NEWS RELEASE

Singapore, 16 November 2023

Realistic talking faces created from only an audio clip and a person's photo using NTU Singapore computer program

A team of researchers from **Nanyang Technological University, Singapore (NTU Singapore)** has developed a computer program that creates realistic videos that reflect the facial expressions and head movements of the person speaking, only requiring an audio clip and a face photo.

DIverse yet **R**ealistic **F**acial **A**nimations, or **DIRFA**, is an artificial intelligence-based program that takes audio and a photo and produces a 3D video showing the person demonstrating realistic and consistent facial animations synchronised with the spoken audio (see videos).

The NTU-developed program improves on existing approaches (see Figure 1), which struggle with pose variations and emotional control.

To accomplish this, the team trained DIRFA on over one million audiovisual clips from over 6,000 people derived from an open-source database called The VoxCeleb2 Dataset to predict cues from speech and associate them with facial expressions and head movements.

The researchers said DIRFA could lead to new applications across various industries and domains, including healthcare, as it could enable more sophisticated and realistic virtual assistants and chatbots, improving user experiences. It could also serve as a powerful tool for individuals with speech or facial disabilities, helping them to convey their thoughts and emotions through expressive avatars or digital representations, enhancing their ability to communicate.

Corresponding author Associate Professor Lu Shijian, from the School of Computer Science and Engineering (SCSE) at NTU Singapore, who led the study, said: "The impact of our study could be profound and far-reaching, as it revolutionises the realm of multimedia communication by enabling the creation of highly realistic videos of individuals speaking, combining techniques such as AI and machine learning.

Our program also builds on previous studies and represents an advancement in the technology, as videos created with our program are complete with accurate lip movements, vivid facial expressions and natural head poses, using only their audio recordings and static images."

First author Dr Wu Rongliang, a PhD graduate from NTU's SCSE, said: "Speech exhibits a multitude of variations. Individuals pronounce the same words differently in diverse contexts, encompassing variations in duration, amplitude, tone, and more. Furthermore, beyond its linguistic content, speech conveys rich information about the speaker's emotional state and identity factors such as gender, age, ethnicity, and even personality traits. Our approach represents a pioneering effort in enhancing performance from the perspective of audio representation learning in AI and machine learning." Dr Wu is a Research Scientist at the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore.

The findings were published in the scientific journal *Pattern Recognition* in August.

Speaking volumes: Turning audio into action with animated accuracy

The researchers say that creating lifelike facial expressions driven by audio poses a complex challenge. For a given audio signal, there can be numerous possible facial expressions that would make sense, and these possibilities can multiply when dealing with a sequence of audio signals over time.

Since audio typically has strong associations with lip movements but weaker connections with facial expressions and head positions, the team aimed to create talking faces that exhibit precise lip synchronisation, rich facial expressions, and natural head movements corresponding to the provided audio.

To address this, the team first designed their AI model, DIRFA, to capture the intricate relationships between audio signals and facial animations. The team trained their model on more than one million audio and video clips of over 6,000 people, derived from a publicly available database.

Assoc Prof Lu added: "Specifically, DIRFA modelled the likelihood of a facial animation, such as a raised eyebrow or wrinkled nose, based on the input audio. This modelling enabled the program to transform the audio input into diverse yet highly lifelike sequences of facial animations to guide the generation of talking faces."

Dr Wu added: "Extensive experiments show that DIRFA can generate talking faces with accurate lip movements, vivid facial expressions and natural head poses. However, we are working to improve the program's interface, allowing certain outputs

to be controlled. For example, DIRFA does not allow users to adjust a certain expression, such as changing a frown to a smile."

Besides adding more options and improvements to DIRFA's interface, the NTU researchers will be finetuning its facial expressions with a wider range of datasets that include more varied facial expressions and voice audio clips.

###

Notes to Editor:

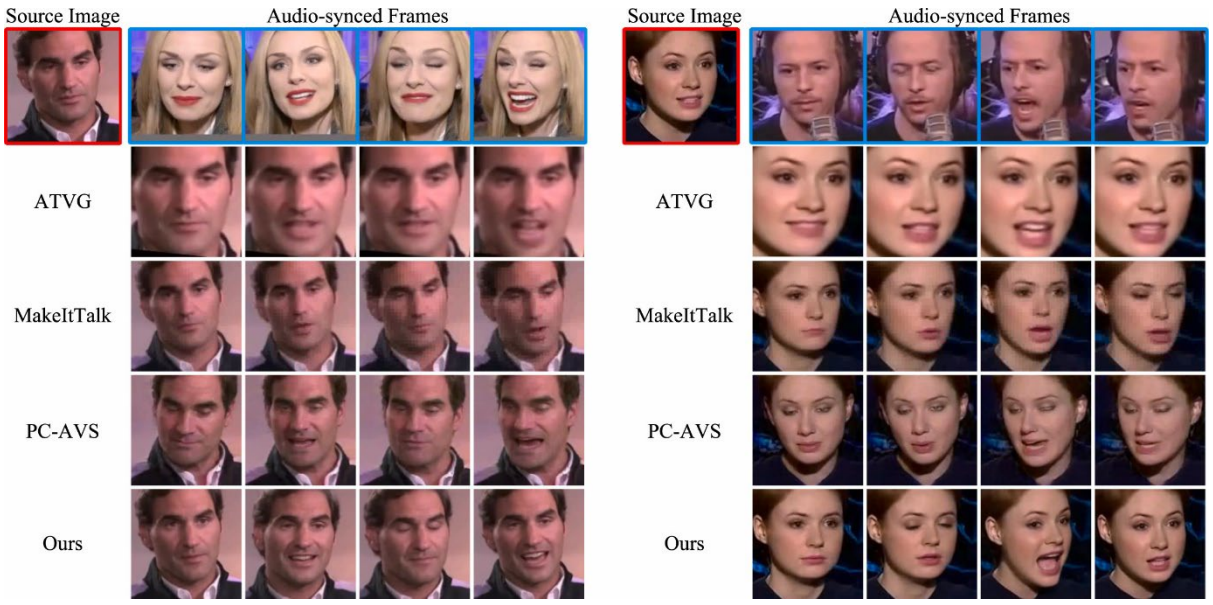

Figure 1: Comparisons of DIRFA with state-of-the-art audio-driven talking face generation approaches

Explainer video: How DIRFA uses artificial intelligence to generate 'talking heads'

Video 2: A DIRFA-generated 'talking head' with just an audio of former US president Barrack Obama speaking, and a photo of Associate Professor Lu Shijian.

Video 3: A DIRFA-generated 'talking head' with just an audio of former US president Barrack Obama speaking, and a photo of study's first author Dr Wu Rongliang.

The research paper titled: "*Audio-driven talking face generation with diverse yet realistic facial animations*" *was* published in *Pattern Recognition* on 16 Aug 2023. DOI 10.1016/j.patcog.2023.109865

*** END ***

Media contact:

Mr Joseph Gan
Manager, Media Relations
Corporate Communications Office
Nanyang Technological University, Singapore
Email: joseph.gan@ntu.edu.sg

About Nanyang Technological University, Singapore

A research-intensive public university, Nanyang Technological University, Singapore (NTU Singapore) has 33,000 undergraduate and postgraduate students in the Engineering, Business, Science, Medicine, Humanities, Arts, & Social Sciences, and Graduate colleges.

NTU is also home to world-renowned autonomous institutes – the National Institute of Education, S Rajaratnam School of International Studies and Singapore Centre for Environmental Life Sciences Engineering – and various leading research centres such as the Earth Observatory of Singapore, Nanyang Environment & Water Research Institute and Energy Research Institute @ NTU (ERI@N).

Under the NTU Smart Campus vision, the University harnesses the power of digital technology and tech-enabled solutions to support better learning and living experiences, the discovery of new knowledge, and the sustainability of resources.

Ranked amongst the world's top universities, the University's main campus is also frequently listed among the world's most beautiful. Known for its sustainability, NTU has achieved 100% Green Mark Platinum certification for all its eligible building projects. Apart from its main campus, NTU also has a medical campus in Novena, Singapore's healthcare district.

For more information, visit www.ntu.edu.sg.