

Creating Virtual Characters

Marco Gillies

Department of Computing,
Goldsmiths, University of London
London, UK
m.gillies@gold.ac.uk

ABSTRACT

An encounter with a virtual person can be one of the most compelling experiences in immersive virtual reality, as Mel Slater and his group have shown in many experiments on social interaction in VR. Much of this is due to virtual reality's ability to accurately represent body language, since participants can share a 3D space with a character. However, creating virtual characters capable of body language is a challenging task. It is a tacit, embodied skill that cannot be well represented in code.

This paper surveys a series of experiments performed by Mel Slater and colleagues that show the power of Virtual Characters in VR and summarizes details of the technical infrastructure used, and Slater's theories of why virtual characters are effective. It discusses the issues involved in creating virtual characters and the type of tool required. It concludes by proposing that Interactive Machine Learning can provide this type of tool.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**; *Collaborative and social computing*; • **Computing methodologies** → **Animation**; *Machine learning*; Motion capture;

KEYWORDS

Virtual Reality, Virtual Characters, Animation, Machine Learning

ACM Reference Format:

Marco Gillies. 2018. Creating Virtual Characters. In *MOCO: 5th International Conference on Movement and Computing, June 28–30, 2018, Genoa, Italy*. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3212721.3212835>

1 INTRODUCTION

One of the most compelling experiences in virtual reality is a face to face encounter with another person. This is quite unlike simply seeing an image of a character on a 2D screen. The person can be life-size; right in front of you. They can face you just as in real life. But perhaps most importantly you share a space with them. This seemingly simple fact enables the social cues that are central to face to face conversation. Body language, non-verbal communication, is

central to how we interact with other people. The shared 3-D space of virtual reality allows many aspects of body language to work in a way that they simply cannot on screen. Making eye contact feels real, it feels like a character is looking at you not simply looking out of the screen. Turning towards you or away from you means the same thing as in real life, moving close to you feels intimate, not simply a camera zoom. Body language gives social interaction in VR a power and intensity that is only otherwise possible in a real face-to-face encounter. I will therefore focus on body language in this paper, as it is the particular element that makes VR interaction more compelling than other media (and verbal interaction has been discussed extensively elsewhere by authors much more qualified on that subject than I am).

In virtual reality we can encounter other real people, or at least their graphical representation: avatars. We can also encounter fully virtual people, agents whose appearance and behavior are both generated by computer, what, in games, would be called non-player characters. In both cases it is body language that can give realism to our social interactions. In the case of avatars this implies that we need good enough tracking to communicate our body language. Agents on the other hand need algorithms that can generate realistic body language in real time. These are both important challenges for the future of VR. While avatars will be vital for the equivalents social media in virtual reality, agents will be the characters in narrative VR, the equivalents of film and story driven games.

Given the huge importance of human characters in almost all existing media as well as the far greater impact of social interaction in virtual reality, it seems likely that virtual characters will be hugely important as VR transitions from a technology to a medium. This raises a huge challenge of how to create compelling virtual characters that are capable of believable social interaction with real people. This position paper presents a personal view of how to create such characters grounded in 20 years of research in virtual humans and virtual reality, particularly the time when I was privileged enough to work with Mel Slater, the great thinker on virtual reality.

2 SLATER'S EXPERIMENTS ON VIRTUAL CHARACTERS

From the late 1990s to the present, Mel Slater and his teams in London and Barcelona performed a series of experiments that demonstrated the impact of social encounters with life sized virtual humans and virtual reality.

One of the first of these was the virtual audience[14]. Participants were asked to give a short oral presentation while immersed in a virtual reality display. In front of them they could see an audience composed of a small number of animated virtual characters.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MOCO, June 28–30, 2018, Genoa, Italy

© 2018 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

ACM ISBN 978-1-4503-6504-8/18/06...\$15.00
<https://doi.org/10.1145/3212721.3212835>

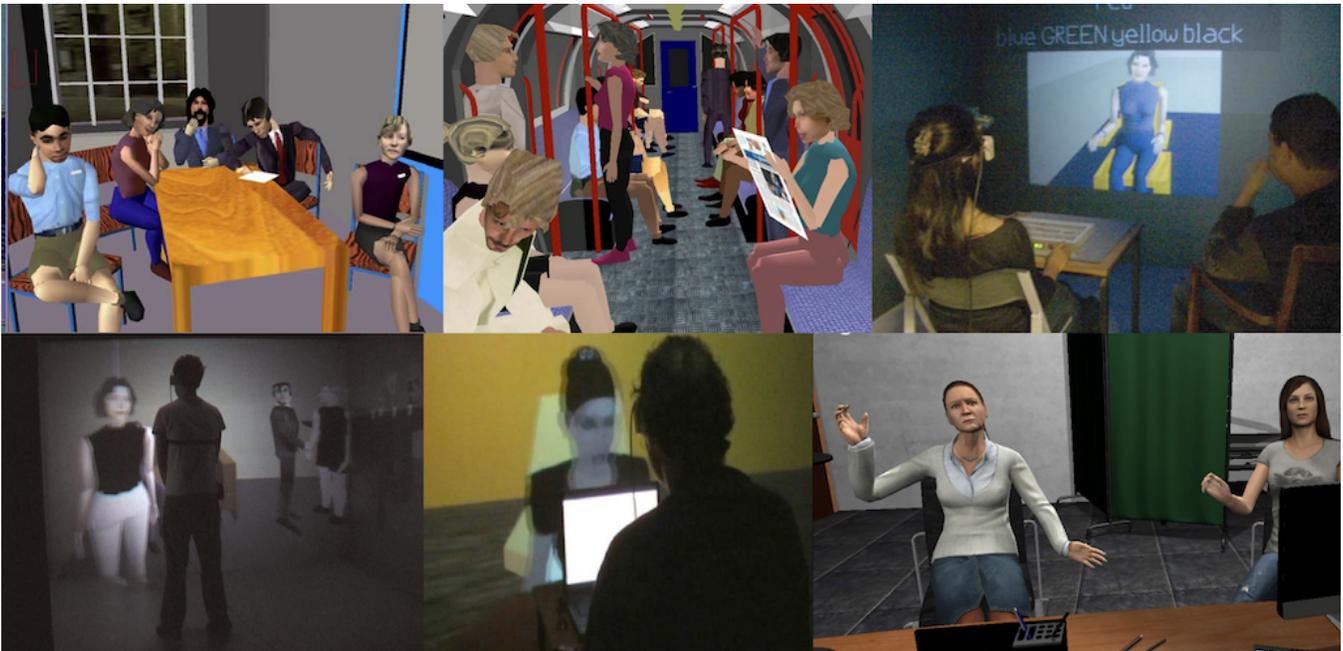


Figure 1: Slater’s experiments on Virtual Characters. (left to right and top to bottom): the virtual audience[14], the London underground[2], the Milgram recreation[16], the virtual bar[11], the socially anxious virtual character[12] and the medical consultation[13].

While the characters were not particularly graphically realistic, even by the standards of the time, they were animated with body language typical of an audience at an academic or business presentation. This body language could be varied between two conditions. One was extremely positive smiles, nodding and a lot of direct eye contact (which was possible because the participants’ heads were tracked). The other was very negative with characters slouching, looking around, muttering and falling asleep. The animation itself was largely pre-scripted though an experimenter was able to control it to some degree so that responses happened at an appropriate time in the speech. Participants showed dramatically different responses in the two conditions. Those that had the negative audience evaluated their performance significantly worse than those with a positive audience, even though both were fully aware that neither audience was real. Anecdotally experimenters noted that participants struggled to talk to the negative audience with some even directly criticizing the audience for their behavior. This experiment for the first time showed the dramatic effect of virtual characters on peoples behavior and in particular the importance of body language.

Freeman *et al.*[2] ran an experiment in which participants stood inside a virtual London underground train containing a number of virtual characters that were programmed to have neutral behavior and not interact with the participant. They found that participants with a history of paranoid delusions and also a sizable proportion of healthy members of the public had paranoid thoughts about the virtual characters thoughts and behavior towards them even though the characters behavior was objectively unrelated to the

participant. For example *“There was an aggressive person - his intention was to intimidate me and make me feel uneasy”*. A number of the healthy participants also interpreted the characters behavior as directed towards them but in a positive way for example *“It was nice much nicer than a real experience - people aren’t so forthcoming with their feelings in a real situation. Thought they were pretty friendly”*. These experiments show that people are predisposed to read self related attitudes into entirely preprogrammed behavior of virtual characters.

Perhaps the most dramatic of Slater’s experiments is the recreation of Stanley Milgram’s notorious obedience experiments[16]. In the original experiment participants were led to believe that they were giving electric shocks to a fellow participant. They would arrive at the laboratory at the same time as another supposed participant, who was in fact a confederate. They were assigned seemingly at random to one of 2 roles, the teacher and the learner. Though in fact the true participant was always the teacher. The participant was watched by an experimenter in a formal lab coat as they asked questions of the learner. If the learner made a mistake they were instructed by the experimenter to press a button which they were told would administer an electric shock to the learner. The learner would respond with obvious signs of pain and distress, though in fact there were no electric shocks and the learner was simply acting. The shocking finding of these experiments was that the majority of participants continued to administer shocks even when the learner appeared to be in extreme pain and possibly dying. The experiments gave an important insight into how obedience to authority can result in shocking behavior from ordinary members of the public, however that also resulted in an ethical outcry due

to the extreme distress and guilt caused to the participants. The resulting ethical guidelines for research ensure that a similar experiment could never be conducted again. Slater and his team realized however that it would be possible to run the experiment in virtual reality and with a virtual character because it would be entirely obvious that there was no real person coming to any harm. The new experiment largely followed the original protocol except the learner was a virtual character viewed in the VR display. While the character displayed distress with verbal and non-verbal responses modeled on those described by Milligram, at no point was it implied that this character was real. It is not surprising the majority of participants complied with the experimenter and continued to administer “electric shocks”, after all it was clear that no one was being hurt. What is surprising is that the majority of experimental participants experienced distress at their actions and a minority did, in fact, feel so distressed that they refused to continue the experiment. This shows that even though the virtual situation is not as emotionally powerful as the equivalent situation with a real person it can still generate strong responses emotions some people even if it is very clear that no real person is involved. The responses therefore do not appear to be due to an intellectual understanding of the situation but a lower level more emotional response to a virtual human’s verbal and primarily non-verbal behaviors.

While all these experiments show the impact of virtual humans in virtual reality, in all cases the participants interaction with those virtual humans was rather indirect. In one case, the London Underground experiment, the characters did not respond at all to the participants. In the case of the virtual audience it was a group response to a monologue from the participant and in the case of the Milgram experiment the interaction on the participant was simply pressing a button (though the impact was implied to be considerable). None of these cases would quite like an ordinary face-to-face conversation which is a symmetric, two way flow of dialogue and body language. Simulating this kind of full interaction is technically challenging in a number of ways as the character must very genuinely respond to the participant. Despite these challenges Sylvia Pan and Mel Slater ran a number of experiments that investigate face-to-face conversations between participants and virtual characters. The characters appeared to respond both verbally and nonverbally to what the participant said and how they behaved. I will describe the details of how this works technically later in this paper.

In the first of these experiments we produced a very normal and pleasant, though potentially somewhat stressful, social interaction[11]. Male participants were asked to enter a virtual environment that looks like a typical bar with several virtual characters in it and they were approached by a virtual young woman that began to talk to them. She made pleasant conversation asking questions about the participants and making some flattering comments. The experiment was therefore an interaction with a stranger of the opposite sex, which could potentially be stressful, but that had a positive outcome. The experiment was conducted with two groups of participants one with high levels of social anxiety and the other with very low levels of social anxiety, with the aim of studying and potentially helping treat social anxiety. The experiment seemed to be stressful for both groups, at least at the beginning, as measured by biometric markers of arousal. This was particularly pronounced for participants who

reported that they were not in a relationship (they were asked this by the virtual character). However, participants with high social anxiety showed a significant decrease in anxiety after the experiment as measured by a post questionnaire. This experiment shows that a virtual social interaction can have a strong emotional impact but can also be sufficiently positive to affect peoples attitudes to social situations. This is particularly interesting as the previous experiments showed much stronger impacts in negative situations.

A follow-on experiment reversed the situation, with a virtual character that displayed behavior associated with either low or high social anxiety[12]. Participants were asked to interview the character asking questions on the personal nature that were potentially embarrassing. While the content of the characters speech was the same in both conditions, non-verbal behaviour, both body language and tone of voice, conveyed either social anxiety or confidence. At the end of the interview the character left the room to answer her telephone and did not return. Participants were where were left to wait alone. They have been told that they could ring a bell to ask her to return. Participants who had interacted with the anxious character waited significantly longer before in the bell. This experiment shows that people obey norms towards a virtual character, being hesitant to disturb her and also that the characters non-verbal behavior can have a significant effect on peoples’ responses to them.

The realistic responses the participants make to virtual characters raises interesting possibilities for social skills training. Pan’s recent work has investigated this possibility for training medical practitioners how to interact with patients[13]. One of the most important medical challenges at the moment is antibiotic resistance. Prescribing antibiotics in large quantities increases the possibility that a new bacteria will emerge that are resistant to these drugs, raising the specter of many deadly and untreatable diseases. So antibiotic resistance is in many ways caused by doctors prescribing antibiotics when they should not. This tendency to prescribe is in large part due to pressure from patients. Pan *et al.*[13] conducted an experiment in virtual reality where family doctors were in a consultation session with a virtual woman and her virtual mother who is presented as suffering from an illness that clearly should not be treated with antibiotics. However, the daughter was extremely insistent that her mother should have antibiotics. In the experiment all of the participants who were trainee doctors and the majority of experience doctors prescribed antibiotics to the virtual woman, despite it clearly not being warranted by the symptoms and despite being observed which would get put pressure on them to act correctly. This is a clear demonstration that people respond to social pressure from virtual characters just as they would to real people.

Taken together, these experiments show the wide range of realistic social responses that people make to virtual characters when they interact with them in the immersive virtual virtual reality. They show just how powerful virtual characters can be in the situation and the importance of developing effective visual characters for the new generation of virtual reality experience. Before discussing possible explanations of this impact and how we should develop virtual characters in future I would like to briefly describe The technology behind the characters used in these experiments.

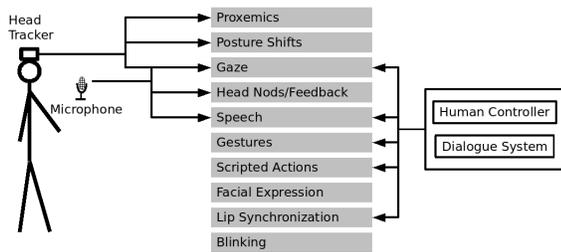


Figure 2: The model of virtual character behavior used in most of the experiments discussed in this paper.

3 THE PLAVCA MODEL

The majority of these experiments used a model of virtual character behavior that I developed around the software platform called PLAVCA[3] and shown in figure 2 (The exceptions being the virtual audience which was conducted before the development of this model and the bystander and medical doctors which use other platforms[8] though the model of behavior was relatively similar).

Generating social interaction and behavior is extremely challenging. It requires producing appropriate verbal and non-verbal responses to a wide range of unpredictable free-form interactions from a real person. At the time generating such free-form verbal dialogue was considered unfeasible (it still remains highly challenging but with technological advances might be starting to be possible now). We therefore took a Wizard of Oz approach in which an experimenter selected appropriate responses from the menu whenever the participant said something.

However, using a Wizard of Oz would not have been possible for non-verbal responses (body language). In part this is because they require very fast responses, faster than would've been possible for an experimenter to respond, but also because body language is largely subconscious, we do it without thinking about what we do, and so an experimenter would not necessarily know what actions would be appropriate in a given context (I will return to this theme later).

We therefore took a semi Wizard of Oz approach inspired by the work of Hannes Vilhjálmsón[18]. An experimenter could control the verbal aspect of interaction that were sufficiently slow and sufficiently conscious but the non-verbal aspects were automated. The verbal responses consisted of a library of recorded audio utterances with associated animations. The experimenter's role was to select one of these at any point in the conversation.

The non-verbal responses, on the other hand, were generated in real time in response to sensor data coming from the participant. The sensors used varied between experiments, for example the London Underground experiment had no responsive behavior at all. Even the most sophisticated set ups, for example Pan *et al.* 2012[11], use a very limited set of sensors: A microphone to detect the presence of speech (but not its content) and a 6 DOF head tracker. The head tracker enabled a number of different types of interaction based on the participant's movement all the distance

to the character. With these sensors it was possible to generate a number of types of response:

- Gaze behaviour: characters alternate between looking at the participant and looking away. The frequency of each could be varied depending on whether the participant was speaking
- Back channel feedback: The character could produce nods when the participant was speaking.
- Posture shifts: The head tracker could be used to detect posture shifts in the participants and the character could respond with their own posture shift as a form of coordinated behavior
- Social distance: The character could use the head tracker to determine distance to the participant, and maintain an appropriate social distance for conversation.
- General behaviour: as well as these responses the characters would make small movements by playing back pre-recorded animations

All of these behaviours were very simple relying simply on immediate reactive responses or at most a basic state machine. However, as we have seen above they could be extremely effective. Though the behaviors are very generic, it was possible to create distinct personalities for four characters by changing the animations (often with motion capture) and tweaking the parameters of the responsive models, but this was not easy to get right.

4 SLATER'S MODEL OF PLAUSIBILITY

These experiments with virtual characters had an important impact on Slater's theories of the psychological illusions that underpin virtual reality. Often collectively called *Presence*. The experience of virtual reality is based on reproducing real world sensorimotor contingencies, the relationships between our movements and what we perceive. The primary example being head tracking as we turn our heads the display updates and out the new changes just as it would in the real world. These sensorimotor contingencies create a sense of being physically in another space, what Slater calls *Place Illusion*.

However, these low-level sensorimotor contingencies are not enough to explain the complex interactions that happened between human participants and virtual characters. For that reason, Slater introduced a second illusion: *Plausibility*. An interaction with a virtual environment, or in this case a virtual character, is plausible depending on the degree to which it fulfills three conditions:

- The character responds to the behaviour of the participant
- The behaviour of the character refers personally to the participant
- The character's behaviour is credible in the sense that it fits expectations from similar situations in the real world

For example, in Pan *et al.* 2012 [11] the virtual character fulfills these requirements: her speech and body language both respond to the participant, she speaks directly to the participant and her behavior was scripted to be typical of conversation in a party situation.

The responsive body language model described in the previous section seems central to plausibility, as it enables responsive

behavior that refers personally to a human participant. Realistic, responsive body language is a key part of what makes an interaction with a virtual character in VR so compelling, and a key challenge for creating compelling characters is designing that body language.

However, the experiments also show that perceived responsiveness is enough. In the virtual audience and London underground experiments, participants read responsiveness into the behavior of the characters when it was very limited or even completely absent. People have a considerable capacity to read more response into a character than is really present (though it is easier to make behavior appear responsive if it really is responsive). So we can rephrase the requirements for plausibility as

- The character *appears* to respond to the behavior of the participant
- The behavior of the character *appears* to refer personally to the participant
- The character’s behavior *appears* credible in the sense that it fits expectations from similar situations in the real world

5 MAKING CHARACTERS NOT “AI”S

This leads us to an important point about the creation of virtual characters. While at first glance it appears to be dauntingly “AI complete”, in the sense that it seems to require a complete model of human intelligence. However, the aim is not to create an accurate simulation of a person, but to create the perception of interacting with a person. I would go so far as to say that the task is not the (intractable) scientific one of simulating a person, but the artistic one of creating a character in the sense that a novelist or actress will create a character. After all, Hamlet is not a complete simulation of a person, in many ways it is the incompleteness of the character that makes him interesting, because it opens up ambiguity and space for interpretation.

We can therefore view the problem of designing body language for a character as an artistic process, and predict that, those with experience of creating characters are likely to provide important contributions. In particular, actors are extremely good at creating compelling body language and are likely to play a key part in designing interactive characters for VR.

5.1 Tacit and Embodied

There is another important challenge in simulating and recognising body language. Body language is a tacit skill, in the sense meant by Polanyi[15], we are able to perform body language effortlessly and without thinking about it, however we are largely unable to describe consciously what we are doing when we do body language. To use Polanyi’s phrase, “we know more than we can say”, we know how to do body language we simply cannot say how we do it.

The first implication of this is for the interaction itself. When we interact with a virtual character we must do so largely implicitly. Unlike a traditional user interface where we explicitly choose to press a button, a virtual character must understand the non-verbal behavior that we are producing without thinking about it. Great care must be taken with this type of interaction. There are risks of misinterpretation, particularly when working with a wide range of different people and cultures. Elsewhere I have argued[5] that

more aspects of non-verbal interaction with characters should be made explicit where they can be.

The second implication, and the one I will focus on for the rest of this paper, is for designing body language interaction. Writing software is often very much about making explicit the processes and interactions that we perform. They must be made explicit because they must be described in computer code. Body language on the other hand, is very difficult to make explicit, if not impossible, since we ourselves only understand it implicitly. How can we implement the rules of body language. If we do not know what they are?

If creating virtual characters is an artistic process, as we said in the last section, then it should be artists that create our characters. Actors, in particular, have an excellent understanding of the nuances of body language and can perform body language in a highly expressive way. However, this understanding is very much implicit, a performance is not described explicitly in words except in the most vague terms. Modern schools of acting [17] emphasize entering into a character psychologically in order to be able to perform the actions of that character with the appropriate emotional tone. The non-verbal performance therefore comes from a process of psychological empathy not from an explicit script. This is the result of a long process of improvisation, rehearsal and refinement which would be very hard to capture explicitly in code.

5.2 Designing by doing

The process of designing a character by acting and of designing software by programming in many ways could not be more different. Programming involves understanding a problem, making it explicit, abstracting it and putting it into a symbolic form. While preparing a character does involve reading and thinking through the character, in general it is very much more grounded in action. Actors create characters by improvising scenes, i.e. they design the performance by performing it, physically acting out the actions of the character.

It is through this embodied process of improvisation that the character is designed. The fact that actors rehearse using their bodies enables them to make use of the tacit and embodied skills they have. It is not a problem that their knowledge of body language is not wholly explicit. Quite the opposite, improvisation and rehearsal require that knowledge to be embodied form. In this paper I would like to propose that actors methods of improvisation are an excellent basis for designing interactive virtual characters for virtual reality. The design process should not be like software design, but should be embodied and improvisational. We should design our characters by doing, by physically acting out interactions between characters and players. Only by physically performing these actions can we make use of our tacit body knowledge. Only by using real interactions between people can interactions with virtual characters feel later.

To summarize, body language is key to compelling social interactions in virtual reality. It is a tacit embodied skill, meaning that we all know how to do it but we do not know how to describe what we do when we do it. Creating virtual characters does not need to mean creating a complete AI simulation of a person. It is about creating the perception of a response of character, something much more like the artistic process of creating a fictional character. Actors are experts in creating characters but also in the tacit and embodied

knowledge of body language. That is why I want to propose actors processes as an ideal model for creating virtual characters.

6 DESIGNING VIRTUAL CHARACTERS

So how are virtual characters designed at the moment? How much do these methods cater for embodied knowledge and improvisation? How could we create design tools that are modeled on actors' processes?

Before we try to answer these questions we need to distinguish between the three key elements of virtual characters:

- Appearance: the graphics that determine how a character looks, normally consisting of polygon meshes, textures and possibly quite advanced materials
- Animation: the movements of the character which are normally represented as a sequences of frames of a skeleton rig. The animation moves the graphical mesh of the character.
- "AI": the algorithms that control how the character responds to inputs and how it pro-actively selects behaviour. The "AI" determines the animation of the character in response to inputs from players or the environment. (In this context I am using the term AI in the way that it is used in the games industry as opposed to the academic discipline of artificial intelligence, hence the quotation marks).

The appearance of a character can be created with a wide range of excellent artist focused tools. These tools are aimed primarily at visual artists, rather than animators, as this is the visual side of the character. However, the appearance of actors can be captured using photogrammetry or other scanning techniques.

There are two main approaches to animation, both of which are artist drive. The first is keyframe animation, which draws on computer animation skills. This consists of laboriously posing characters at different frames which are interpolated to create a full movement. This process is very much not embodied, and it is notable that animators often find it challenging to go from an embodied knowledge of movement to animation frames (though it can produce excellent results, particularly for stylized motion). The other approach is motion capture, which is the recording of the 3D movements of a persons body, or performance capture which is an enhanced motion capture process in which all elements of a performance: movement, voice and facial expression are captured simultaneously. Performance capture in many ways epitomizes the approach I am suggesting in this paper, it allows actors to define the movements of characters simply by moving, by performing. It is embodied and allows for improvisational and rehearsed processes. It is no wonder that it is now considered a gold standard for creating realistic and nuanced character movement. However, it can only produce linear animations that are designed to be played back, as they are. It cannot produced the real time responsive behavior described in the PIAVCA model above.

"AI", on the other hand, is still very much a programmer driven process. It uses custom algorithms which have to be hand coded or at the very least hand tweaked, and there is very little in the way of artistic tools beyond visual programming environments and certainly nothing that could be described as embodied design. However, "AI" is key to compelling virtual character body language in VR. It is not enough to simply play back motion capture data,

it will feel wrong as the character will not feel responsive. There must be real time responses to the behaviour of the player.

So current approaches have very embodied ways of producing animation that can be played back statically, but when it comes to real time responses, they fall back on programmer oriented tools. The PIAVCA model was able to include realistic movement, which in many cases was created by actors via motion capture. But this was restricted to the characters fixed utterances and did not change in response to the participant (beyond selecting which utterance to play). The real time body language was based on state machines and other algorithms whose parameters needed to be tweaked manually and so did not allow for artist oriented or embodied design processes.

6.1 Machine Learning

Is there a way of turning the creation of "AI" into an embodied, artist oriented process? We can go back to the development of the field of artificial intelligence over the last decade. Traditional AI approaches consisted of designing sophisticated algorithms to mimic human reasoning. However, over the last twenty years, and even more so in the last five, academic, and increasingly industrial artificial intelligence has moved towards machine learning, which instead of relying on hand designed algorithms, uses large data sets (for example of photographs of people) to train computers to perform tasks such as face recognition. The task when designing an artificial intelligence system is no longer about designing and tweaking algorithms but of curating data that serves as input to one of several machine learning algorithms (I would not want to deny the excellent work of many researchers developing new learning algorithms, however, providing data to an existing algorithm is a much more common task).

What does this imply for designing characters? Their AI could be designed with data not code. By data I mean a number of examples of the types of interaction that we desire from a character: for example we might have an example in which the character nods in response to the player talking with a certain tone of voice, but simply looks at the player when the tone of voice changes. How could these examples be created: using performance capture. Increasingly performance capture is being performed on two or more actors simultaneously to ensure realistic interactions. Currently the resulting data is simply played back, but it could be used as input to a machine learning algorithm.

Machine learning has already been used to great success for interactive virtual characters, for example Huang *et al*[9] have created a model of backchannel feedback based on data from real conversations and Levine *et al*. [10] have used machine learning for gesture generation.

I would like to propose a general framework (based on that proposed in Gillies 2009[4], called *Interactive Performance Capture*. This framework generalizes Performance Capture, an approach that has been highly successful for creating realistic animation by recording the movement of actors. Interactive Performance Capture extends Performance Capture by modeling the interaction between two actors. It uses Machine Learning to create a data driven approach to the "AI" of virtual characters.

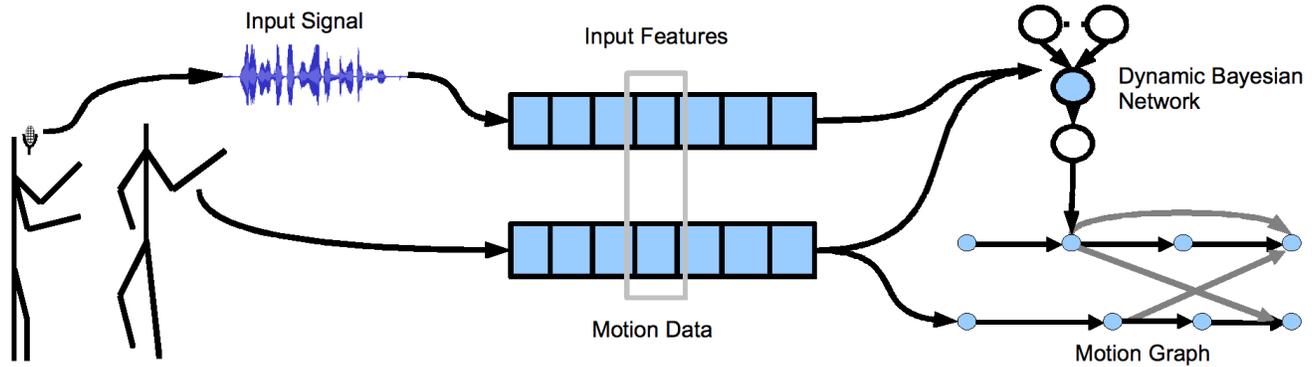


Figure 3: Interactive Performance Capture

The Interactive Performance Capture framework is shown in figure 3. It begins with standard performance capture (figure 3 left): two actors perform an interaction (they have a conversation). One actor plays the part of the character, the other the part of the player. The performance is captured via a number of streams of data: audio, facial expression and body movement capture. It is very important that the “player” is captured using tracking technology as close as possible to those available to a real player in VR (perhaps a microphone, plus the head and hand tracking that is built in to a head mounted display). However, the “character” can be captured with as high fidelity equipment as possible as their data will be used to generate animation. Improvisation is key to this process. While in traditional film and performance capture, the aim is to achieve a single “best” performance, the aim in interactive performance capture is to capture the many different responses that a player might make to the character and include ways in which the character could respond in turn.

The resulting data from the actor playing the character and the actor playing the player are synchronized (figure 3 center) so we can know which actions of the player created which responses in the character. The data are used in two ways. The data from the “character” as used as the basis of an animation model (for example a move tree) that is used to animate the character, just as they would be in a traditional work flow (figure 3 bottom right). The synchronized data are also used to train a machine learning algorithm (figure 3 top right). This algorithm generates a mapping from the player’s behavior to the character’s behavior which can be used to select animation clips to play from the animation engine. For example, in Gillies (2009) an Input-Output Hidden Markov model was used as it is equivalent to the finite state machine models used in traditional game AI. This model was able to select states based on input from the player and select animation clips based on the state.

6.2 Interactive Machine Learning

Machine Learning opens up the possibility of embodied character “AI” design, allowing actors to design the interactions of characters through movement, by using their embodied, tacit knowledge of

body language. However, there are still problems. Machine Learning is currently not a very usable technology, and is considered the realm of technical experts. There is no innate reason for this, since it is really about curating data, which should be possible by anyone who understands the domain (be it photographs, writing or human movement). I believe that there needs to be a considerable push to make machine learning usable[7].

Another problem is that it is not really an artistic process designed to create individual characters. Current machine learning approaches rely on large datasets captured systematically and designed to be broadly representative. As they normally include a large number of people tend towards modeling average behavior rather than individual nuance. Also, the need to capture data systematically and representatively, tends to suppress the experimentation, improvisation and iteration central to the artistic process.

However, this is largely due to how machine learning is used, rather than the algorithms themselves. The work of Rebecca Fiebrink on Interactive Machine Learning is very inspiring in this context. She developed the Wekinator[1], a machine learning tool aimed at electronic musicians wanting to design novel gestural musical instruments. They could design the gestures by giving examples, rather than by programming rules. Many of her collaborators found this a very embodied design experience:

the Wekinator emphasizes an embodied approach to composition ... by allowing the user to generate training examples by actually performing gestures in real-time, as opposed to approaching the mapping creation process rationally or mathematically, as is required in the design of explicit mapping functions

Her work shows that machine learning can be an artistic tool and can support embodied, tacit knowledge. It also shows that artists used her tool in a very different way from traditional machine learning. For example, they used a small number of carefully chosen data points, rather than large datasets. They also worked very iteratively, rather than collecting all of the data at the beginning, their concept of what they were designing often evolved during

the design process, resulting in both the data and the final gestures changing.

Together with Andrea Kleinsmith and Harry Brenton[6] I have tried to apply some of Fiebrink's ideas to designing virtual characters with actors. We found that machine learning can fit well with actors' working processes, but there are still problems. The biggest being debugging: initially training a model is relatively easy, but seeing how to fix it if it does not work as expected can be much harder. There needs to be a lot of research on providing visual feedback to creators about how machine learning algorithms work and so how to debug and refine them.

7 CONCLUSION

This paper has provided a personal view point on the future of virtual character technology, but I hope one that will be illuminating to people. Virtual reality is now at a cross roads. For three decades it has been a technology, restricted to research labs, it is now becoming a medium which is able to reach thousands and in future millions of people. At this point technologists need to hand over the baton to the artists who will be creating the virtual reality experiences that will move people deeply, making them laugh, making them cry and possibly changing their whole perspective on life. The responsibility of technologists is now to create tools that enable artists to create the best work they can. In no area of VR is this more important than Virtual Characters, which I believe will be central to the most profound and most moving VR experiences that we will see as the medium moves to maturity.

ACKNOWLEDGMENTS

I would like to thank all members of the Virtual Environments and Computer Graphics Group at University College London, particularly those that took part in the work presented here. I would specifically like to thank Mel Slater and Xueni (Sylvia) Pan. I would also like to thank my current institution, Goldsmiths, University of London, for all the support in my research.

REFERENCES

- [1] Rebecca Fiebrink, Perry R. Cook, and Dan Trueman. 2011. *Human model evaluation in interactive supervised learning*. ACM Press, New York, New York, USA, 147 pages. <https://doi.org/10.1145/1978942.1978965>
- [2] Daniel Freeman, Katherine Pugh, Angus Antley, Mel Slater, Paul Bebbington, Matthew Gittins, Graham Dunn, Elizabeth Kuipers, David Fowler, and Philippa Garety. 2008. Virtual reality study of paranoid thinking in the general population. *British Journal of Psychiatry* 192, 4 (2008), 258–263. <https://doi.org/10.1192/bjp.bp.107.044677>
- [3] Marco Gillies. 2008. Pivca: A Framework for Heterogeneous Interactions with Virtual Characters. In *IEEE Virtual Reality Conference 2008 (VR 2008), 8-12 March 2008, Reno, Nevada, USA, Proceedings*. IEEE, 255–256.
- [4] M Gillies. 2009. Learning Finite-State Machine Controllers From Motion Capture Data. *IEEE Transactions of Computational Intelligence and AI in Games* 1 (2009), 63–72.
- [5] Marco Gillies. 2016. What is Movement Interaction in Virtual Reality for?. In *Proceedings of the 3rd International Symposium on Movement and Computing - MOCO '16*. ACM Press, New York, New York, USA, 1–4. <https://doi.org/10.1145/2948910.2948951>
- [6] Marco Gillies, Harry Brenton, and Andrea Kleinsmith. 2015. Embodied design of full bodied interaction with virtual humans. In *Proceedings of the 2nd International Workshop on Movement and Computing - MOCO '15*. ACM Press, New York, New York, USA, 1–8. <https://doi.org/10.1145/2790994.2790996>
- [7] Marco Gillies, Rebecca Fiebrink, Atsu Tanaka, JÁrÁlmi Garcia, Saleema Amershi, Bongshin Lee, FrÁldÁldric Bevilacqua, Nicolas D'Alessandro, JÁuelle Tilmanne, Alexis Helior, Fabrizio Nunnari, Todd Kulesza, Wendy Mackay, and Baptiste Caramiaux. 2016. Human-Centered Machine Learning. In *ACM SIGCHI Workshop (full citation to follow)*.
- [8] Marco Gillies and Bernhard Spanlang. 2010. Comparing and Evaluating Real Time Character Engines for Virtual Environments. *Presence* 19, 2 (2010), 95–117.
- [9] Lixing Huang, Louis-Philippe Morency, and Jonathan Gratch. 2010. Learning backchannel prediction model from parasocial consensus sampling: a subjective evaluation. In *Proceedings of the 10th international conference on Intelligent virtual agents (IVA'10)*. Springer-Verlag, Berlin, Heidelberg, 159–172. <http://dl.acm.org/citation.cfm?id=1889075.1889095>
- [10] Sergey Levine, Christian Theobalt, and Vladlen Koltun. 2009. Real-time prosody-driven synthesis of body language. *ACM Transactions on Graphics* 28, 5 (12 2009), 1. <https://doi.org/10.1145/1618452.1618518>
- [11] Xueni Pan, Marco Gillies, Chris Barker, David M Clark, and Mel Slater. 2012. Socially anxious and confident men interact with a forward virtual woman: an experimental study. *PLoS one* 7, 4 (1 2012), e32931. <https://doi.org/10.1371/journal.pone.0032931>
- [12] Xueni Pan, Marco Gillies, and Mel Slater. 2015. Virtual Character Personality Influences Participant Attitudes and Behavior - An Interview with a Virtual Human Character about Her Social Anxiety. *Frontiers in Robotics and AI* 2 (2 2015), 1. <https://doi.org/10.3389/frobt.2015.00001>
- [13] Xueni Pan, Mel Slater, Alejandro Beacco, Xavi Navarro, David Swapp, Joanna Hale, Paul Alexander George Forbes, Catrina Denvir, Antonia F de C Hamilton, and Sylvie Delacroix. 2016. The Responses of Medical General Practitioners to Unreasonable Patient Demand for Antibiotics - A study of medical ethics using immersive virtual reality. *PLoS ONE* (2016).
- [14] D P Pertaub, M Slater, and C Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence-Teleoperators and Virtual Environments* 11, 1 (2002), 68–78.
- [15] Michael Polanyi. 1966. The Tacit Dimension. *Knowledge in Organizations* (1966), 135–146. <https://doi.org/10.1016/B978-0-7506-9718-7.50010-X>
- [16] Mel Slater, Angus Antley, Adam Davison, David Swapp, Christoph Guger, Chris Barker, Nancy Pistrang, and Maria V. Sanchez-Vives. 2006. A virtual reprise of the Stanley Milgram obedience experiments. *PLoS ONE* (2006). <https://doi.org/10.1371/journal.pone.0000039>
- [17] Konstantin Stanislavski and Jean Benedetti. 1938. *An Actor's Work: A Student's Diary*. Routledge.
- [18] H H VilhjÁlmsson and J Cassell. 1998. BodyChat: Autonomous Communicative Behaviors in Avatars. In *second ACM international conference on autonomous agents*.