for physical environments to be reset or reframed.

Technical language as a form and a particular kind of vision tool is an essential part of understanding Performance Capture for the gaming industry. As a tool this language needs to be mastered, not only by the operators of the interface but also by the users at the beginning (the animators/actors) and the users at the end (gamers). Both in and of itself, and as particularly focused (especially as developed over recent years to keep pace with the advances of digital technology) the language/s of the environment to be described here must be deployed appropriately to discuss the sorts of ideas related to performance for video game production. This new technicity of language is a vital part of the actor's navigation of the technologised performance capture environment and key to the developed understanding of the dilemma actor's encounter in the navigation of the gaming spaces associated with content production.

This relationship of technicity and language is best introduced through Larry Hickman in his discussion of the work of 1930s educational pragmatist John Dewey in *Philosophical Tools for Technological Culture – Putting Pragmatism to Work* (2001), Hickman is paraphrasing Dewey's philosophy:

Knowing is also relative in the sense that it involves connections to other knowers. Knowing is sharpened and extended by taking the stances or viewpoints of others within a community of inquiry, that is, by considering a problem from as many differing perspectives as

possible. Thinking, language, and knowledge are all community enterprises, both in terms of their historical development and in terms of their ongoing function of construction and reconstruction (Hickman 48)

The connection to other knowers, specifically in relation to the motion captured environment, can be re-appropriated to include all users of the system. This point is important when we consider that within the unique ecology of a motion capture studio there are many parties that may have never had the opportunity to encounter each other before. Particularly when we place animators alongside actors and technical operators alongside fight scene directors (who themselves may have never directed formally trained actors before), this connective knowing needs to be established within this habitat. The particular technicity of the environment is the common community, and as Motion Capture (and indeed the capture of an actor's movement for games) remains a relatively new form, there still exists an absence of common knowing that all participants are a collective part of. The standard frame capture rate (or frequency) deployed in film - that is the rate at which individual frames or images are captured consecutively, is 24 frames per second (or fps). An individual Motion Capture Camera deployed as part of a Motion Capture system (that may involve up to 100 individual cameras in a dedicated network or array) captures at a standard rate of 250 fps. If we consider that the first silent film camera was capturing at around 16fps (moving up to 24fps with sound) in the chemical/mechanized scenario and now the actor is faced with a capture rate that is 10 fold that recent experience (in historical terms), it my suggestion that 'more' of the actor is now being captured by the device. As the captured 'aura' of an actor's performance can be broken down into 250 frames of individually consecutive images per capture (per second) then this creates an astoundingly large body of material that is capable of being reproduced, reframed, targeted and manipulated in the creation of a characters real time response in the game. This in and of itself represents a unique and disciplined challenge for the actor working in this space and highlights a new focus that resides outside of the traditional training actors receive.

The first two layers of the actor's *aura* as traditionally captured, can be classified as the actor's image (the first layer) and the actor's voice (the second layer). As we move beyond the capture, relay and storage of the visual and the aural in traditional framing terms, and into the gaming territory that now also captures the plotted movement of performance within omniscient framing environments, we enter a landscape populated with a third layer of captured aura. This third element, captured frameless movement, is an addition to the layers of performance aura to be stored in the capture device and the associated effects of this are central to informing the background of this discussion. A primary connection exists between the integrated perception of environment that this article discusses (based on the visual map developed through movement about a 3D gaming landscape), and the concept of an actor's movement that can now be captured outside of any formal frame.

Where frames set the optical terms of cinema, in relation to the more contemporary technology of motion-capture, we should consider the concept of the 'capture volume', which in motion capture is the amount of 3D space that the system can 'see.' A translation of physical space to screen based 3D space; this is determined by the placement and settings of the capture devices (cameras) and their distinct relationship with each other as separate units. Depending on the capture that is being undertaken, the size of the volume will be adjusted. The variables involved could include: the amount of objects to be captured, the nature of the performance that is to be captured (either game cut scene, in game action) or the physical properties that are required in the space for performers to interact with. On this point it is worth noting that if a particular character needs to be captured sitting at a desk writing, climbing a rope or performing any other task that will involve interactions with static 3D objects, then the best way to achieve this is to physically have the actor sit at a desk or indeed climb a rope placed in the volume, remembering that it is only their movement in space that is recorded and not a visual image of the physical object. The establishment of the volume is a vital early step in the profilmic setup as any character or object

performing outside of this volume (in whole or in part) will either not be captured at all or their individual template will turn into an unmarked data stream or cloud of ghost markers.

Initially the term animation is not as simple to define as we might think. In brief terms and in a broad collation it can refer to the various conditions of living, the properties of being able to grow, the qualities of being active or vigorous. Animation can be used in reference to the display of a set sequence of images, to the creation of mimetic movement or a visual illusion of activity due to the phenomenon of continually focused vision. Animation is primarily concerned with the endowment of the features of *animal life* as distinguished from *plant life* and most importantly of all the giving of energy to (and the activation of), a situation or circumstance. Animation is tied to the act of movement, specifically animal movement within a landscape. So then to animate a character is to imbue that character with movement, to bring it to life and to energise the inanimate into being. Out of these ideas about animation we can find a link to the dilemma associated with the camera obscura's first *animated* scenes. The camera obscura occupying the space of the first live stream, the first connection between spaces in real time and the first passive gaming experience of looking through the window of the real via the virtual. It is from this first discovery of the power of animation (the real replicated onto the screen) that has identified the beginning of the actors challenge to populate this animated space and in the present day it is within the motion capture studio, where the actor's movement is captured for later manipulation in the gaming environment that the dilemma resides today.

The work of ecological psychologist James J Gibson and his theoretical pursuit for a redefinition of the conceptual frame that determines how visual perception is formed, offers a unique perspective to the discussion of the actor in the performance captured environment. Throughout this discussion and with this in mind, I specifically define *environment* in ecological terms; that is: the place where the cognate live, or 'the surroundings of those organisms that perceive and behave, that is to say animals' (Gibson 1976 p 7).

It is often neglected that the words animal and environment make an inseparable pair. For Gibson, each term implies the other: 'If no animal could exist without an environment surrounding it, equally though not so obvious, with an environment implies an animal to be surrounded' (Gibson p9). This quote goes in some way to explaining the ecological significance of the spaces constructed and used for live performance (that is performance to be viewed through a live, unedited human to human interaction like the theatre) and performance capture environments (in general terms modes of performance for film production, audio production and other digital mediations like Motion Capture that involve a machinic or digital interface) that can only be framed as potential environments due to their undeniable reliance on the animal (the actor). This connection, the binary of the actor and audience in the theatre, or the ternary of the actor, operator and audience in the mediated scenario, establishes performance-ready spaces as ecological in nature.

Within performance we address this environmental question continually. The theatre actively acknowledges the presence of both performer and spectator occupying the same space at the time of performance. Through this mutual occupation of the same space they create, in Gibsonian vernacular, *environmental conditions*; that is the condition of the animal surrounded by an environment. In film, the place of the animal in the environment has less to do with the audience in the live context (this relationship is formed after an editing process) and is more immediately concerned with mechanical invariants that frame the landscape (a term I will describe later in this section) and the place and co-function of the operator(s) environmentally.

It is in the digitally mediated motion capture scenario, where the relationship with the environment is on the whole less concerned with a capture of framed image or mechanized invariance, that an overall integration of supplied movement by the animal affords the most complete environmental integration. Through Motion Capture, the movement dilemma associated with visual environmental navigation can be explored and captured, confirming the link between sight and movement as key components of the establishment of vision.

Through a lens borrowed from ecological psychology (specifically Gibson's *An Ecological Approach to Visual Perception*, 1976), there are particular areas of performance landscapes that can be defined in terms of *surfaces, mediums and substances*. It is worth spending a little time defining these terms, as they are central to an understanding of the ecological perspective. This is particularly important in relation to recognising notions of difference in the various forms of performance environments that currently exist and the potential differences in work that is demanded from the actor in these environments. The ecological reference to a surface refers to Gibson's characterisation of 'interfaces', specifically in ecological terms the interface being the area between two substances (or mediums) i.e. earth and water (at the bottom of an ocean), water and sky (at the horizon):

'the surface is where most of the action is, the surface is where light is reflected or absorbed, not the interior of the substance. The surface is what touches the animal, not the interior. The surface is where chemical reaction mostly takes place. The surface is where vaporization or diffusion of substances into the medium occurs. And the surface is where vibrations of the substances are transmitted into the medium'. (Gibson 1979, 19)

A surface is primarily the point of this interface, where two mediums intersect or meet. In the Gibsonian world the intersection, or meeting, is observable only through the presence of light and can also only be completely defined through a moving observation (as opposed to an observation that remains static and framed). This concept of movement is fundamental to the establishment of a thorough visual perception of a landscape, and is a key factor in the link between the work of Gibson and the initial navigation of current vision-based performance capture environments.

A substance is the easiest of the three terms to define. Within ecological psychology, a substance is opaque, cannot be seen through and reflects light. The substances that exist in performance landscapes can be broadly (and

admittedly generally) described as lighting fixtures, drapery and screens, floor coverings; including tarket and portable staging/rostra as well as other smaller objects used for particular functions like speakers, props, costumes and cameras. Every substance has a unique surface, beyond the definition of the interface, and this surface is subject to change dependent on the position from which it is perceived.

In the ecological frame a medium is best described as that which facilitates movement through the environment. For different types of animals this can mean different things. For fish water is a medium, for birds the air is a medium and for other animals the ground is a medium. In a performance environment the concept of the medium or *media* is particularly vital and central to the deployment of the ecological terminology.

The thing which facilitates the movement through the environment, whether that environment be a theatre, film studio or motion capture lab at its most fundamental remains the ground, but in the consideration of performance environments existing beyond the general ecological definition the concept of a medium could also be related to an actor's particular singular or combined set of performance techniques and training. In Gibson's ecological terminology this idea of media(um) is best exemplified through the following: 'I have described the environment as the surfaces that separate substances from the medium in which animals live' (Gibson 1979,127).

For Gibson, the conventional scientific approach to visual perception reduces images to 'flattened out objects, a sort of pancake of a solid body' (Gibson 1979, 119). Gibson states that we in fact do not see "images" at all, indeed the concept of the snapshot image is a human construct thoroughly influenced by the hangover of a complicated history of optical studies. Notions of snapshot vision, which involves the momentary exposure of a stimulus or pattern to the eye, or aperture vision involving the scanning of the pattern by the eye, all assume that the eye works like a camera, a camera that remains completely still within a particular setting. While in some ways there may be elements relevant in broader optic studies, Gibson suggests that the visual

system is a lot more complex: 'evidence suggests that visual awareness is in fact panoramic and does in fact persist during long acts of locomotion' (Gibson 1979, 1).

For Gibson what we actually see, or acknowledge as sight, is best framed by what are called 'affordances' provided by the ecological environment. The features that we see in the environment are inferred or learned, and are concerned with particular modes of movement, memory, recognition, nesting and scale. Gibson classifies vision as inextricably linked to two particular modes of movement and further argues this type of sight as central to environmental navigation and perception; these modes of sight are *ambient* and *ambulatory* vision. Ambient vision is where the viewer scans an environment by moving their head about in a stationary position, while ambulatory vision involves the same continuous scan but is undertaken while moving. These are described as *normal vision* techniques that sit outside of the traditional lab test of vision where a subject sits in a chair with a headrest limiting any movement. Instead of viewing vision as a series of particular frames or snapshots in sequence, ambulatory vision provides a particular visual flow that completely discards notions of a flat visual field.

The next key concept to understand in an ecological approach to visual perception is invariance. Within the environment there are objects that move (like other animals) and objects that do not. The objects that do not are what Gibson calls invariable and are used as anchors to establish a particular visual field. The horizon in particular is an invariant often used as a visual locator. Gibson also refers to the parts of our own body that we can see as we move about as invariant, like the part of our nose that we can constantly see or our hands, feet and legs that come in and out of sight as we move. This notion of the invariant in the landscape is also key to an individual's understanding of personal balance and place environmentally, as it is the constant in the occupied landscape that serves as a marker for defining a sense of place and provides the necessary stillness that is at the heart of providing balance.

Further philosophical nuance can be given to Gibson's ecological theory via the work of the epistemologist Avrum Stroll. Stroll has written extensively on the work of Gibson and there are questions to be asked about some of Gibson's complete definitions, particularly of a surface, when described as the interface between two substances. Stroll questions the physical existence of the interface. The horizon, for example Stroll states cannot be physically seen, or touched, or cut: 'what divides the atmosphere from the water must be a common boundary, which is neither air nor water. Such a boundary is therefore not part of either state of matter... moreover such a common boundary must be without substance' (Stroll 1987, 450). In this instance Stroll refers to surfaces as 'conceptual entities only', (ibid) yet as we enter performance spaces that are virtual in nature, that are accurate representations of real space in scale and geometry, though cannot be physically touched or cut (yet can most definitely be seen), we are no longer in conceptual territory. Screen based representations of the virtual, discount the idea of conceptual surface existence only. The screen interface between the real and virtual moves to the heart of the profound influence Gibson's ideas of environmental navigation can have on our discussions on current digital performance environments. It is the navigation of the screen-based environment (and indeed the classification of the screen as a potential environment) in real time Motion Capture scenarios, which demonstrate the banishment of the conceptual, placing the representation of self on the screen into an environment that is real. The animal occupies the landscape. When an actor's movement is fed live from a real space into a 3D space, the interface between the actor and screen environment remains both the surface of the ground (the real) and the surface of the screen (the virtual).

The ecological account of vision is primarily concerned with what can be seen with the naked eye and is not a perception enhanced by any mechanical or digital apparatus (or *techne*). It is an environment where ordinary persons act and interact with familiar objects in mind and is subject to particular notions of what Gibson calls ecological *nesting* and *unit constants*. It is within similarly applicable landscapes that nesting and unit constancy take the place of tool-based measurement techniques and this is an important contribution to our discussion of the actor in the environment.

Nesting and constant unit principles, where ‘canyons are nested within mountains, trees nested within canyons, leaves nested within trees’ (Gibson p9), allow for a relative sense of unit scale to be deployed. This sense of scale imbeds the measurement of the individual components of an environment within each other to afford an imbued sense of universality applicable to other comparable environments. This notion of affordance, the concept of what the ecological, or familiar environment *affords*, is a key element of Gibsonian theory applicable to the initial navigation for the actor working within the geographically, socially or professionally unfamiliar but recognizably performative habitat. It is this common environmental affordance, based on a particular theatrical visual perception of a unit constancy that makes it possible for the actor to locate himself or herself. There are many examples of these sorts of unfamiliar but still navigable performative spaces; the sorts of environs that could be described as found performance spaces, i.e. repurposed and converted industrial spaces deployed by performance collectives like Shunt (see www.shunt.co.uk). Where the particular *affordance* for the performer is based on the temporarily installed theatrical fittings or the place and existence of a live audience. An actor’s learnt environmental perception enables them to reorient successfully, and appropriately *nest* themselves within performance environments (either live, mediated or for capture) through an active deployment of these ideas.

The actor *nested* in the space is applicable to Gibson’s “tree in the canyon”, with attributable concepts of relative scale occurring (though man made), as described in the *unit constant*. This assertion that in the terrestrial environment there is a sense of universal scale and measurement is one that can be applied to the un-navigated actor (i.e. the actor encountering a particularly unique performative environment for the first time), dependent on the environment described. This environment could be a theatre in a foreign country, an exotic outdoor location on a film shoot or a studio environment where movement for cut scenes in video games is created. Gibson’s *constant unit* principle prescribes that a grain of sand, or a pebble, or a boulder or even a mountain is more or less recognizable and classifiable anywhere ‘these

natural units are not of course perfectly uniform...nevertheless even if their repetition is not metrically regular, it is stochastically regular, that is to say regular in a probabilistic way.... a blade of grass is a blade of grass' (Gibson 1979, 10). This principle, applicable to performance environments, may provide the fundamental solution to navigation and universal classification within the digitally mediated space like Motion Capture environments. Before the current study there has been no focused study that has gone beyond the identification of the potential dilemma the performer faces when navigating the digitized performance space. The work of Gibson provides a particularly focused insight into how we might begin to frame the navigation dilemma for actors, how we might approach a discussion that places the actor in the familiar as opposed to the foreign and begin to use the ecological as a way defining the 'unnaturalness' of the performance environment encountered in Motion Capture studios. This approach however only goes a small way in unraveling the expectations that we might come to expect from the actors placed within these hyper real environs. There has been no methodological approach developed that has sought to address this through either:

- Proposing a specifically tailored mode of examining current spatial similarities and differences in technologised performance environments,
- An interrogation of the required (potential) training actors require for these spaces,
- A detailed survey of performance styles from the past may be re-appropriated and revitalized within these environments
- An approach to targeting artistic, dramatic, scenographic and spatial direction that remains in step with the technological shift associated with this age; specifically in the captured performance of actors in animated features and video games

While Peter Brook has walked into empty spaces and called them theatres since the 1960's and Anne Bogart has appropriated the work of Mary Overlie in the mapping of personal processes for actors through the 1980's there has been little contemporary discussion addressing the actor negotiation of the physical and virtual properties associated with performing in hyper

environments like Motion Capture for gaming and while important for the future of the industry, this discussion has only just begun.

The Actor and the environment: three parables

I

The actor walks into a space. It is a theatre, an environment he understands. He understands because he can perceive himself within this landscape. The elements that compose this environment are not only known to the actor, but also have set and established habitual relationships with each other. The actor brings to this place his training, his personal and professional history and memories of this and similar environments. These are his reflective surfaces and substances that can be applied in this space. He has a set focus in this theatre based on the known relationship between the stage and audience and can be confident in his knowledge of the theatre *nest*.

There is a common *nesting* in all theatres, a common sense of *an ecological scale*; this allows the actor to perceive his environment, to inhabit this space and locate himself within it. The actor is aware how an audience will perceive him in this environment. While he deploys an ambulatory vision on the stage (where the audience provide invariance), for the audience he is the invariant and supplies a unit constancy that allows them to deploy their own individual ambient vision. His behavior is dictated by his localized and trained sense of visual perception and this perception provides him with a memory beyond an internal kinesthetic awareness and is more focused on a complex external understanding of the surroundings.

The complex nature of the theatrical environment is unique in environmental terms. It is a specialized environment that maintains its own sense of scale, unit constancy and invariance. Standing alongside the norm of human habitats the theatre takes a significant amount of attention to navigate, learn and inhabit. It is spatially, technically, hierarchically and socially unique. The same of course could be said about any specialized environment, like an automated car plant or yoghurt factory, but for the theatre (unlike the factory) there is a relative sense of nesting and unit constancy that can be applied

across most performance environments. This is where Gibson's terms are so helpful in ascribing a solution to the dilemmas associated with the place of the actor in the digitally mediated performance environment. For the trained actor, the theatre is a known type of habitat, complete with all of the environmental associations that enable an active sense of visual perception and negotiation to be deployed. There is a recognized unit constancy of comparable stage lighting systems, counter weight fly systems, stage prosceniums and orchestra pit boundaries in most theatres. This existence of unit constancy facilitates the actor's entry and ongoing location in the theatre environment. Obviously however not all performance spaces contain all of these units. Many found spaces enabling performance have none of these units present. The constant (or invariance) for the actor (regardless of theatrical trappings) remains the audience. It is this fundamental connection between audience and actor that is key to the deployment of the actors developed visual sense of perception in the theatre. This primary and essential affordance not only allows the actor to enter the theatre environment ecologically but also to exist and thrive within it.

II

The actor walks into a space. It is a film studio. This is also an environment he understands. He can locate, behave and respond to his surroundings within this environment. However this space asks the actor to deploy his evolved sense of visual perception of performance environments in ways that are unlike his theatrical habitat. There are different forms of nesting at play here, and unlike the theatre; this environment is prone to constant rapid change and adaption (especially in terms of variable surfaces, substances and technology). There is a constant invariant the actor can rely on here; the camera. As long as the general environmental surfaces are not subject to drastic change and the camera remains invariant the actor is able to locate himself within this environment and work. The usual surfaces and mediums the actor uses to place and locate himself within the studio environment are constantly adapting here. Unlike the theatre, there is a varying degree of focus needed in this style of performance, which are all dependant on where the camera, the capturing device, is.

In the theatre the focus is always to the observer, the audience; it is mostly set as part of the initial navigation of the space and sits as a constant ecologically. In the film studio, however the observer is not constant, the observer is more substance than invariant and is movable within the environment. The visual perception of the actor requires constant shift as there is little invariance (beyond the camera) and a constant manipulation of objects taking place within the film studio. The actor must compensate for the large number of active participants in this habitat and for the focus required with the introduction of the intermediary possessing a particularly unnatural framed snapshot vision (the camera). The actor's position within the hierarchy that manages the massive numbers of people that are involved in film production needs to also be negotiated environmentally. This is especially vital ecologically, as the others that occupy this place are constantly locating and relocating themselves (along with a tremendous array of portable *techne* that they carry). As long however, as the invariant stays in place, the actor is able to work.

III

The actor walks into the space to work, though now this is not a *space* as he has formerly understood it, it is a volume (a volume marked out physically on the floor and virtually on a screen located within a motion capture studio). This is not an environment he understands or has been prepared for in his training and not a habitat he has navigated. The environmental volume is visually imperceptible in many ways. The actor will not be 'himself' here, cannot locate himself because he does not know who or where he is. How can the actor work? Can the notions of nesting be applied in this environment? Are there any markers for the actor to use as navigation points? Can this foreign place become a known environment? There is a capture device at play here, but what is the point of focus? Is there invariance at play, is there an observer? Can nesting and notions of scale be applied to technique? Is there a visual marker for the actor to view himself (a mirror?), can the actor rely on their own body as the invariant? How can this space be navigated, populated?



Fig 1: Image from chest mounted virtual camera in 3D space – Image M Delbridge (2012)

Works Cited

- Gibson, J (1979), *The Ecological Approach to Visual Perception*, Lawrence Erlbaum and Assoc, New Jersey
- Hickman, L (2001), *Philosophical Tools for Technological Culture*, Indiana University Press, Bloomington
- Stoll, A (1987), *Surfaces*, University of Minnesota Press, Minneapolis, USA