

Mark Sagar, the Sci-Tech award winning (ex-Weta Digital) facial expert has just left Auckland University and got first round funding for his new company Soul Machines. Mark Sagar and his team at Auckland University developed by Baby X a working intelligent, emotionally responsive avatar.

Soul Machines attracted US\$7.5 million in a series A financing round led by Horizon Ventures with Iconiq Capital. The company is developing human-like avatars that will be a part of the developing world of artificial intelligence-based platforms.

There is a major push starting to develop worldwide for both avatars and cognitive agents. Avatars are representations of ourselves in either the virtual world or online.

We have written stories about the huge push Facebook is doing in this space to allow people to join friends in a VR space and express emotion and push for virtual presence. Agents are stand alone AI digital humans, where the logic of the emotional and cognitive response comes an AI simulation. While Mark's work might be used for avatars, it is in the area of cognitive agents that he is leading the world.

Soul Machine's first commercial project will be made public early this year. Our own Mike Seymour joined Mark and Annette Henderson to publish a look at the inner workings of Baby X, developed by the engineering research team at the Laboratory for Animate Technologies based in the University of Auckland's Bioengineering Institute.



A close up of the latest BabyX's eyes

The December issue of [Communications of the ACM outlines the Baby X project](#).

Unlike some other projects, BabyX is designed as both a stepping stone and an expandable base to even more complex agents. While BabyX does have her own AI backend, she was designed so that she could be expanded to connect to new AI engines as they are developed such as IBM's Watson (now Intu Project) AI project or others.

Why does Siri or Amazon Echo need a face? Of all the experiences we have in life, face-to-face interaction fills many of our most meaningful moments. The complex interplay of facial expressions, eye gaze, head movements, and vocalizations in quickly evolving "social interaction loops" has enormous influence on how a situation will unfold. Facial expressions are a rich and subtle way to convey meaning. Yet in a vastly computerised world, the dominant metaphor for dealing with the computer is a desktop. But people don't put a photo of their desk on the fridge or carry it around in their wallets. People respond to faces. Asked to view a photo of a friend's child or spouse and one will expect to view a picture of their face, – anything else would be creepy or odd. But even a poorly photographed image of a loved one naturally smiling is considered a 'great photo'.

From birth, we are genetically programmed to respond to faces. The natural interactions in the first months and years of life are a fundamental element of learning and lay the foundation for successful social and emotional functioning through life.

If faces are so important, then why have we not developed facial user interfaces before now? The answer lies in the uncanny valley and the notion of partial realism being worse than nothing in terms of affinity with the face. Much is written about the theory of the uncanny valley, but it is sometimes confused with a notion of a 'visual Turing test'. People conversationally discuss uncanny valley as if it means perfectly human, undetectable from real. This is not the core intent of the original 1970s robotic paper. BabyX is not meant to pass as a real baby being filmed in some remote location. But she is real enough that when interacting with her, she seems pleasant, and you enjoy the interaction.

Successful agents don't need to appear 100% real in the same way as they might need to in a major feature film when the digital human is giving a performance in place of an actor. For an agent to be successful, they need to invoke a sense of affinity and BabyX does this very strongly in those who have interacted with her. Far from 'being creepy' she causes laughter and delight.

But how does this virtual human baby work?

“BabyX” was created as an autonomously animated psycho-biological model of a virtual infant. This presents two clear problems, the representation of the baby; does she look and emote correctly, and the second challenge of interacting? In simple terms, BabyX is able to see you when you stand in front of her, via camera input that does facial tracking and voice analysis. She then has a bio-based ‘brain’ that reacts and she displays her emotional response. You can show her a pretty picture of a sheep and she will smile and say ‘sheep’.

Far from being a novel form of party trick, BabyX has a full developed model of a biochemical brain that processes and reacts. In more complex terms, you become present with her as you both interact. The simple notion of action-reaction as an intellectual ping pong match misses the point of BabyX. She is modelled to accurately provide you a sense of her being there as you talk and interact. She is not responding with predetermined scripted responses nor is she tricking the viewer with elaborate misdirected pre-programmed emotional cues.

To create an interactive scenario involving unscripted life-like interactions, is challenging. The issue is not only how to represent the complex appearance and movement of the synthetic face in real time. For an embodied agent’s behavior to be believable, it must be consistent and contextually appropriate. If the Agent can be affected by the interaction, and respond in a way that can affect the user, then each partner is more invested in how the interaction unfolds, creating engagement and emotional connection.

What creates all these fleeting movements that communicate so much? And how can a simulation keep them consistent, appropriate, and adaptive? Everything that happens on the face reflects a brain state or in human terms ‘one’s thoughts’. Because the behavior of the face is affected by so many factors—cognitive, emotional, and physiological—Mark Sagar’s BabyX explores a more-detailed, holistic and biologically based approach than has previously been attempted in facial animation.

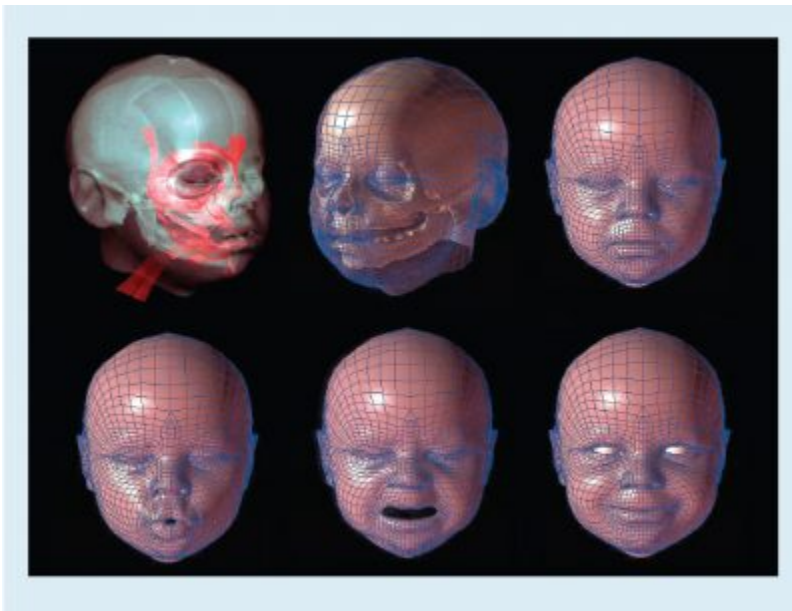
To actually build such a system the team came up with a “Brain Language” (BL) to create autonomous expressive embodied models of behavior driven by ‘neural’ models based on affective and cognitive neuroscience theories. The goal was to integrate different state of the art theories and models to create a holistic “sketch” of basic aspects of human behavior, with a focus on the face and interactive learning.

Most research in building autonomous human agents has been as “embodied conversational agents” (such as Siri) and this is generally at a language level, not specifically focused on the subtler details of facial expression and nonverbal behavior.

Much of the work on virtual humans has an unfortunately robotic “feeling,” particularly with facial interaction. This is possibly due to most virtual human models not focusing on the microdynamics of expression or on facial realism. These microdynamics are considered particularly critical in learning contexts. As child developmental cognitive experts [\(Katharina Rohlfing and Gedeon Deak\)](#) have commented, “when infants learn in a social environment, they do not simply pick up information passively. They respond to, and learn from, the interaction as they jointly determine its content and quality through real-time contingent and reciprocal coaction.”

Facial expression

BabyX’s computer-graphic face model is driven by muscle activations generated from motor-neuron activity. The facial expressions are created by modeling the effect of individual muscle activations and their non-linear combination forming her range of expressions. Mark Sagar was a champion of introducing FACS expressions to VFX pipelines for Gollum and King Kong, while he was at Weta Digital.



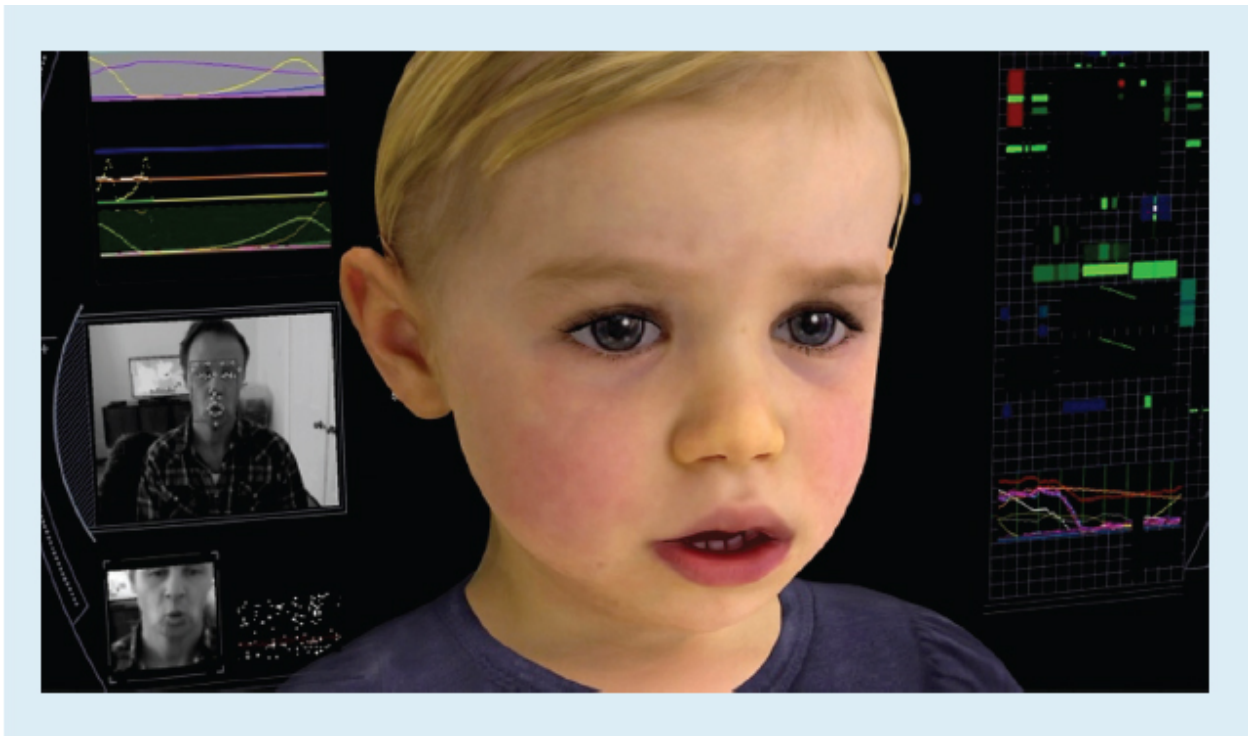
For BabyX, the modeling procedures involve biomechanical simulation, scanning, and geometric modeling. Fine details of visually important elements such as the mouth, eyes, eyelashes, and eyelid geometry, were painstakingly modeled. Mark scanned his own sleeping daughter as the basis of BabyX’s appearance.

A highly detailed biomechanical face model has been constructed from MRI scans and anatomic reference. Skin deformation is generated by individual or grouped-

muscle activations. The team have modeled the deep and superficial fat, as well as muscle, fascia, connective tissue, and their various properties. BabyX uses large-deformation finite-element elasticity to deform the face from rest position through simulated muscle activation. Individual and combined muscle activations were simulated to form expressions interpolated on the fly in BL as the face animates. The response to muscle activation is consistent skin deformation and motion.

The BabyX framework models the current best thinking on how the brain works. As such there is a direct correlation between the structure of BabyX and a human brain. For example, expressions are generated by neural patterns in both the subcortical and cortical regions of the brain. The subcortical area is involved in laughing and crying. Evidence suggests certain basic emotional expressions like these do not have to be learned. In comparison, voluntary facial movements such as those involved in speech and culture-specific expressions are learned through experience and predominantly rely on cortical motor control. The Auckland team's psychobiological facial framework aims to reflect that facial expressions consist of both innate and learned elements and are driven by quite independent brain-region simulations.

Mark Sagar believes that to drive a biologically based, life-like autonomous character, one needs to model multiple aspects of a nervous system. Such as including models of the sensory and motor systems, reflexes, perception, emotion and modulatory systems, attention, learning and memory, rewards, and goal or decision making systems. So the BabyX project sort to define an architecture that is able to interconnect all of these models as a virtual nervous system.



Unlike a robotic system where so much of the effort is focused on mechanical limitations of being able to replicate the complexity of the face, the BabyX approach is to building a biologically based model of behavior, with particular emphasis on the importance of rapid face-to-face interaction. Such immediate interaction is difficult to achieve in robotics due to mechanical constraints. BabyX can be reduced to more biological level, as well as being expandable to allow adding newer higher-level complex systems, if and when they are developed, in a Lego-like manner.

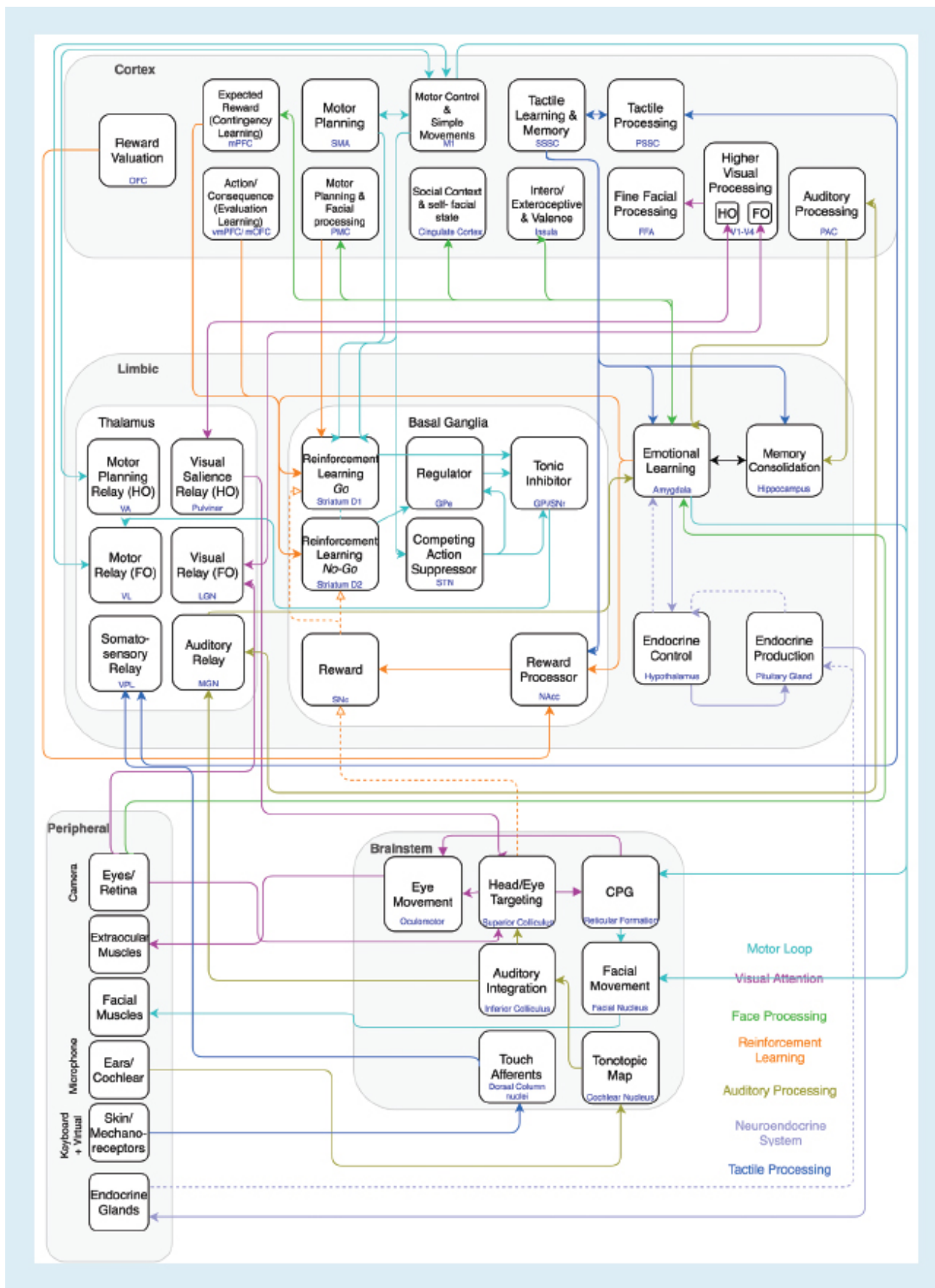
BL is the key to this modular simulation framework and was developed over five years to integrate neural networks with real-time computer graphics and sensing. It is designed for maximum flexibility and can be connected through a simple API. It consists of a library of modules and it is designed to support a wide range of computational neuroscience models. As stated in the CACM article, *“the models supported by BL range from simple leaky integrators to spiking neurons to mean field models to self-organizing maps. These can be interconnected to form larger neural networks (such as recurrent networks and convolutional networks like those used in deep learning)”*.

A huge aspect of BabyX is the notion of a complete two way emotional interaction, referred to in computer circles as Affective Computing. To achieve this, BabyX has to sense those talking to her. Sensory input is typically through camera, microphone, and keyboard to enable computer vision audition and “touch” processing, but data can be input from any arbitrary sensor. BabyX’s face is output through OpenGL and the

OpenGL Shading Language. But much like the AI subsystem, the neural network system could drive any sophisticated 3D animation system.

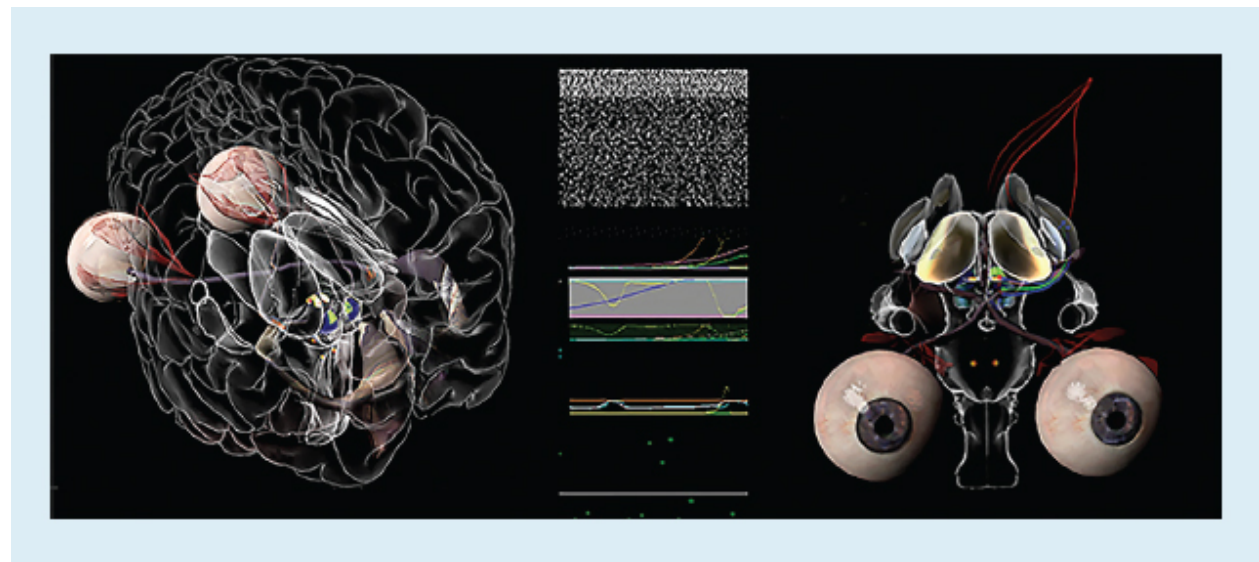
Nervous system

BabyX's biologically inspired nervous system consists of an interconnected set of neural system and subsystem models. The models span the neuroaxis and generate muscle-activation-based animation from a continuously evaluated neural network.



An architectural diagram above shows some of the key functional components, neuroanatomical structures, and functional loops. A characteristic of this approach is the representation of subcortical structures (such as the basal ganglia) and brainstem nuclei (such as the oculomotor nuclei). The structures are functionally implemented as neural network models with particular characteristics for example activation of the hypothalamus releases ‘virtual hormones’. When working with BabyX it is possible to enter a mode and play directly with her virtual brain chemistry, thus simulating a mother who might have been on drugs, or a disease such as Parkinson’s. Virtual dopamine, for example, plays a key role in motor activity and reinforcement learning. It can also modulate plasticity in the neural networks and have subtle behavioral effects such as pupil dilation and blink rate.

BabyX’s emotional states affect the sensitivity of behavioral circuits. For example, stress lowers the threshold for triggering a brainstem central pattern generator that, in turn, generates the motor pattern of facial muscles in crying. Stress BabyX by leaving her alone and she might cry.

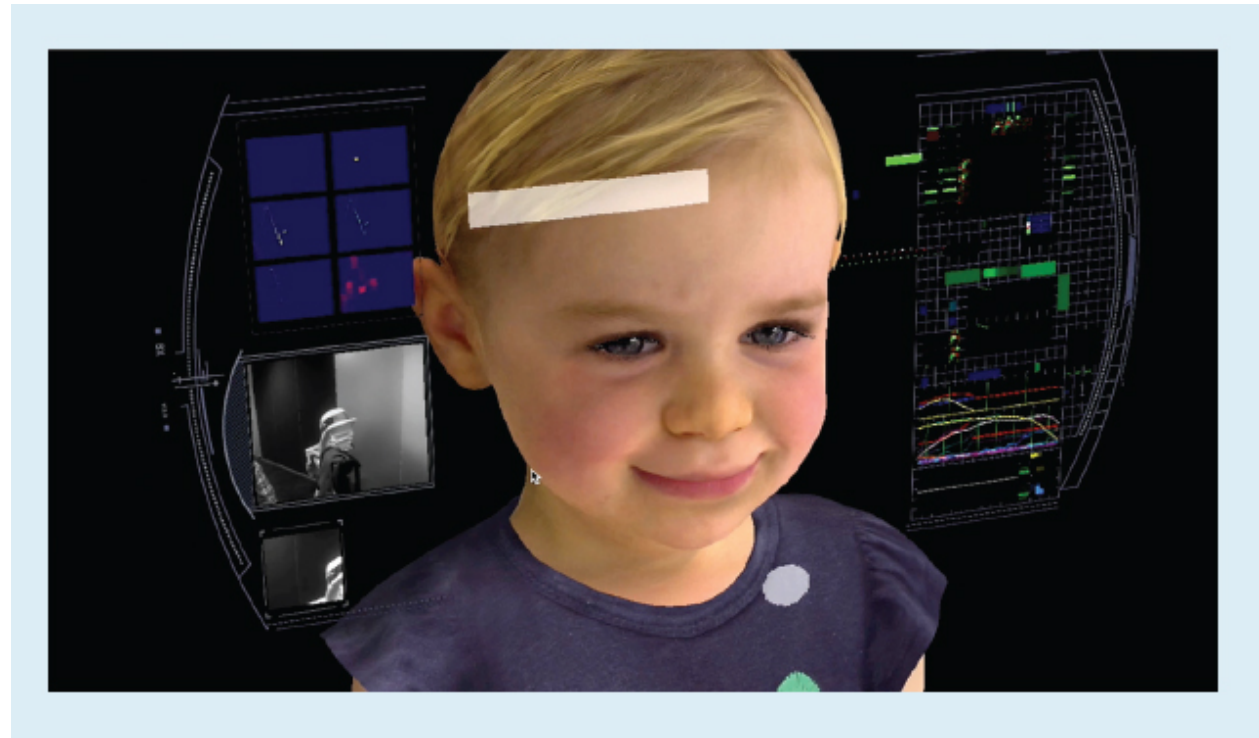


All of this ‘under the hood’ simulated brain is presented in a graphical brain with each function in the right geographical or anatomically correct, parts of the brain.

Learning

An example showing learning through interaction with the environment (known as autonomous action discovery) was to have BabyX learn to play the classic video game Pong. Mark and the team connected motor neurons in BabyX to the bat controls and overlaid the visual output of the game on the camera’s input. Motor babbling causes the virtual infant to inadvertently move the bat, much like a baby might flail its

arms about. Trajectories of the ball are learned as patterns on neural network maps. If the bat hits the ball, a rewarding reaction results, reinforcing the association. This further resulted in the bat being moved in anticipation of where the ball is going. Without specialist programming or explanation, but with the releasing of 'virtual dopamine' when she hit the ball, BabyX taught herself to control the bat and play Pong.



Teaching herself Pong (bat at top of screen)

Effect on people

At Siggraph 2015 Mark Sagar demonstrated the current version of BabyX. The educated and informed hall full of graphics specialist were shown how she works and watched as Mark interacted with BabyX on stage, as he has done in similar seminar around the world. And just as in those other instances, the audience was shown BabyX reading simple words and getting upset when Mark left her alone (by leaving her field of view) but also they saw 'under the hood' just how BabyX was designed and modeled.

While the audience was informed professionals, their reaction was audible and visceral to BabyX. People get 'caught up in the moment'. Significantly, their

responses turn sharp negative reaction when the Mark offered to demonstrate the pain response, as if someone was about to “hurt” the baby.

There was no sense that the audience rationally thought the baby was real, but they reacted immediately as if she were. Modeling the pain response is a completely sensible part of any virtual human at this level, and yet the instant response was one bordering on horror that Mark would ‘hurt’ BabyX. Even within a formal academic presentation, and with the pain response a valid part of any brain model, the audience reacted as if Mark was about to be cruel.

Interestingly, this was followed by an emotional display of relief in the form of laughter. A split second after the audience gasps in horror, it acknowledges how visceral this response was and laughs. As laughter is infectious, Mark laughed, which was registered by BabyX’s sensory inputs, causing her to be “happier” and she smiles. Which only feeds the audience’s collective amusement as they watch BabyX laugh about a joke at her expense. The audience thus became a part of the feedback loop that changed both parties’ emotional states. The implication is that a witness to a BabyX session is becoming a part of the holistic environment and the interactive experience. There may be ethical implications as well, and further research is needed to investigate the co-defined dynamic interaction that allows such strong “in the moment” emotional responses, as such responses may have long-term interface implications.

Moving Forward

BabyX exists just in the computer but she could be given input sources to explore her environment. These do not need to be limited to just robotic arms. The Lab explored giving her standalone mini robots that could explore and map a room she is in and there is no reason why she could not have a dozen or more of these all inputting into her ‘seeing’ the room and mapping it.

Much work has also been done into her ability to speak. BabyX babbles with a synthesized voice sampled from phonemes produced by a real child. The team have been implementing techniques so BabyX can learn an acoustic mapping from any arbitrary voice to construct new words using her own voice. This could lead to BabyX being capable of learning arbitrary lip sequences, and allow BabyX to construct her own sentences.

Finally, there is no reason BabyX needs to be a Baby. As part of the wider problem of Agents, the team has developed the [Auckland Face Simulator](#) which we have covered

in the past here at fxguide. This is an engine to allow for a much faster production of high quality faces, removing much of the manual time consuming artist repetitive work to produce any reasonable adult face.



Auckland Face Simulator

While the popular press are quick to jump to science fiction robo-apocalypse themes, BabyX and her future successors offer a tremendous opportunity to expand the nature of a computer interface, providing the aged and restricted with much more flexible and perhaps appealing interaction modes. The opportunities for both learning and medical simulation training are great, as they are in the area of remote learning. BabyX is a landmark piece of serious research that aims to further these UI goals but in the process she is also helping us better understand human development cognition and what drives facial acceptance in the human mind. While we learn how to make her, we are learning a lot about ourselves.

fxphd production blog

[NEW COURSE: ADVANCED LIGHTING PRACTICES FOR MAYA & ARNOLD](#)



Philip Maddock leads a new course exploring production-level shot lighting techniques to create beautiful, realistic and faithful lighting.

[FIND OUT MORE](#)

get the fxguide app



Related Posts

- [Conversational Agents with Cate Blanchett: the...](#)

Get email updates whenever new articles are posted

[Sign up](#)

