Multi-sensory Data Representation in Virtual Worlds: abstraction or pragmatism?

Ifan D H Shepherd¹, Iestyn D Bleasdale-Shepherd²
1. Middlesex University, London, UK
2. Valve Corporation, Seattle, USA

Abstract

Contemporary data visualisation and GIS software continue to make almost exclusive use of the visual sensory modality in representing data to the human analyst. The potential roles of other sensory modalities in virtual geographical environments (VGEs), and especially those designed to support interactive data exploration, are identified. A trans-sensory model of sensory variables is proposed, built on Bertin’s original graphical sign system, as a basis for the fusion of multiple sensory representations of data in multi-sensory data representation systems (MSDRSs). This is followed by a detailed evaluation of the (largely intuitive) adoption of sensory variables in contemporary videogame technology, in order to identify lessons which may be used to embed this model in future VGEs that aim to provide facilities for multi-sensory data representation.

Keywords:
Multi-sensory data representation, multi-sensory data visualization, multi-sensory data perceptualization, multi-sensory geo-visualization, multi-sensory GIS, multimodal data display, multimodal interfaces.

1. INTRODUCTION

“If the task of the display is to serve as the looking glass into the mathematical wonderland constructed in a computer memory, it should serve as many senses as possible.” (Sutherland, 1965)

“A standard GUI is a mirror that reflects back a severely misshapen human being with large hands, huge forefinger, one immense eye, and moderate sized ears. The rest of the body is simply the location of backaches, neck strain, and repetitive strain injuries.” (Rokeby, 1998, p.38)

If we take as our starting point the human sensory system, then the most fundamental feature characterising that system is the way in which the senses operate together in providing perceptual clues to the individual in pursuit of survival and fulfilment. How odd, then, that the overwhelming majority of currently available exploratory and analytical tools available for geographical analysis are unisensory, addressing primarily the human visual sense. It is over forty years since Sutherland (1965) proposed that computer graphics research look beyond “the picture in the window”, and attempt to create virtual environments in which observers could immerse themselves. It is a similar period of time since Helig designed working multi-sensory cinematic displays in the form of the Sensorama and the Telesphere Mask (Helig 1998). However, although considerable progress has been made since that vision was outlined to harness the power of multi-sensory pathways in many kinds of interactive software, virtual geographical environments (VGEs) remain, for the most part, almost exclusively preoccupied with generating visual signals for the eye-brain system.

This paper is organised as follows. First, a brief summary is provided of the main arguments in favour of multi-sensory data representation. The paper then proposes a trans-sensory model of multi-sensory variables as a basis for the effective fusion of multiple sensory outputs. Attention is then turned on a highly interactive type of virtual world available in the mass consumer market -- the videogame -- in which multi-sensory information is routinely displayed. A number of contemporary games are evaluated to determine whether they provide a suitable model of multi-sensory data representation that may be adopted by mainstream VGEs. The paper concludes with a challenge for mainstream GIS and data visualization vendors. Throughout the paper, applications software that provides multi-sensory display will be referred to as a multi-sensory data representation system (MSDRS).
2. WHY MULTI-SENSORY DATA REPRESENTATION?

The first reason for advocating multi-sensory data representation is that each of the human sensory modalities has particular strengths in receiving and interpreting information from the outside world. Eyesight is highly effective for taking in both wide views and for inspecting objects in high precision up close; it can detect motion in areas of peripheral vision; and it is able to discriminate colours and textures (Ware, 2004). Hearing is highly effective in recording non-verbal sounds that provide a sense of place; it signals background activities that may not be visible; it does not require user orientation or attentional focus; it is able to perceive highly transitory events and is sensitive to particular kinds of repetition and correlation that that may escape vision; it may be better at comprehending time-varying and multidimensional data (this makes it possible for auditory mappings of data to be used to reveal patterns and relationships that might not be visible in visual displays); and the binaural auditory system is capable of omnidirectional scanning of the environment (Bregman, 1994; Kramer, 1994; Cohen & Wenzel, 1995; Eldridge, 2006).

The tactile sense organ is far more spatially extensive than any other sense apparatus; it is highly sensitive to vibrations; and it can measure minute distances between stimuli. Finally, the kinesthetic sense has the unique property of both receiving and issuing force or stress, and is thus uniquely able to engage in active exploration of environments. Finally, each of the senses is in some way able to compensate for deficiencies in, or complete absence of, other senses.

In addition to individual sensory modalities having individual strengths that equip them for interpreting appropriately rendered data, additional advantages can accrue from combining two or more senses in a single data representation system. Table 2 summarises the main advantages of combining multiple senses; a more detailed discussion may be found in Shepherd (1994; 1995a). These broad advantages may be illustrated by the ability of the binaural auditory system to undertake omnidirectional scanning of the audio environment, which may be useful in several ways when undertaking spatial tasks (Cohen & Wenzel, 1995):

-- it assists when visual cues are limited
-- it can take over when the visual workload is high
-- it can undertake monitoring independently of the direction of gaze
-- it can help to direct the attention of the eyes towards particular locations that require more finely tuned visual investigation (i.e. cross-sensory guidance)

Many other interactions and synergies exist between most pairs of senses (Lewkowicz & Lickliter, 1994).

<table>
<thead>
<tr>
<th>Senses may be integrated for two distinct (but overlapping) purposes:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Sensory substitution or replacement</strong></td>
</tr>
<tr>
<td>This involves the replacement of one sensory modality by another, either on a regular or permanent basis, or during part of a work session. Sensory switching can serve several purposes:</td>
</tr>
<tr>
<td>• <em>Provides an alternative sensory source for impaired users</em></td>
</tr>
<tr>
<td>• <em>Circumvents poor operational conditions</em></td>
</tr>
<tr>
<td>e.g. strong ambient light; loud environmental noise; platform or personal motion (e.g. because of use of wearable or mobile computing facilities).</td>
</tr>
<tr>
<td><strong>Avoids over-use of a particular sense</strong></td>
</tr>
<tr>
<td><strong>2. Sensory combination</strong></td>
</tr>
<tr>
<td>This involves the manual or automatic combinations and substitutions of sensory modalities. Sensory combination has two primary uses:</td>
</tr>
<tr>
<td>• <em>For reinforcement or redundancy</em></td>
</tr>
<tr>
<td>By representing a data variable through more than one sense the effect is reinforced, either to underline the importance of the variable, or because the effect of single-sense encoding is relatively weak</td>
</tr>
</tbody>
</table>
For enhanced data representation

- by distributing the representation of data variables across multiple senses, it may be possible to encode a greater number of variables than using a single sense.
- by distributing the representation of data variables across multiple senses, it may be possible to reduce the cognitive overload on a sense that previously had too much information channelled to it.

Table 1. Benefits of representing information to multiple sensory modalities.

Several strategies suggest themselves for the use of MSDRSs by analysts. One strategy is the simultaneous use of two or more sensory modalities. As suggested in Table 2, this may involve the combination of sensory representations that improve perception through redundant or enhanced multisensory coding of data. A second strategy is interleaved use of multiple sensory modalities. In this approach, the user’s exploration strategy might involve rapid switching between one sensory modality and another, or between one combination of senses and another, in pursuit of a particular navigation, interaction or interpretation task. A third strategy involves sequential switching between sensory modalities. For example, a user might undertake one phase of investigation favouring one sense, or combination of senses, before moving on to a subsequent phase that favours a different sense, or combination of senses. The important point here is not that one strategy is necessarily better than the others, in some absolute sense, but that analysts are able to choose between various sensory approaches to the exploration of their data.

3. TOWARDS A TRANS-SENSORY MODEL OF SENSORY VARIABLES

A number of essential components are required to construct an effective MSDRS. These include: output devices for each sensory modality; software drivers to enable these to work with major operating systems; standard software libraries for each of the main senses (i.e. ‘engines’ or middleware); and high-level application software. The application software, at a minimum, must implement a model of multi-sensory data with rules as to how such data are to be rendered for effective human interpretation. The model we propose as a basis for this functionality is a trans-sensory extension of Bertin’s graphical sign system (Bertin, 1973; 1978; 1979; 1981; 1983), which had at its heart a system of visual variables. By encoding different kinds of data in appropriate visual variables, various graphical products (maps, diagrams, etc.) are able to express information in a principled manner.

Bertin initially described seven such variables (position, size, shape, value, orientation, colour and texture), although this set has subsequently been expanded, largely as a response to the development of advanced and affordable computer graphic technology since the time that Bertin identified his variables. Additional visual variables include: the third positional variable of height (i.e. the z-axis), saturation, and focus or blur (MacEachren, 2001). Additional dynamic visual variables have also been suggested (MacEachren, 1994), and a scheme for unifying several types of time-varying symbols within Bertin’s approach has also been proposed (Shepherd, 1995b). Table 2 summarises the major sensory variables that may be simulated digitally to provide feedback to the major human senses.

Bertin’s second major contribution was to use perceptual characteristics of the visual variables (e.g. their selective, associative and quantitative roles) to link them to what are nowadays referred to in statistical terms as the levels of measurement of the data being visualized (i.e. categorical or nominal scale, ordered or ordinal scale, interval scale, and ratio scale). The linking of data variables to visual variables by matching their fundamental characteristics to construct appropriate and effective data representations lies at the heart of modern data visualization. This model has also been exploited in a number of subsequent attempts to automate the data visualization process (e.g. Mackinlay, 1986, 1987; Roth & Mattis, 1990; Casner, 1991; Murray, 1994; Senay & Ignatius, 1994; Jung, 1995; Zhan & Buttenfield, 1995), which is another reason for placing it at the heart of attempts at multi-sensory data representation.
<table>
<thead>
<tr>
<th>Sight/Vision</th>
<th>Hearing/Audition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal position</td>
<td>Location</td>
</tr>
<tr>
<td>Vertical position</td>
<td>Direction (azimuth, elevation)</td>
</tr>
<tr>
<td>Depth position</td>
<td>Distance/range</td>
</tr>
<tr>
<td>Shape</td>
<td>Pitch/frequency</td>
</tr>
<tr>
<td>Size</td>
<td>Intensity/volume/loudness/amplitude</td>
</tr>
<tr>
<td>Length</td>
<td>Timbre</td>
</tr>
<tr>
<td>Area</td>
<td>Damping</td>
</tr>
<tr>
<td>Volume</td>
<td>Duration</td>
</tr>
<tr>
<td>Angle</td>
<td>Repetition rate</td>
</tr>
<tr>
<td>Curvature</td>
<td>Tempo</td>
</tr>
<tr>
<td>Colour</td>
<td>Rhythm</td>
</tr>
<tr>
<td>hue</td>
<td>Rest</td>
</tr>
<tr>
<td>value</td>
<td>Attack rate</td>
</tr>
<tr>
<td>saturation</td>
<td>Decay rate</td>
</tr>
<tr>
<td>Orientation</td>
<td>Phase</td>
</tr>
<tr>
<td>Texture</td>
<td>Envelope (volume shape)</td>
</tr>
<tr>
<td>Pattern</td>
<td>Attack</td>
</tr>
<tr>
<td>Orientation</td>
<td>Decay</td>
</tr>
<tr>
<td>Roughness</td>
<td>Sustain</td>
</tr>
<tr>
<td>Transparency</td>
<td>Release</td>
</tr>
<tr>
<td>Focus/blur</td>
<td>Wave shape</td>
</tr>
<tr>
<td>Reflectivity</td>
<td>Ambience</td>
</tr>
<tr>
<td>Time-varying parameters</td>
<td>Reverberation</td>
</tr>
<tr>
<td>Blinking</td>
<td>Resonance</td>
</tr>
<tr>
<td>Movement</td>
<td>Presence</td>
</tr>
<tr>
<td>Change</td>
<td>Spaciousness</td>
</tr>
<tr>
<td></td>
<td>Harmony</td>
</tr>
<tr>
<td>Touch</td>
<td></td>
</tr>
<tr>
<td>Position</td>
<td>Kinesthesis</td>
</tr>
<tr>
<td>Size</td>
<td>Position</td>
</tr>
<tr>
<td>Shape</td>
<td>Movement</td>
</tr>
<tr>
<td>Elevation</td>
<td>Effort, tension/force</td>
</tr>
<tr>
<td>Intensity/sharpness</td>
<td></td>
</tr>
<tr>
<td>Texture/roughness</td>
<td>Vestibular/proprioception</td>
</tr>
<tr>
<td>Frequency</td>
<td>Direction</td>
</tr>
<tr>
<td>Outline</td>
<td>Orientation</td>
</tr>
<tr>
<td>Density</td>
<td>Acceleration</td>
</tr>
<tr>
<td>Interval</td>
<td>Rotation</td>
</tr>
<tr>
<td>Orientation</td>
<td></td>
</tr>
<tr>
<td>Vibration</td>
<td>Smell/Olfaction</td>
</tr>
<tr>
<td>Locus</td>
<td>Fragrance</td>
</tr>
<tr>
<td>Frequency</td>
<td>Concentration</td>
</tr>
<tr>
<td>Amplitude</td>
<td>Essence</td>
</tr>
<tr>
<td>Duration</td>
<td></td>
</tr>
<tr>
<td>Temperature</td>
<td>Taste/Gustation</td>
</tr>
<tr>
<td>Pain</td>
<td>Sweetness</td>
</tr>
<tr>
<td>Pleasure</td>
<td>Saltiness</td>
</tr>
<tr>
<td></td>
<td>Bitterness</td>
</tr>
<tr>
<td></td>
<td>Sourness</td>
</tr>
</tbody>
</table>

Table 2. Sensory variables available for data representation [1].

It should be noted that a trans-sensory model of sensory variables is a necessary, but not sufficient, requirement for building an effective MSDRS. When sensory outputs are combined ready for output, there are several fundamental differences between the sensory signals that pose significant problems. Two examples illustrate this. First, the difference between static visual images (as in maps) and inherently time-varying audio output may lead to problems of synchronisation. Second, the different
frequency rates at which humans process visual, vestibular and haptic signals can lead to lags in the various sensory information streams that may lead to virtual motion sickness. (These issues are discussed in standard texts on virtual reality and virtual environments, such as Durlach & Mavor (1995) and Stanney (2002).) There is, therefore, still some work needed to construct a fully fledged trans-sensory sign system.

4. VIDEOGAMES AS MSDRSS

In this section, we examine the extent to which videogames express the rule-based approach to multi-sensory data representation outlined in the previous section.

4.1 Videogames and VGEs compared

There are several reasons for looking to videogames for guidance on best practice in the design of MSDRSs for geographical analysis. The obvious reason is that both videogames and VGEs construct highly complex and often extremely realistic models of real and synthetic worlds. A second reason is that, unlike VGEs, videogame technology has been a significant experimental laboratory for new data representation technologies over many years. Hundreds of new game titles are released annually, and these are acquired by generally young consumers with a thirst for the latest gadgets. A third reason is that most modern videogames combine elements of the visual, the aural, the tactile and the kinesthetic in their highly interactive virtual environments.

Before analysing what may be learnt about the technology of multi-sensory data representation from videogames, it is relevant to consider several significant contrasts between videogame and VGE technology, and in the role of the various forms of sensory information these two families of software output to their users. Perhaps the most notable difference is that while videogames are primarily a means of entertainment, providing players with challenges and sensory stimulations to deliver adrenaline-coated fun, VGEs are primarily a tool for evidence-based enlightenment, providing tools that help users answer burning questions and solve significant problems. In videogames, various sensory stimuli are orchestrated to deliver experiences, while in VGEs they are (or could be) coordinated to reveal information. (Gibson’s (1968), p.47) suggestion that experiences are acquired through channels of sensation while percepts are obtained through perceptual systems need not detain us here.) This distinction has a direct bearing on which of the senses are engaged in games and VGEs, and on the kinds of information that might be output to each sensory modality within these two families of software.

A second difference is that most videogames are designed to induce a highly immersive experience, within the constraints of desktop technology. Although many game players, like cinema-goers, have their sense of immersion enhanced by switching off the lights during gameplay, the games themselves also adopt various tricks to encourage the suspension of disbelief and immersion in the virtual world. One such trick is the adoption of a first-person viewpoint in many action games. A second trick is the minimisation of visual and auditory elements which distract from those directly representing the main virtual environment. (For a fuller discussion, see Shepherd & Bleasdale-Shepherd, 2008.) A third trick is careful synchronization of the visual and auditory outputs with each other and to the world’s behavior/simulation – e.g. visuals and audio tied closely to creature behaviour and physics simulation. Games also mimic the effects of the environment upon the player, such as by jostling the viewpoint and emitting large sounds when something hits the player.

In many videogames, the content of each sensory output stream is tailored to sustain the sense of immersion in the gameplay. Images and sounds are of primary importance in achieving immersion, as they are in cinema and TV drama, with images and sounds in consort being designed to elicit strong emotional responses. Haptic stimuli play a relatively limited role in eliciting immersion videogames, partly because of the lack of availability (or cost) of specialist peripherals, but also because haptically displayed information does not always serve the experiential goal of immersion. (The realism of even run-of-the-mill visual and auditory output far exceeds any such available haptic analogues.) In VGEs, by contrast, various combinations of images, sounds and haptic stimuli are devised to sustain the analyst’s search for answers to questions, and a variety of ancillary information (menus, information boxes, status data, etc.) commonly appears on screen to assist the process of exploring information spaces.
It is not the authors’ intention to suggest that immersion is not of value in information exploration environments, nor that information transmission is unimportant in videogame environments. (It will be shown below how videogames encode information for delivery to various sensory modalities.) Nevertheless, a principal aim in the design of videogames is to enhance the total user experience through immersion, while a principal aim in the design of VGEs is to enhance the analytical ability of the user through the output of maximally relevant information. Videogames display information in ways that support their entertainment goal, while VGEs display information in ways that support their analytical goal. It is probably fair to suggest that while games address both the human affective and cognitive systems, with the latter supporting the former, VGEs primarily address the human cognitive system, with only a minimal role for the affective system. It should be noted that most existing research into affective computing (Picard, 1997) involves the role of sensory systems in the interface rather than in the representation of information and data.

A third contrast between videogames and VGEs is one suggested by Joiner (1998), who proposed a continuum from interactivity to storytelling (or narrative) in classifying videogame software. Examples of contemporary games may be found at most points along this continuum, and many games combine storytelling and interactivity in often highly creative creative ways (Clanton, 2000). In contrast, VGEs are self-evidently positioned at the interactivity end of the spectrum. Admittedly, a number of researchers (e.g. Monmonier, 1994) have suggested that narratives may be introduced into dynamic cartographic systems by means of guided tours and other scripted facilities. However, such systems are primarily aimed at the communicational, expository or educational roles of VGEs, rather than for their exploratory or analytical roles.

### 4.2 Multi-sensory data representation in videogames

Perhaps without really being aware of it, videogame designers are trailblazers in normalising the multi-sensory representation of information in the context of virtual environments (VEs). The sensory representation of information may be found in many videogames that involve players in exploring 3D VEs. Such games belong to several genres, including: first-person shooter (FPS) games, third-person platformer games, and real-time strategy (RTS) games. Our analysis of a number of such games reveals several ways in which designers have achieved effective multi-sensory representation of information. We focus particularly on four sensory modalities: the visual, the aural, the tactile and the kinesthetic, which involve the sensations of images, sounds, feeling and force. Because the main concern of this paper is with how information is conveyed to software users by means of various sensory representations, our analysis of videogames largely ignores sensory information (such as music) which typically serves other roles. Although computer-generated music is frequently used in interface icons in general applications software, much of it in a videogame context has a strongly emotive content, which lies outside the scope of the current analysis. (It should be noted, however, as discussed below (see Table 4), that many games use music to convey categorical information.) Our focus is on what may be called environmental sounds, which are constructed to display information about various aspects of the game’s virtual environment, including its characters, inanimate elements and events. Other uses of music, such as to provide changes in mood, are not considered further. Our remaining discussion focuses on how videogames implement what might we have called a trans-sensory sign system.

#### 4.2.1 The display of sensory feedback in videogames

Our analysis of videogames therefore begins with data (or, more broadly, with information), and with ways in which two broad types of data (categories and measurements) are converted into the four sensory outputs identified above. Categorical data may be sensorised to signal or identify individual members of a category of things. For example, a unique visual or aural signature may be used to identify different types of monster, or to alert the player to different types of event (e.g. varied sounds to signal the footsteps of different characters). Measurement data may be used to provide ordered or quantitative information about various elements in the game. These may range from the relative strengths of opponents, the current level of health of the player, the amount of ammunition remaining in a weapon, the degree of proximity of an adversary or target object, or the degree of danger in the current location or situation. Tables 3 to 6 provide examples of the various ways in which these two broad types of information are represented to each of the four main sensory modalities in videogames.
Categorical data
In most videogames, the visual design of entities is carefully undertaken so as to allow at-a-glance disambiguation between the many entity types present. Two commonly used techniques may be illustrated with the team-based, multi-player game Team Fortress 2, in which the player may take on 9 roles (scout, spy, medic, heavy (weapons), soldier, sniper, pyro (maniace), demo (lition) man, or engineer). Because the capabilities and behaviours of players varies greatly between each of these roles, it is important for the player to be able instantly to identify another player’s team and role (Mitchell et al., 2007). Two kinds of visual coding are used to ensure this:

• The avatar for each role has a unique silhouette, easily distinguishable from all the others.
• Players on opposing teams are differentiated through the use of simple colour schemes - e.g. one team wears red clothing, the other blue. The use of color-coding to relate groups of entities is a very common practice in videogame design.

A third kind of visual coding technique widely used in games is associated with particular actions or events:

• Individual actions or events are associated with a unique visual signature, often created using ‘special effects’ based on particle systems. This is important, because it provides clear feedback to the player as to when they have (or have not) successfully completed a specific action (e.g. blowing up a vehicle).

Measurement data
• Analogously to 3D audio spatialisation, games which use perspective projection provide implicit scalar visual feedback - relating an object's size on-screen to its distance from the viewpoint (as well as the obvious relation between on-screen position and virtual 3D location)

• Location and size (size is often related to strength of enemies or value of items)

Table 3. Examples of visual feedback.

In many videogames, some forms of sensory information are often used on their own to provide relevant information to the game player. Visually coded information is perhaps most frequently used in this way. In Portal, for example, specific status information about health is conveyed by using object colours to vary with the level or state (active/inactive) of the relevant variable. Other games apply similar coding to indicate remaining ammunition, energy levels, and ‘magic’. Sound, too, is sometimes used on its own to encode information, as for example in the countless games which use 3D audio spatialisation to encode the proximity and direction of objects (seen or unseen) in relation to the viewpoint. Tactile sensation, in the form of rumble, may sometimes be provided as the sole signal alerting players to certain game states, for example to convey the player’s proximity to an unseen and unheard monster. This is uncommon in multi-platform games, however, which tend to include only those facilities likely to be available across all platforms on which they are released. (This is because not all game controllers have rumble.) Finally, force-feedback may sometimes be used to indicate a game’s internal physical simulation which may not be visually apparent. For example, force-feedback steering wheels may encode a vehicle’s friction on the driving surface through forces opposing the player’s steering action. In the final console game F-Zero GX, for example, a force-feedback steering wheel means that when an enemy racer slam into the player’s hovercraft, the wheel will jerk appropriately, in addition to pulling against the player on tight turns.
Categorical data
It is very common in videogames for classes of entities to be identified by means of audio 'tags'. In Half-Life 2, for example, each type of enemy has a distinctive aural signature, and will emit easily recognised sounds when distinctive events occur:

-- idling (the enemy is not yet aware of the player)
-- awakening (as the enemy notices the player)
-- attacking (as the enemy prepares an attack on the player)
-- death (as the enemy dies)

Similar audio tags are used to identify types of item (e.g. medical supplies or ammunition), providing aural confirmation as the player collects them.

• Music is often used to indicate different states of play, including the presence of danger and a lack of remaining time, as well as to provide notification of certain events, including the completion of a major goal or discovery of a secret area.

Measurement data
• 3D spatialisation of audio implicitly provides scalar feedback by changing audio volume on the basis of a sound source’s distance from the player or viewpoint.

• Another common audio feedback mechanism is to increase the pitch of a sound, or rate of a periodic sound, to give timing information to the player. For example, there are many puzzles in the Zelda series of games involving timing. A common conceit is for a switch, once thrown, to emit a ticking sound which increases in rate and pitch as the player's remaining time to complete the puzzle runs out.

Table 4. Examples of audio feedback.

Although videogames output data from both categorical and measurement variables in most of the sensory modalities considered in Tables 3 to 6, it is equally clear that displays in most of the sensory modalities are is restricted to a relatively small number of the sensory variables summarised in Table 2. Thus, for example, of the visual variables, orientation and focus (or blur) are rarely if ever used; in terms of audio variables, phase and reverberation tend not to be used; of the tactile (or tactual) variables, few are used apart from a general rumble; and of the kinesthetic variables, only force is reflected. Although the absence of several of these sensory variables is a result of designer choice, in many games it is primarily due to the limitations of the hardware devices available to display the required sensory signals. The absence of direction-feedback in the kinesthetic domain is largely because most game platforms provide only force-feedback devices, direction-feedback devices being largely the domain of more specialised VR applications. Force-feedback steering wheels do provide this, though it is only a 1D directional (i.e. left/right) cue. Reverberation is an interesting case because, although it is common in the sound technology available to game players, designers would not think to use it to convey directional information because it is so imprecise. It would, however, be sufficient to encode binary categorical data.

Categorical data
Controller 'rumble' is often used to identify events occurring in the virtual environment, with the 'profile' of a rumble (that is, how its strength varies over time) being used to differentiate between different events. For example, in the game Jak & Daxter, rumble is used to identify moments at which the player avatar lands on the ground, when they fire a weapon or when they take damage from an enemy attack.

Controller rumble is also often used to generate a sensation which implies that the player's avatar is in contact with certain surfaces or objects in the virtual environment. In Ratchet & Clank, for example, the player character can don magnetic boots in order to walk along metal tracks that loop around at arbitrary orientations with respect to gravity. Whilst walking on such surfaces, controller rumble provides the sensation of the heavy, reverberating impact of the player's boots against the metal track.
In *Super Mario Galaxy*, the Wii controller rumbles when the player’s avatar lands from a great height.

**Measurement data**
- Controller rumble is sometimes used to convey scalar feedback independently of other sensory channels. Typically amplitude or pulse rate is used to convey proximity to otherwise secret (e.g. underground) items within a game.

**Table 5. Examples of tactile feedback.**

**Categorical data**
Force-feedback devices are not normally used to reflect categorical information.

**Measurement data**
- Force-feedback joysticks have long been available for flight simulators. Feedback forces are derived from physical simulation of the virtual aircraft, in an attempt to mirror the forces felt through a joystick in real aircraft.

- Force-feedback steering wheels are widely available for driving games. Forces are used to convey the resistance of the car's wheels to turning based on physical simulation of the vehicle in contact with the road.

**Table 6. Examples of kinesthetic feedback.**

### 4.2.2 Multi-sensory feedback in videogames

Our analysis of videogames now turns to focus on how two or more sensory outputs are combined. The first set of questions concerns the frequency with which *sensory combinations* occur in videogames, which senses are most likely to be combined, and how often are they combined during the course of a game. The evidence suggests that multiple sensory combinations are extremely common, with combinations of the visual and audio being the most frequently used, and combinations of the visual, audio and tactile coming not too far behind. It should be noted, however, that these combinations, and the frequency with which they occur, vary considerably between games and during the course of playing particular games.

Some examples of multi-sensory interaction include:

- The use of audio tagging of significant game events or changes of state (as indicated in Table 4) avoids introducing unnecessary visual clutter in the game's head-up display (HUD), thus improving immersion.

- In *Zelda: Twilight Princess*, the Wii controller combines both sound and tactile feedback (by emitting a 'twang' sound and rumbling) when the player fires a bow and arrow.

- Categorical tactile feedback (e.g. in the form of rumble) is not normally given independently of information from other sensory channels. It is usually used to reinforce audio or visual cues, to aid the sense of immersion in the game's virtual environment, and to increase the sense of connection between the player and their avatar in the game.

- Visual cues and audio volume are often used together in videogames to help distance the camera from the player avatar, or conversely to make them feel more tightly bound to them.

- Audio cues can help players focus on a critical visual element, for example, by teaching the player to associate a distinctive audio cue with a given enemy action (and its related animation, such as raising an arm high prior to striking the player). This audio cue can then be used to help the player quickly locate an attacking enemy from among many of the visualized enemies on screen (i.e. by quickly searching for a raised arm when they hear the audio cue). This is useful in a game where
the use of obvious visual attention-drawing (such as making the enemy glow) would be out of keeping with the realistic visual style, and would thus hurt immersion.

An interesting case of the decoupling of visual and audio information can be found in third-person games, where the player controls an avatar. Audio effects are typically spatialised as if sensed from the viewpoint (i.e. as if a virtual microphone was positioned there). They are not located at the viewpoint, but are located wherever it makes sense -- e.g. for a character coughing, it will be positioned at their mouth. This can sometimes lead to a sense of emotional detachment of the player from their in-game avatar, as may be witnessed in games such as *Ico*.

A key research question is whether the provision of additional sensory information provides improved environmental intelligence to the player, helping them to undertake relevant tasks more effectively. It is common for surface materials in a virtual environment to impact on gameplay, and in these situations varying footstep sounds provides surface type information. Many games, such as *Half Life 2*, also modify sounds using DSP (digital signal processing) to produce effects such as echo or underwater damping, which serve to reinforce more obvious visual cues.

In terms of the *sensory combination* strategies used in videogames, it is typical to find that categorical representations are exclusively combined with other categorical representations, and measurement representations exclusively combined with measurement representations. However, because some game entities or events combine attributes at different levels of measurement, multi-sensory representations of a categorical attribute may occur alongside multi-sensory representations of a measured attribute. An example might be where the type of monster is encoded using categorised visual and aural cues, while its level of health may at the same time be encoded using a combination of scaled visual and aural cues. There are, of course, particular gameplay situations in which sensory combinations have greater relevance and utility. For example, sensory combination is of particular help to players in situations where one sense is impaired. Thus, if an event is associated with both audio and visual cues, then it will still be recognisable even in a dark area (where vision is impaired), or at a distance or behind thick glass (where hearing is impaired), or even if hearing is temporarily reduced, as happens in *Half Life 2* when an explosion temporarily deafens the player.

This brings us to a consideration of *sensory substitution*, which refers to situations in which one sensory output is replaced by another. For example, in some games when the lights go out or darkness falls, audio outputs may be brought into play to compensate for the absence of visual information. It is common practice for modern videogames to teach players to associate a given sound with a given image (such as an enemy or item), so that when one sensory representation is displayed on its own, its equivalence to the other is automatically assumed. However, there are probably few games in which one sensory representation is deliberately switched to another by the software, even to compensate for a change in game circumstances. In most games, this is unnecessary, because particular sensory outputs are rarely lost altogether. (When a player’s avatar dives into water, for example, the audio output is usually modified rather than switched off entirely).

However, many contemporary games have responded to the possible lack of a particular sensory ability on the part of the game player, commonly by providing closed-captioning as part of a mature system. This means, for example, that sound effects are voiced on screen through text subtitles for the hard of hearing or deaf player, and certain visual effects (e.g. brightness, contrast and font size) may be adjusted by the player to compensate for their poor vision. Some forms of sensory impairment are considered at an even more fundamental design level in those games which have been created or adapted especially for certain groups of impaired player (e.g. Spoonbill, 2007). (In addition, the games company Deaf Gamers (www.deafgamers.com) reviews games software from a deaf person’s perspective, indicating the importance of visual and textual feedback to substitute for lack of hearing, Audyssey (Audyssey, 2007) provides a discussion list for blind gamers, and AudioGames.net maintains an extensive database of games based on sound, and blind-accessible games.)

Where one sensory output may be switched to another, it is unlikely for the replacement to be brought into play automatically by the software itself. It is more typical for the switch to be invoked by the player when they consider that the circumstances of the gameplay make it appropriate. For example, in dark environments, players of *Metroid Prime* and *Splinter Cell* can switch visual modes (e.g. to night-vision or thermal vision), but there are few, if any, examples of games where players can switch between complementary aural modes, or are able to switch between one sensory output and another.
Some games do switch off certain sensory clues (e.g. turning off the lights, obscuring the view, or making solid objects invisible), sometimes unexpectedly, but usually as a form of challenge for the player. This game behaviour is taken to extremes in Eternal Darkness, which takes numerous sensory actions (e.g. turning off audio-visual information, unplugging the game controller, transforming the player into a zombie, displaying cockroaches crawling across the screen, replaying the previous scene...), to suggest to the player that they are losing their mind.

One particularly important category of visual output in videogames and VGEs alike is that which specifies locational information. In the virtual environments of most videogames, however, a great deal of visual information is not inherently spatial. (In the Portal game, for example, colour is part of the visual coding that helps players identify specific characters, but it does not, of itself, convey much spatial information.) As is well known from the use of the third positional dimension in geovisualization (Shepherd, 2007), the spatial cues delivered by dat visualization software may be deficient in various ways. For example, symbols may be occluded, distances may be distorted by perspective, objects displayed in the vertical axis may conflict with one another, and spatial information in the form of lighting may interfere with information in the form of surface colouration. A previous generation of videogame (including Sim City, Railroad Tycoon and Populous) used isometric projections of the game world to reduce some of these problems (such as occlusion and perspective distortion of distance), though this was largely an unintended benefit of the then immature development of fully 3D game engines. More modern 3D videogames attempt to compensate for some of these deficiencies in other, often clever, ways. In some videogames, for example, players can apply a form of transparency to the game environment to temporarily reduce object occlusion, either through x-ray vision, or by using a thermal vision mode (as in Metroid Prime). Some games (such as Viva Pinata) also automatically use transparency to avoid undesirable occlusion caused by foreground objects. The temporary collapsing of the dimensionality of the display from 3D to 2D in Crash however, works in the opposite direction. Rather than helping to make the player’s positioning more precise by removing perspective and occlusion, this particular game mechanic creates puzzles by altering the configuration of the environment (when viewed from different angles), and the player’s ability to traverse it.

When it comes to the other three sensory modalities, however, the amount of location-specific information provided is typically far less. Most videogames are able to exploit stereophonic sound to provide some indication of direction, particularly to the left and right of the player. However, this is the lowest level at which directional information may be provided through audio output. With the advent of surround sound (or 5.1 stereo), sound can be made to appear to come from the rear of the player, though the quality of the positional audio information depends significantly on the quality and positioning of the player’s speakers. (Headphones often provide the best listening quality.) An example of a game which uses some form of spatialised sound to provide locational information to players is The Legend of Zelda: The Wind Waker, which uses it to reveal the location of secret underwater treasure. Most combat games now provide effective spatialised sound (stereo or surround) to indicate the direction of gunfire, and game reviews are increasingly disparaging of new titles that do not support these forms of spatialised sound.

While stereo sound merely provides information about left and right in front of the player, and surround sound does not adequately convey information about the location of objects and events above, below or behind the player, binaural sound (Foster & Schneider, 1996) or positional 3D sound (Microsoft) render full 3D sound information. However, even binaural sound is unable to provide absolute sound positioning information. Fortunately, in most games it is sufficient for players to receive information about the relative distance that objects or events are located from them. In Half Life 2 and Zelda Wind Waker, for example, the locational information is sufficiently rich to provide enough cues for the player to act appropriately (e.g. moving away from danger or moving towards a source of treasure).

There are two broad approaches to providing sound within videogames and VGEs: through several emitters spread throughout the scene, or through a single emitter at a single location. In the former case, if the hardware and software can synthesise multi-channel sound, then the player will perceive a multidirectional soundscape; in the latter case, the player will hear just a single, highly localised sound. In most 3D games, sounds tend to come from emitters localised in 3D space, not from a single emitter. (Of course, if the scene isn't really very 3D, or if the viewpoint doesn't tend to move around (as in a 3D puzzle game), then no spatialisation is required.) There are also two ways of hearing sounds. In most games, the sounds are generated more or less continuously by the software, and are played through the
speakers without the player having to do anything about them. The second method is for the software user to wield a ‘sonic probe’ (as in several VR systems), so that sounds are heard from the object they touch in the scene. In most games, however, audio is almost always tied to the viewpoint, not to a ‘sonic probe’ or ‘microphone’. One of the few examples of the latter is provided by *Viva Pinata* (a type of gardening game), in which the player’s shovel acts as the cursor/avatar, and it sometimes emits a beeping sound which gets louder based on how close the shovel (not the viewpoint) is from buried treasure.

Tactile sensations, which are displayed almost exclusively in the form of vibrations applied to the hand or fingers (usually known as ‘rumble’), are currently able to provide minimal locational information. For example, rumble is not normally differentially applied to the left and right joysticks on a game console controller to indicate positional information. The provision of locational information is more commonly delivered through the other haptic sense of kinesthesia. However, most examples of force-feedback (or force-reflection) in videogames are of a simple variety, involving the application of force to the muscles, tendons and joints of the hand. However, as with rumble, there are no examples of the use of differential force-feedback to the left and right joysticks of a game controller to indicate direction or location of force. Few, if any, games (with the exception of arcade games) involve the application of forces to other parts of the body.

Most of the previously given examples indicate successful approaches to displaying information through individual or multiple sensory modalities. However, there are some cases where a particular sensory display is not entirely successful. A common occurrence is that the virtual microphone in third-person games tends to be at the viewpoint and not the avatar. One drawback of this arrangement occurs when audio cues are used in cases where proximity to the avatar (not the viewpoint) really matters, such as proximity to danger (e.g. nearness of a projectile or enemy to the avatar), or proximity to a discovery (e.g. a secret item which emits a tinkling sound when the player gets close to it). The point here is to discover this with the avatar, not from the viewpoint, and some games attempt to solve this problem by differentiating these proximity alerting sounds from others by using a second virtual microphone at the avatar to capture them.

Another classic problem (which most good games address) is when the player character dies but the player does not know why they died. The standard response is for enemies to emit a ‘preparing to attack’ sound in the moments before they attack, so that the player knows that something is coming even if the source is off screen. A rather frequent blunder is building up visual/aural conventions (e.g. dangerous things are red, or make high-pitched sounds) and then breaking the convention and hence confounding the player’s expectations. For example, the player may be expected to pick up a red/high-pitched object, but they don’t want to, so they get stuck not knowing how to proceed even though the poorly-designed answer is right in front of their faces.

We are now able to draw together some of the results of these analyses, in order to indicate the purposes for which sensory modalities are most useful when used individually, and the purposes for which they are most useful when used in combination. A continuing problem, however, is that individual videogames do not provide a comprehensive range of examples, nor do they do so in ways that enable VGE analysts to undertake desk experiments with the games to determine independent answers to this question. The current uses of multi-sensory output in videogames are still at an early stage, so that the technology is currently unable to provide a definitive answer to this question. Videogames can make at least three contributions to VGEs. Firstly, they demonstrate that a reasonable degree of immersion can, if and when required, be achieved on a desktop platform, without resorting to the often expensive, and usually tiresome, head-mounted displays so emblematic of contemporary VR. Secondly, videogames clearly demonstrate that multi-sensory data representation is possible using relatively standard computer technology. Thirdly, videogames indicate the emergence of a trans-sensory sign system, albeit one that is informally expressed and operationalised. We develop this last point further in the next section.

### 4.3 Where videogames don’t lead

Despite the major lessons on multi-sensory data representation provided by videogames, the architecture of multi-sensory display adopted by most videogames is not necessarily a model to be followed by VGEs. Although videogame designers may be considered to have intuitively adopted a simplified version of the trans-sensory sign system, this has not been formalised, and has certainly not
been used to develop an abstraction layer for multi-sensory display. Some of the reasons for not doing so are briefly outlined below.

4.3.1 **Pragmatics rule OK**

Game developers tend to do only as much as is needed to get the product out of the door. Although most videogame designers are passionate advocates of ‘clean’ code and maximum clarity in creating software architectures, they would argue that there is little need for a multi-sensory abstraction layer in their products. Their advocacy of a pragmatic approach to multi-sensory integration, especially in the context of videogames, is based on three main principles or requirements. The first is that game software developers need considerable flexibility in representing their virtual worlds, and this might be compromised by the adoption of a high-level abstraction layer, with its insistence on a rule-based approach. The second is that an integrated multi-sensory model is still insufficiently mature to warrant adoption for game development. The third is that current game platforms and interface devices are so diverse as to prevent the implementation of such a layer. A final consideration is one of cost: games developers rarely allocate resources to developing facilities that might be useful some time in the future.

4.3.2 **Middleware: the shackle of software libraries**

The development of most modern videogames relies heavily on the availability of standard specialist software libraries. Indeed, most modern videogames integrate a number of software ‘engines’ in their code, including a graphics engine, an audio engine, an AI engine, a physics engine, and a scripting engine. The modularity of these engines would seem to provide an excellent model for the development of a multi-sensory abstraction layer, in which separate engines would be responsible for the rendering of information for specific sensory modalities. However, modularity alone is insufficient to create a common abstraction layer across all sensory modalities. For this to occur, two other conditions have to be met. First, the individual sensory engines have to be built (at the very least) on the basis of an established model of sensory variables. Although a detailed analysis of current rendering engines is beyond the scope of this paper, it is clear that even in the most mature sensory engine technologies (e.g. graphics pipeline libraries such as OpenGL and DirectX), there is currently no native support for an extended version of the graphical sign system, though OpenHaptics does provide cross-platform support for basic haptic output. Even if there were such support, and it was coordinated between different developer communities (graphics cards and audio cards, for example, are developed by entirely different businesses), the second requirement consists of the resources and effort required to integrate such a model or engine into the software design process. Recent evidence from one major games producer suggests that considerable effort, over a long period, may be required to integrate even a mature physics engine into the game design process (Stelly, 2006). The third requirement is discussed in the next sub-section.

4.3.3 **Information sources and the role of design**

In videogames, output (be it visual, audio, tactile or kinesthetic) is primarily design based, in that the data to be represented (e.g. for an object or event) is created by the game designer to serve a particular purpose (e.g. excitement, challenge, fun) in the context of the evolving gameplay. In contrast, in VGEs, and especially in data visualization software, output (for whatever sensory modality) is essentially rule driven (or should be), and is based on pre-existing information. Indeed, the conversion of the data to the rendered output is automatable, as illustrated by the experimental systems devised, amongst others, by Mackinlay (1986, 1987). In other words, a videogame world involves designers in hand-crafting the sensory products, whereas a VGE will typically have its sensory products created by the software from a database. Given that specialist designers (especially visual and audio designers) tend to work on individual output streams (creating texture images and audio files, respectively), it is perhaps unsurprising that there is little integration of the production of the output streams within the software architecture of the game.

Although the work undertaken by the specialist design teams comes together in the final videogame, their development work is often undertaken separately. In some game companies, the various sensory experiences are integrated by a lead designer. In other companies, the development process is highly focused on user experience, which is continually checked through playtests. These help to ensure that
the multiple forms of sensory output are tied tightly together, where valuable to the play experience, from an early stage. In terms of the technologies that each team uses to create their own sensory displays, the code tends to be entirely separate, though they are usually tied together through an entity-management system and in the content authoring tools. (For example, designers associate visual and aural representations with entities using the available design tools. This is fundamentally a technical matter of associating filenames with objects, for each type of sensory representation, then adding a myriad flags and settings relating to each type of representation.)

These two approaches to the bringing together of sensory outputs during the production of individual videogames suggests that it might not be impossible for a multi-sensory abstraction layer to be developed. Neither precludes the adoption of an abstraction layer, though the working practices of both designers and managers might need to change to accommodate such an innovation.

4.3.4 Sales before features or, new hardware is not enough

In videogames, specialist devices must achieve significant market adoption before they can be properly integrated into games. Individual titles frequently cost tens of millions of dollars to produce. Consequently, in order to ensure a financial return, they must reach out to players who own mass market platforms and interface devices. A novel input or output device that ships with device drivers for common operating systems and hardware platforms might well be usable as a substitute for existing standard interface devices (e.g. substituting for the keyboard, mouse, or console joysticks and buttons), because most of the common gameplay actions controlled by players may be emulated across a number of devices. (A current exception is the Wii controller, though alternative gestural input devices are in the pipeline.) However, the same device substitution approach is far less feasible on the videogame output or display side. Plugging in a haptic or sonic rendering device would do little to advance the cause of alternative sensory output because there would be little (visual or other) output that could be delivered meaningfully through this sensory channel. (In some cases, there is a double hardware bottleneck. For example, few games produce stereoscopic visual output, despite the fact that a number of graphics cards provide hardware support for this form of display. This is because of the absence among gamers of suitable stereoscopic display devices. Crossing one’s eyes for any length of time in front of the computer screen is simply a non-starter.)

On the software side, games would need to be specially written to showcase the use of alternative sensory modalities; indeed, they would need to be hand crafted for specific individual modalities. It is simply uneconomic for this to happen until there is a sufficiently large ownership -- and, importantly, users -- base for new devices, especially among game-playing consumers.

This highlights the two Catch-22 loops that exist in this situation. The first is that new sensory display devices need to sell extremely well before mainstream games can include them in the gameplay, but such sales are largely dependent on games being written to showcase their appeal. The second loop brings the role of the abstraction layer into focus. Because games do not have an abstraction layer for managing sensory output, users cannot simply switch display devices and ‘play’ a game’s information streams through interchangeable sensory renderers. However, it is not simply the absence of the abstraction layer that causes this problem, it is also due to the information base (i.e. the root source) of the emitted sensory information being different from that found in data visualization software. In other words, audio/visual data for games are authored separately, tailored for their respective sensory channels, involving aesthetically crucial, but ultimately non-transferable, details.

This discussion leads inexorably to a crucial question: is it feasible to remove this chicken-and-egg roadblock by writing an abstraction layer to manage videogame sensory output? This question is answered in the next section.

4.3.5 Emotion not information

In videogames, much non-visual display is created to have an emotional rather than a strictly informational impact; it does this by addressing the affective areas of the player’s brain far more than the cognitive. (This point must not be over-stated, however, because a great deal of non-music audio provides players with information, in addition to providing emotional effects.) Because of this, the design of sensory output for games follows different rules to the design of sensory output for data visualization software. There is at least one other domain where this approach is adopted: background
and theme music in films. Few would seriously consider formalising, abstracting and standardising this cinematic component in terms of the semiotics of information transmission, though artists (e.g. Cadiz, 2006) have come close to doing so in some of their creative installations.

Research into audible icons (or ‘earcons’) provides useful evidence on user requirements in a closely related field. In a series of experiments in the 1990s, Gaver (1994) and others found that naturalistic sounds were far more acceptable by users to signify computer operations than sounds composed from data using abstract rules. In a similar vein, there can be few on-screen monsters in contemporary videogames that emit blood-curdling sounds which have been constructed by following a aural version of Bertin’s graphical sign system. The emotional impact of sounds (whether it is meant to chill the player’s heart or to foster endearment and trust) is more a product of creative design than a semi-automatic pipeline process based on multi-variate data being converted into sounds using audio rendering rules.

5. CONCLUSIONS

“Should bad data smell like bad cheese?” (Mark & Frank, 1992)

This paper began with a critical comment on the overwhelming visual bias of contemporary VGEs, and their failure to support the kinds of multi-sensory perceptual interactions that characterise most human behaviour in the real world.

The paper has highlighted the uni-sensory products of data visualizers, and has contrasted it with the multi-sensory products of videogame developers. Although the former appear to be principled while the latter appear to be pragmatic in their design of their information display software, such a reading is inappropriate. There has been a singular lack of progress within the GIS and commercial data visualization communities over the past decade in developing a coherent abstraction layer encompassing all of the main sensory modalities, an incomplete implementation of the 35-year old graphic sign system, and a belated effort to convert 2D GIS to 3D VGEs.

We have demonstrated how a branch of the entertainment industry with little knowledge of the principles of data visualization has managed to produce software that demonstrates a rudimentary, but nevertheless highly effective, form of multi-sensory data representation. Representing data through several sensory modalities offers the promise that all users, and not just the visually privileged, will be able to benefit from a progressive expansion of what has been called 'experiential computing' (Bricken, 1991; Heilig, 1992). It also opens up new possibilities for extracting meaning from data for everyone, whether they be career analysts or occasional information explorers. Although a number of technical and conceptual challenges remain to be solved, it is clear that our understanding of the issues involved far exceeds the efforts currently being made to create systems that harness multiple human senses for the routine interpretation of data. Despite the evident benefits of a multi-sensory approach to data interpretation, we are still some distance away from being able to exercise in our virtual environments most of the perceptual skills we routinely deploy in our everyday lives.

6. NOTES

1. The sources used to construct the original version of Table 3 are given in Shepherd (1995a). These have been augmented by material from Cohen & Wenzel (1995).

7. REFERENCES


