Publicado en Ciudad de México el 01/09/2021

# [Todas las Sensaciones: NVIDIA comparte la Investigación de Síntesis de Voz Expresiva en Interspeech](http://www.notasdeprensa.es)

## NVIDIA permite a los desarrolladores y creadores de IA conversacional con modelos de vanguardia para la síntesis de voz expresiva que pueden expresar personajes, asistentes virtuales y avatares personalizados.

 La IA ha transformado el habla sintetizada del tono monótono de las llamadas automáticas y los primeros sistemas de navegación GPS al tono profesional y fresco de los asistentes virtuales en los smartphones y los altavoces. Pero todavía existe una brecha entre el habla sintetizada por IA y el habla humana que se escucha en la conversación diaria y en los medios de comunicación. La gente habla con un ritmo, entonación y timbre complejos que es un desafío para la IA emular. Para ayudar a los desarrolladores y creadores, los investigadores de NVIDIA están creando modelos y herramientas para una síntesis de voz controlable y de alta calidad que captura la riqueza del habla humana, sin artefactos de audio. Estos modelos pueden ayudar a las líneas de servicio al cliente automatizadas por voz para bancos y minoristas, dar vida a personajes de videogames o libros y proporcionar síntesis de voz en tiempo real para avatares digitales. El equipo creativo interno de NVIDIA utiliza la tecnología para producir una narración emotiva para una serie de videos sobre el poder de la IA. La síntesis de voz expresiva es solo un elemento del trabajo del Area de Investigación de NVIDIA en IA conversacional, un campo que también abarca el procesamiento del lenguaje natural, el reconocimiento automático de voz, la detección de palabras clave, la mejora de audio y más. Optimizadas para ejecutarse de manera eficiente en las GPUs NVIDIA, las herramientas de inteligencia artificial conversacionales de última generación de NVIDIA se lanzan en el kit de herramientas de código abierto NVIDIA NeMo, disponible en NGC. Algunos de los últimos proyectos de síntesis de voz y reconocimiento de voz de NVIDIA Research se exhibirán en la conferencia INTERSPEECH 2021, que se extenderá hasta el 3 de septiembre. Detrás de las Escenas de: I AM AI.Los investigadores y profesionales creativos de NVIDIA no se limitan a hablar sobre la IA conversacional, sino que hacen lo propio y ponen en práctica modelos innovadores de síntesis de voz en la serie de vídeos I AM AI. Estos videos destacan las vastas aplicaciones de la IA en todas las industrias, pero hasta hace poco, fueron siempre narrados por un ser humano. Los modelos de síntesis de voz existentes ofrecían un control limitado sobre el ritmo y el tono de una voz sintetizada, por lo que los intentos de narración de IA no evocaban la respuesta emocional en los espectadores que un locutor humano talentoso podía lograr. Eso cambió el año pasado, cuando el equipo de investigación de texto a voz desarrolló modelos de síntesis de voz más potentes y controlables como RAD-TTS, que se utiliza en la demostración ganadora de NVIDIA en la competencia SIGGRAPH Real-Time Live. Al alimentar el modelo de texto a voz entre 5 y 30 minutos de audio de entrenamiento del habla de un individuo, la IA puede convertir cualquier mensaje de texto en la voz del hablante. Otra de sus funciones es la conversión de voz, en la que las palabras de un hablante (o incluso el canto) se transmiten con la voz de otro hablante. Inspirada en la idea de la voz humana como instrumento musical, la interfaz RAD-TTS brinda a los usuarios un control detallado a nivel de cuadro sobre el tono, la duración y la energía de la voz sintetizada. Con esta interfaz, el productor de video de NVIDIA podría grabarse a sí mismo leyendo el guión del video y luego usar el modelo de IA para convertir su discurso en la voz de la narradora. Usando esta narración de línea de base, el productor podría dirigir la IA como un actor de voz, ajustando el discurso sintetizado para enfatizar palabras específicas y modificando el ritmo de la narración para expresar mejor el tono del video. Las capacidades del modelo de inteligencia artificial van más allá del trabajo de voz en off: la conversión de texto a voz se puede utilizar en games, para ayudar a las personas con discapacidades vocales o incluso para ayudar a los usuarios a traducir entre idiomas con su propia voz. Incluso puede recrear las actuaciones de cantantes icónicos, haciendo coincidir no solo la melodía de una canción, sino también la expresión emocional detrás de las voces. Dar Voz a los Desarrolladores e Investigadores de IACon NVIDIA NeMo, un kit de herramientas de Python de código abierto para IA conversacional acelerada por GPU, los investigadores, desarrolladores y creadores obtienen una ventaja en la experimentación y el ajuste de modelos de voz para sus propias aplicaciones. Las API fáciles de usar y los modelos previamente entrenados en NeMo ayudan a los investigadores a desarrollar y personalizar modelos para conversión de texto a voz, procesamiento de lenguaje natural y reconocimiento de voz automatizado en tiempo real. Varios de los modelos se entrenan durante más de 100,000 horas en sistemas NVIDIA DGX en conjuntos de datos de código abierto y se pueden ajustar para el caso de uso de un desarrollador mediante la computación de precisión mixta en las GPU NVIDIA Tensor Core. NVIDIA NeMo también ofrece modelos entrenados en Mozilla Common Voice a través de NGC, un conjunto de datos con casi 14,000 horas de datos de voz de origen colectivo en 76 idiomas. Con el apoyo de NVIDIA, el proyecto tiene como objetivo democratizar la tecnología de voz con el conjunto de datos de voz de datos abiertos más grande del mundo. Voice Box: Los Investigadores de NVIDIA Desempaquetan el Discurso de la IAINTERSPEECH reúne a más de 1,000 investigadores para mostrar trabajos pioneros en tecnología del habla. En la conferencia de esta semana, el Área de Investigación de NVIDIA presenta arquitecturas de modelos de IA conversacional, así como conjuntos de datos de voz totalmente formateados para desarrolladores. Ver las siguientes sesiones dirigidas por oradores de NVIDIA: ● Scene-Agnostic Multi-Microphone Speech Dereverberation — Martes, 31 de Agosto ● SPGISpeech: 5,000 Hours of Transcribed Financial Audio for Fully Formatted End-to-End Speech Recognition — Miércoles – 1ª de Septiembre ● Hi-Fi Multi-Speaker English TTS Dataset — Miércoles – 1ª de Septiembre ● TalkNet 2: Non-Autoregressive Depth-Wise Separable Convolutional Model for Speech Synthesis with Explicit Pitch and Duration Prediction — Jueves, 2 de Septiembre ● Compressing 1D Time-Channel Separable Convolutions Using Sparse Random Ternary Matrices — Viernes, 3 de Septiembre ● NeMo Inverse Text Normalization: From Development To Production — Viernes, 3 de Septiembre Encuentre modelos de NVIDIA NeMo en el catálogo de NGC y sintonice las charlas de los investigadores de NVIDIA en INTERSPEECH. By Isha Salian Información sobre NVIDIALa invención de la GPU por parte de NVIDIA (NASDAQ: NVDA) en 1999 generó el crecimiento del mercado de juegos para PC, redefinió los gráficos por computación modernos, la computación de alto rendimiento y la inteligencia artificial. El trabajo pionero de la compañía en computación acelerada e inteligencia artificial está remodelando industrias de billones de dólares, como el transporte, la atención médica y la fabricación, y está impulsando el crecimiento de muchas otras. Obtén más información en https://nvidianews.nvidia.com/.

**Datos de contacto:**

Carlos Valencia

MKQ PR Agency

55 39 64 96 00

Nota de prensa publicada en: [https://www.mexicopress.com.mx/todas-las-sensaciones-nvidia-comparte-la](http://www.notasdeprensa.es/educalivecom-disfruta-de-cursos-presenciales-y-clases-particulares-a-traves-de-internet)

Categorías: Nacional Imágen y sonido Inteligencia Artificial y Robótica Marketing Programación Hardware E-Commerce Software



[**http://www.mexicopress.com.mx**](http://www.notasdeprensa.es)