BRINGING VIRTUAL HUMANS TO LIFE

For Dr. Zerrin Yumak, a computer scientist specializing in socially interactive virtual humans at Utrecht University, there's a missing piece in today's crop of digital humans. "I think the most urgent question right now is how we can generate animations for non-verbal communication including with the face and body, but also holistically," she says. "Multimodal generation of social animations is still missing and it's a challenge that we need to address."

Dr. Zerrin Yumak, Assistant Professor, of the Information and Computing Sciences Department, Utrecht University and Director of the Motion Capture and Virtual Reality Lab

Dr. Yumak is an assistant professor in the Information and Computing Sciences Department at Utrecht University and director of the Motion Capture and Virtual Reality Lab. A member of the Human Centered Computing group, she's been working in the field of 3D digital humans for the last 15 years.

The field covers avatars for humans in virtual spaces, but also digital characters that could be used in video games or to embody chatbots or dialogue systems such as ChatGPT or IBM Watson.

"I'm working on how to make these characters interact naturally with us using natural ways of communication such as facial expressions, gestures and gaze behavior," Dr. Yumak explains. "Lately, I've mostly been working on speech- and music-driven animation. Given speech or music as input, we are automatically generating motion. We are aiming for natural and controllable motion synthesis and to automate these processes for faster and less costly pipelines.

"Our recent work on FaceXHubert and FaceDiffuser aims for generating facial motion using advanced deep learning algorithms both for 3D vertex-based and rigged characters. We have worked on a project in collaboration with Guerilla Games to generate gaze motion and our results show the advantage of a data-driven approach with respect to a procedural gaze model."

These behaviors, whether applied to human avatars, non-player characters in video games or a digital character acting as an interface with a chatbot, are informed by motion capture data.

MAKING DIGITAL HUMANS WELL BEHAVED

Dr. Yumak says that believable, natural animation is an important next step in virtual character development. "We are in a situation where we have a lot of developments in the fields of computer vision/graphics, AI and natural language processing, and VR/AR technologies," she says.

"All of these technologies are now merging, which really lets us develop very realistic virtual characters, virtual environments and dialogue capabilities, but there is more work to be done to generate social behaviors that are correct and convincing. That is a very complex and interdisciplinary research field." She organized the MASSXR (Multi-modal Affective and Social Behavior Analysis and Synthesis in Extended Reality) Workshop at the IEEE VR 2023 conference to bring together researchers and practitioners in this field.

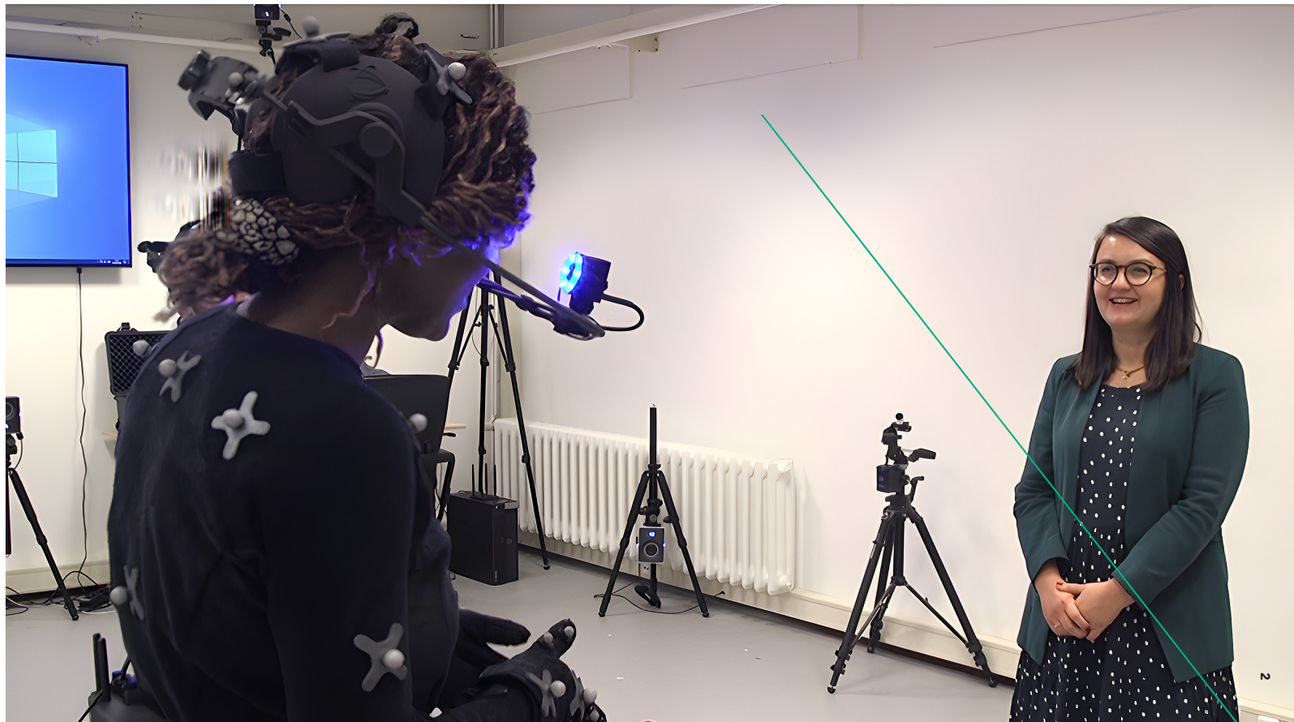
Perhaps the most obvious application for Dr. Yumak's work is in the gaming and immersive tech industry, but there are also opportunities in other fields. "For example, the use of digital humans to deliver training in communication skills, or for business and marketing. For instance, you can have virtual characters as chatbots to advise people on things like mortgages. And there are use cases in telecommunications using Social XR, enabling remote people to interact and work together. Virtual human animations play a crucial role in creating a sense of presence and trust in these applications.

"In health and education, we can have virtual characters or companions that can help children to learn or interact with elderly people. I have worked, for example, on a project where we were developing companions that can play instruments like piano for children with motor disabilities, or a robotic tutor enabled with memory and the appearance of emotions."

Dr. Yumak has been working with Vicon systems dating back to Blade, the precursor to Vicon's Shōgun software platform for digital creators. The lab at Utrecht University consists of 14 Vantage cameras, now running alongside Shōgun. The software, in particular, has done a lot to streamline her work. "Everything is just great out of the box, and the solving and everything works nicely, especially finger-tracking, also tracking multiple people all at the same time. The nice thing is you can have a solver that combines body solving and finger solving together," she says.

"These elements are very important for our research because we're collecting data in the lab and then using this data to train our deep learning models to generate the movement of the virtual characters, and also to conduct perceptual user experiments. Being able to do that fast is very useful for us.

"I want to capture these small group conversations in great detail, at micro-gesture levels. Tracking is important for us to capture all the nuances in the face, the finger movements and the body movements. We have collected a new dataset called Utrecht University Dyadic Multimodal Motion Capture Dataset–10 hours of natural conversations between two actors, including detailed body movements, facial animation and finger movements. That is a crucial step for us towards multi-modal motion synthesis in small group-interactions."



A FRESH APPROACH TO AI AND XR

Dr. Yumak is in the process of putting together a group that will tie some of the disparate strands of the AI and extended reality (XR) fields together. "Currently, we're setting up a new lab bringing together experts and researchers but also public and private partners, and this is called the Embodied AI Lab for Social Good. We want to bring AI and immersive technologies together, and the concept is about embodiment and artificial intelligence. We want to collaborate with technology companies that develop animations, but also with educational or health organizations that use these virtual characters to improve human/machine communication."

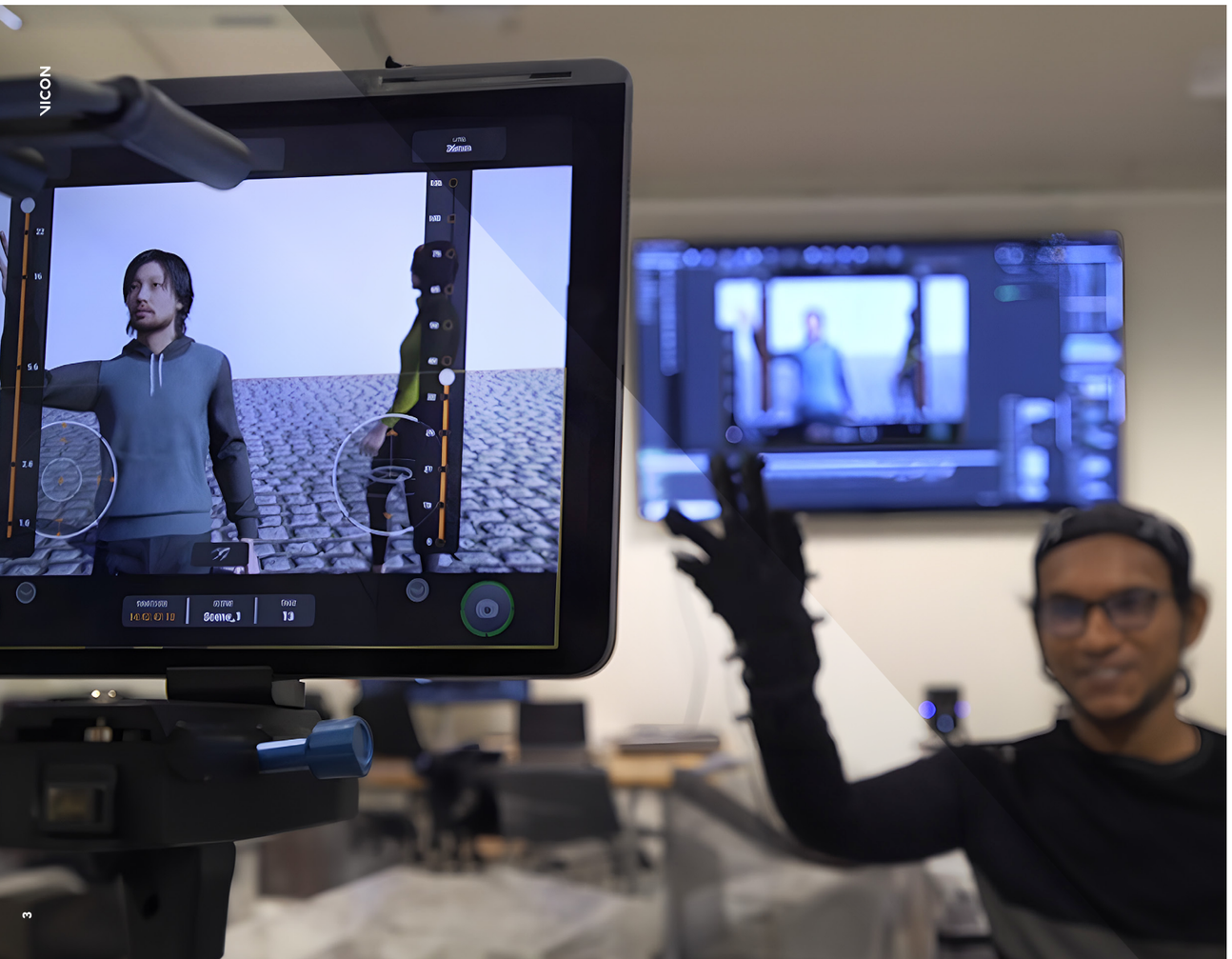
The new lab has its sights set on a number of challenges. "One of the things that we're working on is speech-driven gesture motion synthesis. At the moment this is mostly done using datasets and algorithms that do not include semantic aspects of gestures. In other words, the motions look natural at first glance, but their semantic grounding with respect to accompanying text is lacking. In collaboration with colleagues from Max Planck Institute for Psycholinguistics, we are addressing this challenge"

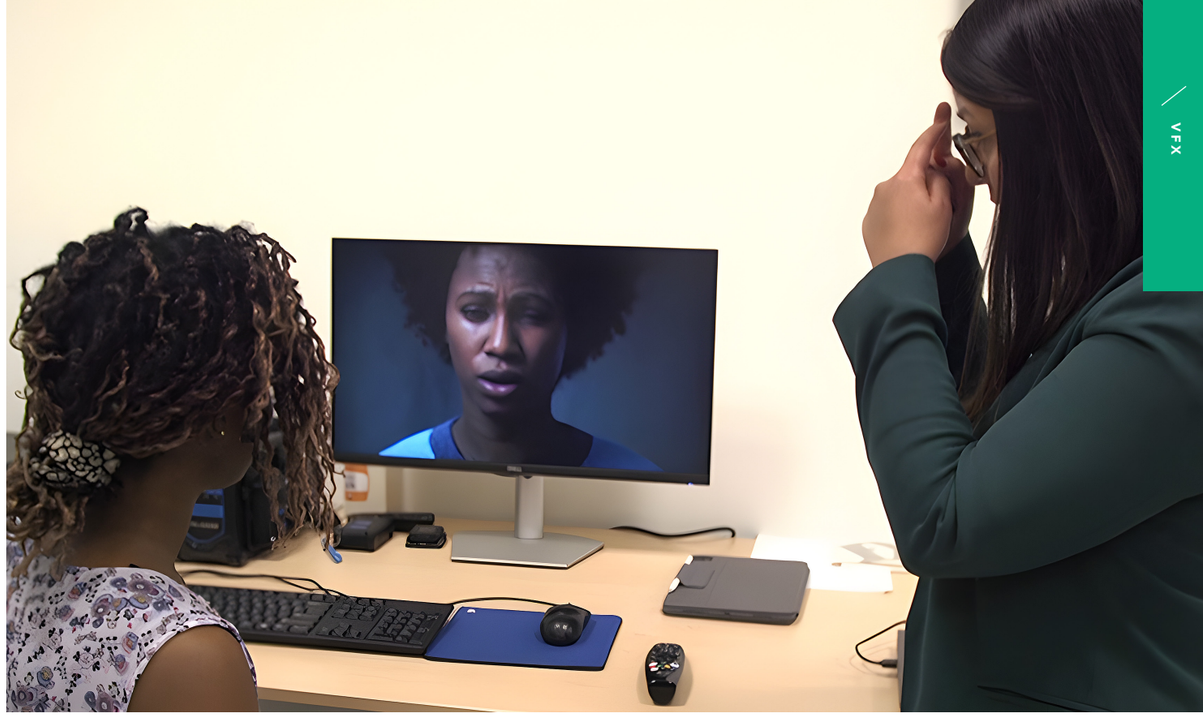
Audio-driven and text-driven gesture generation at the micro level remains a priority, but applying those behaviors to group interaction between characters is increasingly a focus. "At the moment, the current techniques are really directed at monologues," says Dr. Yumak. "You deliver the audio and then you get a character that's talking by itself. But what happens if we are having a two or three-party conversation? Who looks at who, who takes the speaking turn? How do you manage this dynamic?

"Can we automate this process? The listening behavior, talking behavior, but also the style control: Can we create a character that is automatically moving given a certain emotional or personality style?"

Yet another challenge is bringing these behaviors to multi-party interactions between humans and virtual characters. "Can we feed these (facial and body language) signals from the real users to the behavior of the virtual characters? Because everyone is focusing on the language or vision half of the pipeline. The other half, the physical and embodiment, is not really well-studied yet," says Dr. Yumak.

There are applications for the lab's work beyond the digital sphere, too. "The techniques that are relevant for socially interactive digital humans can also be applied to robotic characters. They include very similar pipelines, they use similar algorithms and datasets, but the medium, the embodiment, is different," says Dr. Yumak.



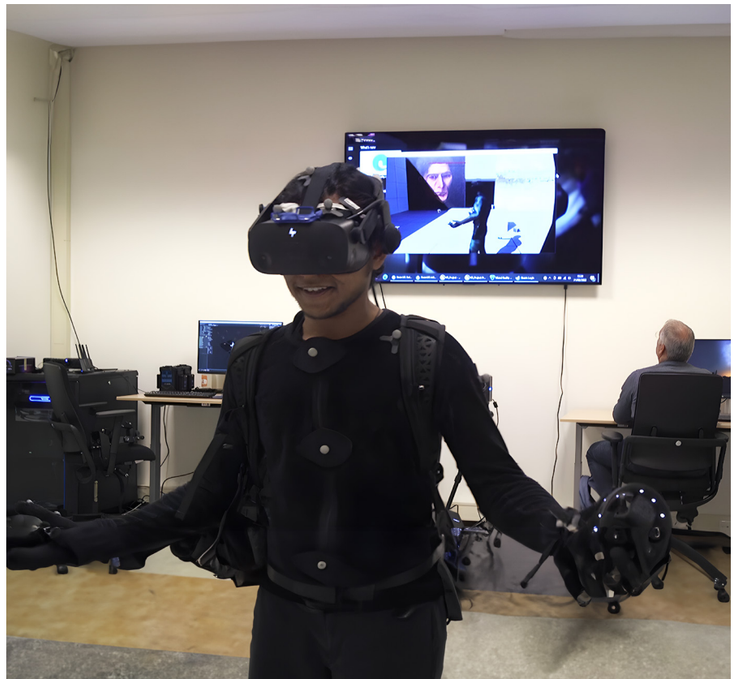


THE CORONAVIRUS EFFECT

Covid did a lot to accelerate Dr. Yumak's work. "I think after the pandemic, we really started to talk about new ways of communication using 3D tools–immersive spaces, virtual reality, augmented reality and spatial computing. That has definitely had a very big impact on my research too," she says.

Dr. Yumak notes that studies have shown 2D video conferencing can create fatigue due to issues like failure of eye contact and a lack of appropriate non-verbal communication, and also notes the environmental impact of reducing travel for meetings.

Thanks to this confluence of technological and social factors, research into immersive telepresence is growing. "We are coming together with a consortium of people to work on the topic of social XR, working on how realistic virtual humans can be used for next generation telecommunication where we can represent ourselves with avatars. It can be really useful, instead of having these 2D screens, to be able to meet in virtual spaces and then interact in a more natural way," Dr. Yumak says.



She's sensitive to the fact that this isn't only a question for engineers and designers. "This is not only about the technology development but how people are perceiving this, whether we can include different perspectives in this kind of discussion when we are developing intelligent behavior in immersive environments.

"So, do we really want to create life-like digital characters or even replicas? To what extent can we ensure privacy and security? We often try to engage in these conversations with people from different perspectives, with backgrounds in humanities and social sciences, and even from philosophy and law. At the end of the day, we should see these technologies as tools to enhance people's quality of life, and not as a replacement for real human communication, and we need to clearly communicate the advantages and potential risks."

For more information on Utrecht University's Motion Capture and Virtual Reality Lab, visit: ***www.uu.nl/en/research/motion-capture-and-virtual-reality-lab***