

New NVIDIA Maxine Cloud-Native Architecture Delivers Breakthrough Audio and Video Quality at Scale

Early-access AI microservices deliver premium-quality communications in the cloud.

Author: Delilah Liu

The latest release of NVIDIA Maxine is paving the way for real-time audio and video communications. Whether for a video conference, a call made to a customer service center, or a live stream, Maxine enables clear communications to enhance virtual interactions.

NVIDIA Maxine is a suite of GPU-accelerated AI software development kits (SDKs) and cloud-native microservices for deploying optimized and accelerated AI features that enhance audio, video and augmented-reality (AR) effects in real time.

And with Maxine's state-of-the-art models, end users don't need expensive gear to improve audio and video. Using NVIDIA AI-based technology, these high-quality effects can be achieved with standard microphones and camera equipment.

At GTC , NVIDIA announced the re-architecture of Maxine for cloud-native microservices, with the early-access release of Maxine's audio-effects microservice. Additionally, new Maxine SDK features were unveiled, including Speaker Focus and Face Expression Estimation, as well as the general availability of Eye Contact. NVIDIA Maxine now also includes enhanced versions of existing SDK features.

Maxine's cloud-native microservices allow developers to build real-time AI applications. Microservices can be independently managed and deployed seamlessly in the cloud, accelerating development timelines.

The Audio Effects microservice, available in early access, contains four state-of-the-art audio features:

Background Noise Removal: Removes several common background noises using AI models, while preserving the speaker's natural voice.

Room Echo Removal : Removes reverberations from audio using AI models, restoring clarity of a speaker's voice.

Audio Super Resolution: Improves audio quality by increasing the temporal resolution of audio signal. It currently supports upsampling from 8 kHz to 16 kHz and from 16 kHz to 48 kHz.

Acoustic Echo Cancellation: Cancels real-time acoustic device echo from the input-audio stream, eliminating mismatched acoustic pairs and double-talk. With AI-based technology, more effective cancellation is achieved than with traditional digital signal processing.

Pexip , a leading provider of enterprise video conferencing and collaboration solutions, is using NVIDIA AI technologies to take virtual meetings to the next level with advanced features for the modern workforce.

"With Maxine's move to cloud-native microservices, it will be even easier to combine NVIDIA's advanced AI technologies with our own unique server-side architecture," said Eddie Clifton, senior vice president of Strategic Alliances at Pexip. "This allows our teams at Pexip to deliver an enhanced experience for virtual meetings."

Sign up for early access .

Maxine offers three GPU-accelerated SDKs that reinvent real-time communications with AI: audio, video and AR effects.

The audio effects SDK delivers multi-effect, low-latency, AI-based audio-quality enhancement algorithms. Speaker Focus , available in early access, is a new feature that separates the audio tracks of foreground and background speakers, making each voice more intelligible. Additionally, the Audio Super Resolution SDK feature has been updated with enhanced quality.

The video effects SDK creates AI-based video effects with standard webcam input. The Virtual Background feature, which segments a person's profile and applies AI-powered background removal, replacement or blur, has been updated with enhanced temporal stability.

And the AR SDK provides AI-powered, real-time 3D face tracking and body pose estimation based on a standard web camera feed. Latest features include:

Eye Contact: Simulates eye contact by estimating and aligning gaze with the camera.

Face Expression Estimation : Tracks the face and infers what expression is presented by the subject.

The following AR features have been updated:

Body Pose Estimation : Predicts and tracks 34 key points of the human body in 2D and 3D — now with support for multi-person tracking.

Face Landmark Tracking : Recognizes facial features and contours using 126 key points. Tracks head pose and facial deformation due to head movement and expression — in three degrees of freedom in real time — now with Quality mode to achieve even higher-quality tracking.

Face Mesh : Represents a human face with a 3D mesh with up to 3,000 vertices and six degrees of freedom — now includes 3D morphable models from the USC Institute of Creative Technologies.

Try out the Maxine SDKs . To directly experience Maxine's effects, download the NVIDIA Broadcast App .

Maxine SDKs and microservices provide a suite of low-latency AI effects that can be integrated with existing customer infrastructures. Developers can tap into cutting-edge AI capabilities with Maxine, as the technology is built on the NVIDIA AI platform and has world-class pretrained models for users to create, customize and deploy premium audio- and video-quality features.

Maxine is also part of the NVIDIA Omniverse Avatar Cloud Engine , a collection of cloud-based AI models and services for developers to build, customize and deploy interactive avatars. Maxine's customizable cloud-native microservices allow for independent deployment into AI-effects pipelines. Maxine can be deployed on premises, in the cloud or at the edge.

Learn more about NVIDIA Maxine and other technology breakthroughs by watching the GTC keynote by NVIDIA founder and CEO Jensen Huang:

Original URL: <https://blogs.nvidia.com/blog/2022/09/20/maxine-cloud-native/>