Will I See Sensor-Based Games Happen Before I Die?

Michael Zyda, University of Southern California

In this, the second "Games" column, we look at the possibility of developing sensor-based games. We do this from a high level, pointing out that all the pieces exist but have not yet been put together.

Welcome to the "Games" column. It is a bimonthly column about topics I'm thinking about with respect to future technology for the games industry. This month's installment is "Will I See Sensor-Based Games Happen Before I Die?" It would be nice if we could. And so this column is about how I think we have the technology to do this mostly in hand, but we have not yet joined all the pieces together in a sufficiently compelling demonstration such that we "have to have this." So, what are the pieces? I will sketch this out for you as if I am standing in my office at a whiteboard.

FROM A HIGH LEVEL, WHAT DOES A SENSOR-BASED GAME LOOK LIKE?

Figure 1 is my quickly thrown together high-level concept of what a networked sensor-based game looks like. Now, I assume the game world is networked and that there are multiple humans and artificial intelligences (AIs) in it, as we are all moving toward the metaverse. And, of course, we could build this all inside one machine and have just one human and one AI if that is the extent of our creativity, but that is so not anyone I know.

The game world is a distributed, shared dataspace representing the state of the game—pieces of the dataspace are in the memory of the game arbiter (think: gameplay server), and some of the pieces are in the computational clients that represent biometrically sensed humans (human i) and emotion-cognizant AI characters (AI i). The humans and AIs live in a sensed world where the world state,

Digital Object Identifier 10.1109/MC.2021.3112320
Date of current version: xxxxxx

0018-9162/21©2021IEEE

EDITOR **MICHAEL ZYDA**
University of Southern California;
zyda@usc.edu

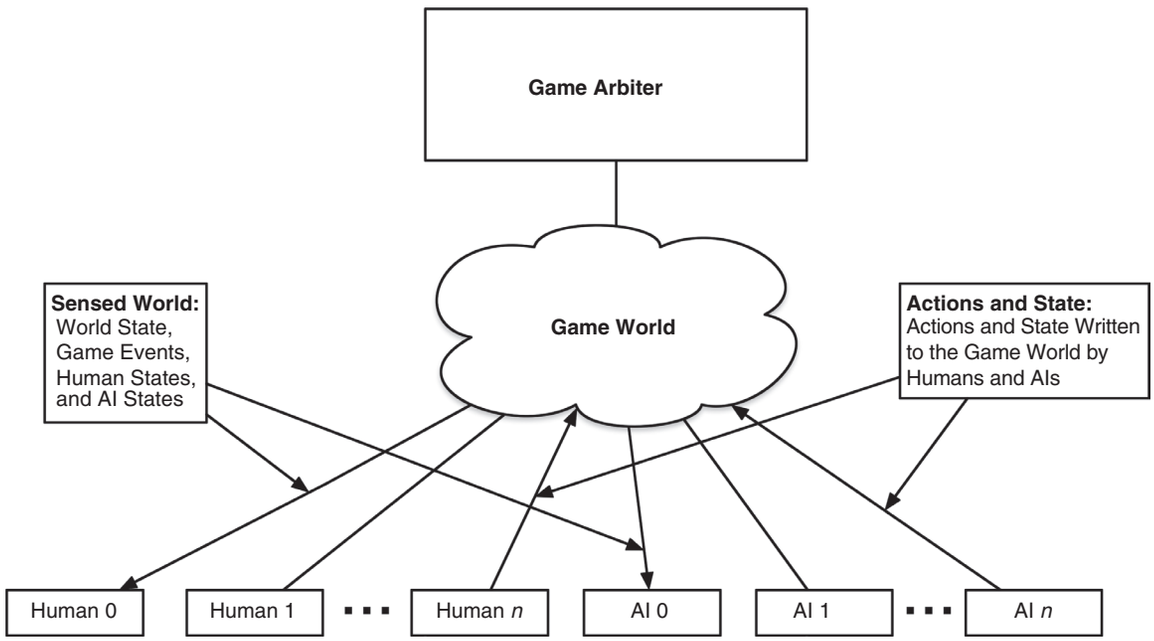

FIGURE 1. The high-level concept for a networked sensor-based game.

COMMENTS?

If you have comments about this article, or topics or references I should have cited, or you want to rant back to me on why what I say is nonsense, I want to hear. What I'm going to do is every time we finish one of these columns, and it goes to print, I'm going to get it up online and maybe point to it on my Facebook (mikezyda) and LinkedIn (mikezyda) pages so that I can receive comments from you, and maybe we'll react to some of those comments in future columns or online to enlighten you in real time! This is the "Games" column. You have a wonderful day!

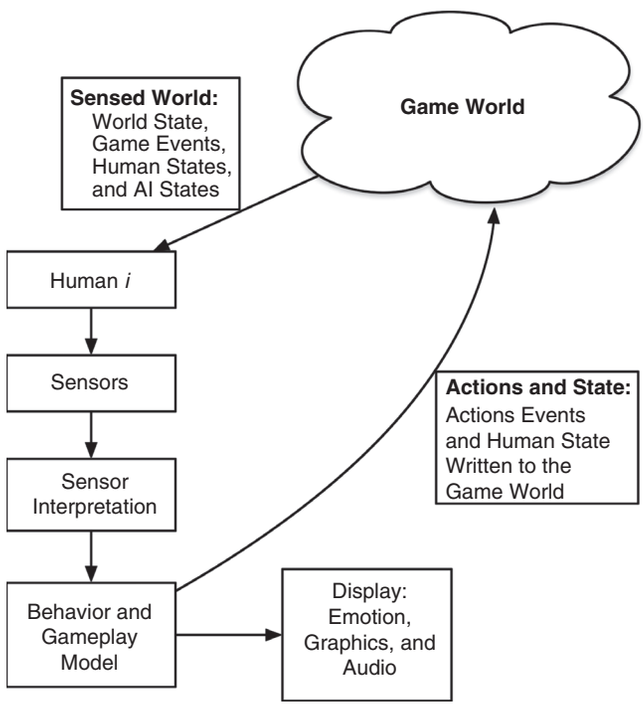

FIGURE 2. The human i entity.

game events, and human and AI states are brought to each entity via packets brought down from the network. Each entity, humans and AIs, writes back to the game world any action and game state changes that have occurred in its client entity. This is high level and simple, and the final architecture may vary.

WHAT DO THE BIOMETRICALLY SENSED HUMANS LOOK LIKE?

Figure 2 offers a brief, high-level architecture of the human entity. Human i is instrumented with a variety of biosensors, biosensors that can provide signals that can be interpreted into the physical and emotional state of the instrumented human. Now, I am not going to be comprehensive and point out the particular sensors you should acquire for this. Rather, I am going to try and convince you that these exist either in research, prototype, or commercial form. You may have to open them up and hack them or program them to make them work for your particular requirement.

Physical state

There are many things that can measure your physical state now—an Apple Watch is one, for example. With the Apple Watch, it knows if we are walking, running, swimming, or engaging in pretty much any fitness type of activity. The Apple Watch can also measure your heart rate and do an electrocardiogram for you. It can even issue a warning message if you stumble and fall and cannot get up. So, some physical measurements are available easily.

Some physical measurements are a bit more work. For example, if we want the current state of all the joint angles of a live human, we can use things like the Microsoft Kinect to get the joint angles of the major articulations of the human skeleton and then apply those in tandem with inverse kinematics to a 3D, personalized avatar so that movement by our live human can be mimicked. Now, the Kinect doesn't quite get us all the joint angles that we want. If we want fingers, there is other technology that comes in the form of an instrumented glove or a high-resolution camera pointed directly at the hands of the live human from a close vantage point.

There are other kinds of things that one can buy or make in prototype, and they all have the same dictum: the more detail you want, the more sensors you must apply and the more bandwidth you will require to get that data from the sensors into a computation platform that is connected to the network that connects to the game world. So, what I am saying is that we can pretty much solve the physical state sensing requirement for the human if we just collect the right sensors and package them and attach them appropriately on the sensed human.

EMOTION STATE

There are many different technological solutions to determining the emotional state of the human, and I briefly cover them in the following. My feeling is that to successfully determine the emotional state, perhaps several of these may be required, with a fusion of the results of each method utilized.

Speech emotion recognition

Speech emotion recognition (SER) is one of the longest-known and most promising methods for determining the emotional state of a human. Schuller, in his 2018 survey, articulates that there are two models used in practice, the first being discrete classes of "emotion categories including anger, disgust, fear, happiness, and sadness often added by a 'neutral' rest-class as opposed to a *value* 'continuous' dimension approach that appears to be the favored approach today." In the second model, Schuller states that "the two axes *arousal* or activation (known to be well accessible in particular by acoustic features) and *valence* or positivity (known to be well accessible by linguistic features) prevail alongside others such as power or expectation."[1] The Schuller paper includes a detailed block diagram of an emotion recognition engine as well as a suggestion that machine learning via the use of a generative adversarial network with a well-authored training set is where most current progress in SER is happening.[1]

The company AudEERING (http://audeering.com) has available now a Unity plug-in solution that listens (captures your player's voice), detects (uses an AI model that analyzes voice features and identifies your players' emotions and their intensity in real time), and interacts (the output values of the AI model are used to trigger events in the game that are fitted to the player's moods). So, basically what we find is that right now, we can reach out and get the emotion state of a speaking human in a straightforward way with a simple Unity plug-in from AudEERING.

Facial expression analysis from webcams

There are many people who have tried using webcams to measure facial expressions and to utilize those expressions for the determination of the emotional state of a human. This is even obtainable via various commercial packages. One easily findable package, on the imotions.com website, indicates that the software provides "20 facial expression measures (action units), 7 core emotions (joy, anger, fear, disgust, contempt, sadness, and surprise), facial landmarks, and behavioral indices such as head orientation and attention." These collected measurements are provided with a probability value representing "the likelihood that the expected emotion is being expressed."[3] The iMotions emotion recognition technology originates from Affectiva, a company cofounded by Rana el Kaliouby and Rosalind Picard, of the Massachusetts Institute of Technology Media Laboratory. Affectiva is now part of Smart Eye (https://www.affectiva.com/what/products/). Now, I'm not trying to be

comprehensive, but it is clear that we can utilize webcams to determine a human emotional state. You can buy this technology.

Hybrid electroencephalography

One of my long-term interests has been low-cost, hybrid electroencephalography (EEG) sensors. They have appeared over the past two decades and are direct descendants of earlier EEG and functional magnetic resonance imaging research. These sensors measure several biometric signals, such as EEG, blood oxygen, and motion. The signals come off the sensors via Bluetooth links and provide a number of human emotional state vectors: mental engagement, physical engagement, surprise/response (how much response there is to new material or events), relaxation, valence (like/dislike), learning/not learning and at what difficulty level, eye blinks, breathing rates, and pulse rates, among others. Additional analysis software indicates that sensed game players can be detected as to when they are "in the zone" (playing a game with little mental effort and fully engaged) and when a game has lost them by requiring too high of a level of mental engagement—basically, when mental engagement is greater than physical engagement.

The company Emotiv has "developed a suite of algorithms that detect human emotions, so that you can get immediate neuroscience insights without an expert knowledge," and the software works with its various EEG caps and wearables. Emotiv suggests that detected human emotions can be leveraged "to uncover the emotions that drive your audience choices."[4] Yes, the greatest minds of our age have all focused on understanding why we buy something so that they can then sell us something else. Now, we clearly can see that there are multiple ways to get emotional states, and all of these return a vector of state information of the detected emotion and the probability of that emotion. So, the next thing we must do is take the physical and emotional state vectors and pass them into an appropriate behavior and gameplay model.

BEHAVIOR AND GAMEPLAY MODEL

So, what does this behavior and gameplay model do with the physical and emotional state vectors that get passed in as measured from a sensed human? Well, we have a sensed human who is playing a game, and we now know what that human is doing physically and emotionally with respect to how that human perceives what he or she should next be doing in the game. That is straightforward, and what we are trying to do with the physical information is create action messages back out to the game world. The client behavior and gameplay model basically contains the methods of interaction the human can have with the game world and the other human and AI characters in that world. We additionally use the emotional state vectors in that action computation in some fashion to impact how well the human player really is allowed to perform in the game. The aggregated human physical and emotional states as well as computed gameplay actions are communicated to the game world network.

OK, the preceding is handwaving and a bit high level, but it is basically what a game client does once it knows what a sensed human wants in terms of gameplay—we have to remember that this game client also receives the rest of the sensed world (the world state, game events, human states, and AI states). And this makes the assumption that inside the behavior model is the gameplay model of what the human does in this particular game with this particular collection of sensed world

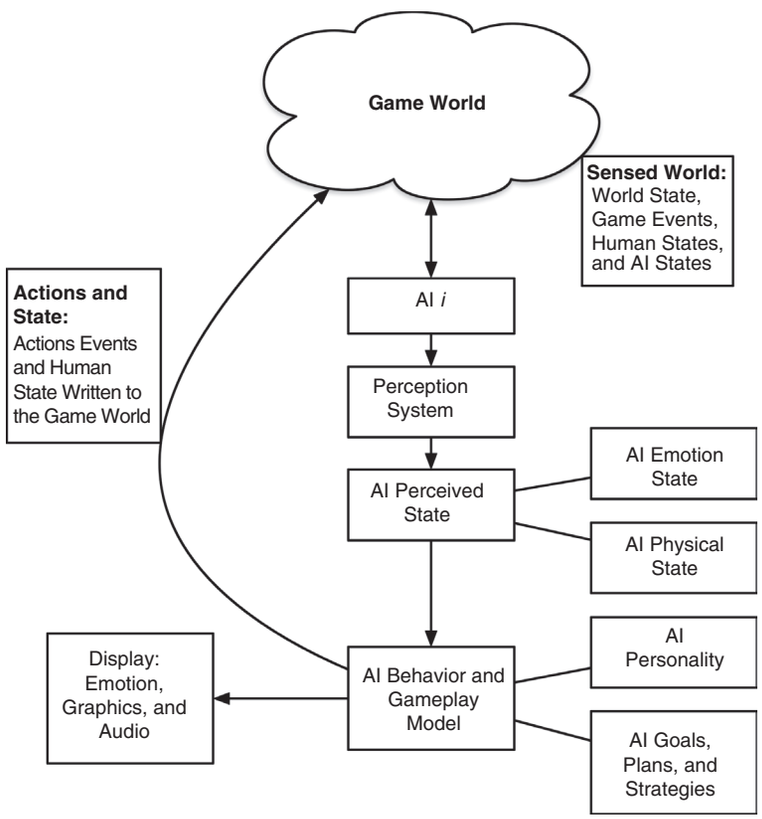

FIGURE 3. The AI i entity.

data (the world state, game events, human states, and AI character states).

AI CHARACTERS

In Figure 3, AI characters are connected to the game world and receive all the same information the human gets: the world state, game events, human states, and AI states. The AI character has to generate action events and physical and emotional states back to the game world. The first module we have for the AI is a perception system of some sort. That perception system receives perfect information about the state of the game world and its characters, and we have to blur that perfect state in some fashion so that the AI cannot act perfectly. Once we have the AI perceived state, we pass it to the AI emotional state module and the AI physical state module and then pass all of that to the AI behavior and gameplay model. The AI behavior and gameplay model uses the AI personality and AI goals, plans, and strategies models to generate the aggregated AI state for this AI character and the AI's actions and send those out to the game world.

GRAPHICAL, EMOTION, AND AUDIO DISPLAYS

Now, once the behavior and gameplay model has determined what the human and AI characters are doing, it then needs to generate a display for each entity. There are graphical displays, emotion displays, and audio displays that must be created. With respect to the human, we know its physical and emotional states and must generate computer graphics for them. What that means is the proper pose for the human both physically and with respect to emotions. The proper emotion display impacts the body language of the human as well as its facial display. With respect to the audio display for the human, this means we must properly represent what the human is saying about its physical and emotional states and expressions. With respect to the human and audio, we could just stream what the live human is actually saying if we don't want to solve the harder problem of adding emotion to a computer-generated voice.

For the AI character, we have the same displays to create—it will be very interesting the first time we generate an emotion display for the AI character interacting with humans. The display components, the physical and emotion displays, are well-known techniques in computer graphics. The proper generation of an emotion-tinged voice for the AI is perhaps a bit harder, but there is research to reference in this area.

DIFFICULTIES, AND ARE WE READY TO PUT THIS ALL TOGETHER?

The preceding is clearly just a high-level look at how we might begin to put together an architecture that allows us to build games imbued with emotional states and interactions. We can get "research-quality sensors and software," but we do not have much in the way of the low-friction sensor/software technologies for game developers and then consumers. We also need authoring tools usable by game developers. We need an authoring tool for specifying the AI character's personality. We might start out with a simple tool with personalities that are exaggerated, as they are usually in film. In fact, we might just select from a few personality types, initially, and provide a personality attribute strength value based on how strong we want that personality to impact gameplay. We don't yet have technology that is capable of taking data from sensor-based gameplay and interpreting it in a way that creates game characters that feel emotionally authentic and … well, human. But we won't get there unless we just start to try.[2]

The harder authoring tool is the one that allows us to write goals, plans, and strategies for our AI characters. Most current game-level authoring tools work by setting simple goals like *walk from point A to B* or *hide if you are in range of an enemy character*. So, the real work in this authoring tool deals with how complex a story we wish to be able to create, a story imbued with strong personalities and emotion states. And the end goal, of course, is to create the subtleties of story and emotion that we find in well-written film, but this time in the game realm. And that is when we can potentially craft games that make us cry. And not just because Metacritic rated them poorly.

ACKNOWLEDGMENTS

One cannot draft a treatise in this area without giving acknowledgment to those who provided discussions that helped form one's ideas. Hans Lee and Mike Lee, formerly of Emsense, provided many lively discussions about what was possible with hybrid EEG back in 2005–2011. I served as an advisor to Emsense at that time. Erin Reynolds, a former student from the University of Southern California and now an advisor to Athanos3D and full time at Disney, read the first draft of this article and provided me with wonderful suggestions to make it more solid. Erin has experience using heart rate sensors with her horror game Nevermind, and her help is greatly appreciated. Dr. Khizer Khaderi, director of the Stanford Human Perception Laboratory, has also provided me with interesting discussions on biometric sensing, starting in about 2005 and continuing to the present. Other, more long-ago people who influenced my career directions and hence my ability to draft this include Prof. Kent R. Wilson, University of California, San Diego, who loaned me his office in the evenings in 1973–1976, where I read through his physical chemistry, quantum mechanics, and neuroscience texts. One of those texts was *The Metaphorical Brain*, by Michael Arbib, which I read in 1973, and it convinced me that there is something important that computing can do for brain modeling and sensing.

REFERENCES

1. B. W. Schuller, "Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends," *Commun. ACM*, vol. 61, no. 5, pp. 90–99, May 2018. doi: 10.1145/3129340. [Online]. Available: https://cacm.acm.org/magazines/2018/5/227191-speech-emotion-recognition/fulltext
2. E. Reynolds, private communication, Aug. 2021.
3. iMotions. "Facial expression analysis: Gain deeper insights into expressed facial emotions." iMotions. https://imotions.com/biosensor/fea-facial-expression-analysis/ (accessed Sept. 13, 2021).
4. Emotive. "Consumer insights." Emotive. https://www.emotiv.com/consumer-insights-solutions/ (accessed Sept. 13, 2021).

MICHAEL ZYDA is a professor of engineering practice in the Department of Computer Science, the University of Southern California, Los Angeles, California, 90089, USA. Contact him at zyda@usc.edu.