

PROFILE

LINKEDIN

AGENTIC AI SYSTEMS

NVIDIA Nemotron RE-USE Test!

AI-Powered Background Music Removal and Vocal Isolation from “We’re Gonna Make It”

AGENTIC AI SYSTEMS — TECHNICAL RESEARCH PUBLICATION
Document ID: 202611-AASPUB-0010 | Publication date: 1 November 2026 | Version 1.2
Research platform: AI3080 — Dual NVIDIA GeForce RTX 3080 | Status: Successful experimental demonstration

1. Executive Summary

Agentic AI Systems (AAS) conducted an experimental evaluation of NVIDIA Nemotron RE-USE, successfully suppressing instrumental background music from a music video while retaining audible vocal content. The source recording was “We’re Gonna Make It” by Humidity Entertainment, featuring E-Double-D, 6Sense and DVS. The experiment used locally operated NVIDIA RTX 3080 GPU hardware and produced enhanced FLAC and MP3 audio artifacts.

The research also established a working CUDA-accelerated Python environment, compiled specialised Mamba-related dependencies and verified the installed software stack through a read-only audit on 1 November 2026. The result is a qualitative proof of concept; objective music-separation and vocal-fidelity metrics were not collected.

2. Original Music Video and Attribution

Field	Original source information
Music video	We’re Gonna Make It
Artist / label	Humidity Entertainment
Featured artists	E-Double-D, 6Sense & DVS
Associated album	The Build Up
Release date	30 March 2010
Copyright	© 2010 Humidity Entertainment

YouTube source: https://www.youtube.com/watch?v=OaGqUO8CKlI&t=88s
Bandcamp album: https://humidity.bandcamp.com/album/the-build-up

Release-associated contributors: ThaBrotherz, E-Double-D, Emerald Sun and DVS. Produced, mixed and recorded by: ThaBrotherz (A. Sutardy & J. Sutardy), thabrotherz@gmail.com, myspace.com/thabrotherzbeatz. Executive Producer: E-Double-D (E. Jamisola), humidityent@gmail.com, reverbnation.com/mcedoubled. These credits are reproduced from the source information supplied for this publication.

3. Research Objectives

The primary objective was to determine whether a neural speech-enhancement model could suppress instrumental accompaniment from the selected music video and preserve audible vocal performances. Secondary objectives were to confirm local GPU execution, validate export to standard audio formats, document software dependencies and assess the feasibility of future automated MP4 workflows.

4. Verified Hardware and CUDA Environment



PROFILE

LINKEDIN

AGENTIC AI SYSTEMS

Component	Verified configuration
Hostname / OS	ai3080 / Ubuntu 26.04.1 LTS
Linux kernel	7.0.0-38-generic
GPU 0 / GPU 1	NVIDIA GeForce RTX 3080 / NVIDIA GeForce RTX 3080
GPU memory	10,240 MiB each (independent memory pools)
NVIDIA driver	580.178.04
CUDA compiler	12.4.131
PyTorch CUDA build	12.4
CUDA availability / devices	True / 2 GPUs detected
GPU architecture	NVIDIA Ampere, compute capability 8.6

The 20,480 MiB aggregate installed VRAM is distributed across two separate GPUs and is not a unified 20 GB allocation. Confirmed RE-USE functional inference used GPU 0.

5. Verified Software Stack and Dependencies

The following versions were collected directly from the AI3080 Nemotron RE-USE virtual environment in AAS Software Stack & Dependency Audit V1.0 (1 November 2026).

Package	Verified version	Function
Python	3.11.17	Isolated runtime
PyTorch	2.5.1+cu124	CUDA neural inference
TorchAudio	2.5.1+cu124	Audio framework
Transformers	4.44.2	Supporting model library
Mamba-SSM	2.2.4	State-space sequence operations
Causal Conv1D	1.4.0	CUDA causal convolution
Safetensors	0.8.0	Model checkpoint handling
NumPy	1.26.4	Numerical arrays
SciPy	1.17.1	Scientific signal processing
SoundFile	0.14.0	Audio file input/output
Librosa	0.11.0	Audio analysis
Packaging	26.3	Version handling
PyYAML	6.0.3	YAML recipes
Einops	0.8.2	Tensor operations



PROFILE

LINKEDIN

AGENTIC AI SYSTEMS

Package	Verified version	Function
TQDM	4.70.1	Progress reporting

5.1 Media Processing Utilities

Utility	Verified version	Executable
FFmpeg	8.0.1-3ubuntu2	/usr/bin/ffmpeg
FFprobe	8.0.1-3ubuntu2	/usr/bin/ffprobe
SoX_ng	14.7.0.9	/usr/bin/sox

The FFmpeg build supports MP3 encoding via libmp3lame and multiple additional audio/video codecs, providing a practical foundation for media extraction, conversion and output preparation.

5.2 Environment Isolation and CUDA Extensions

Virtual environment: /home/edyboy/.venvs/aas-reuse-py311/
Interpreter: /home/edyboy/.venvs/aas-reuse-py311/bin/python

Mamba-SSM 2.2.4 and Causal Conv1D 1.4.0 were compiled and validated with CUDA 12.4. The selective_scan_cuda and causal_conv1d_cuda extensions imported successfully, and a Mamba GPU forward-pass test passed. The audit reported “No broken requirements found” and exited with status 0. It made no software or configuration changes.

6. Model Deployment

Model	Checkpoint	Configuration	Inference script
STANDARD	model.safetensors ~37 MB	config.json 830 bytes	inference.py ~4.4 KB
LAUGHTER	model.safetensors ~37 MB	config.json 921 bytes	inference.py ~4.4 KB

Model root: /home/edyboy/AI/MODELS/NEMOTRON-REUSE/

Both Standard and Laughter model variants completed independent functional audio inference tests. The music-video experiment also generated a Laughter-model output recorded under V2.6.

7. Experimental Methodology

7.1 Source and Preparation

The source material was the Humidity Entertainment music video “We’re Gonna Make It,” featuring vocals mixed with instrumental accompaniment. Audio was prepared for the Nemotron RE-USE pipeline. A recorded test input, NEMOTRON-REUSE-INPUT-66S.wav, was approximately 66 seconds long.

7.2 Neural Processing

The prepared waveform was submitted to the locally deployed Nemotron RE-USE inference pipeline. PyTorch and CUDA accelerated model execution on the RTX 3080. The resulting audio was assessed for suppression of instrumental accompaniment and retention of audible vocals.

7.3 Export and Listening Comparison



PROFILE

LINKEDIN

AGENTIC AI SYSTEMS

Enhanced audio was exported in FLAC and MP3 formats for playback and validation. A qualitative listening comparison indicated that the background music had been suppressed while the vocal content remained audible. No controlled quantitative separation scores were recorded.

8. Experimental Artifacts and Results

Recorded artifact	Approx. size	Purpose
NEMOTRON-REUSE-INPUT-66S.wav	5.6 MB	Prepared source audio
NEMOTRON-REUSE-INPUT-66S.flac	2.8 MB	Enhanced lossless output
AAS-NEMOTRON-REUSE-LAUGHTER-V2.6.mp3	1.1 MB	Enhanced demonstration MP3

The V2.8 audio validation record confirmed the presence of these files after inference. Their existence and successful generation do not independently quantify how much instrumental energy was removed.

8.1 Independent Model Functional Validation

Measurement	Standard	Laughter
Inference status	Pass	Pass
Output duration	5.100 seconds	5.100 seconds
Sample rate	44,100 Hz	44,100 Hz
Peak amplitude	0.663086	0.694397
RMS amplitude	0.029463	0.032130
Audio validation	Pass	Pass

Peak and RMS amplitude measurements verify signal properties, not speech intelligibility or source-separation quality.

9. Audio Processing Architecture

Music video → Audio preparation → Nemotron RE-USE → RTX 3080 CUDA inference → Enhanced vocal-focused audio → FLAC/MP3 export and validation.

The software layers include FFmpeg/FFprobe for media handling, SoundFile/Librosa/NumPy/SciPy for audio and numerical operations, PyTorch for inference, Mamba-SSM and Causal Conv1D for specialised neural operations, and Safetensors for model-weight storage.

10. Key Findings

Successful demonstration: The selected music-video recording yielded an output with suppressed background music and audible vocals. Local execution: RTX 3080 GPU inference succeeded. Operational models: Standard and Laughter passed functional tests. Reproducible environment: Exact dependency versions, model files and CUDA configuration were verified. Output portability: FLAC and MP3 artifacts were produced for listening and inspection.

11. Engineering Considerations



PROFILE

LINKEDIN

AGENTIC AI SYSTEMS

The project required correct selection of CUDA 12.4 for native extension builds, compilation of Mamba-SSM and Causal Conv1D, a dedicated Python 3.11 virtual environment, model-specific checkpoint and configuration handling, and audio-format conversion. Separating this environment from other AI3080 workloads reduced dependency-conflict risks.

12. Limitations

The experiment was a qualitative proof of concept on a selected recording. It did not measure source-to-interference ratio, signal-to-distortion ratio, residual instrumental energy, vocal timbre preservation, perceptual quality or repeatability across music genres. The enhanced result should not be described as a perfect reconstruction of an original studio vocal stem. Performance may vary with instrumentation, recording quality and vocal style.

13. Future Research

Next-stage work may compare Standard and Laughter outputs, evaluate additional songs and genres, benchmark against dedicated vocal-isolation systems, quantify residual music and vocal artifacts, and test automated MP4 extraction, enhancement and remuxing. These are proposed research directions, not completed validations.

14. Conclusion

Agentic AI Systems successfully demonstrated NVIDIA Nemotron RE-USE suppressing background musical accompaniment from the Humidity Entertainment music video “We’re Gonna Make It,” featuring E-Double-D, 6Sense and DVS, while retaining audible vocals. The supporting AI3080 environment was independently audited for package versions, CUDA availability, model files and dependency integrity. This establishes a practical technical foundation for further local AI audio-restoration research.

15. Research Sources, Credits and Disclosure

Original recording: “We’re Gonna Make It” — Humidity Entertainment, featuring E-Double-D, 6Sense & DVS; associated album The Build Up; released 30 March 2010. Produced, mixed and recorded by ThaBrotherz (A. Sutardy & J. Sutardy). Executive Producer: E-Double-D (E. Jamisola). © 2010 Humidity Entertainment.

Source links:
https://www.youtube.com/watch?v=OaGqUO8CKlI&t=88s
https://humidity.bandcamp.com/album/the-build-up

Research evidence: AI3080 Nemotron music-video test V2.6; audio output validation V2.8; Software Stack & Dependency Audit V1.0 dated 1 November 2026. The software versions derive from the provided terminal audit; the music-suppression result derives from the reported listening demonstration.

Disclosure: Independent research by Agentic AI Systems using NVIDIA technology. NVIDIA and Nemotron are trademarks of NVIDIA Corporation. This publication does not imply NVIDIA sponsorship or endorsement. Original music rights remain with their respective rights holders.

About Agentic AI Systems

Agentic AI Systems is an Australian technology initiative focused on artificial intelligence, intelligent automation, GPU-accelerated computing and practical AI integration. Intelligence. Automation. Advantage.
Website: www.agenticaisystem.com.au

Demonstration video: https://www.agenticaisystem.com.au/videos/watch/nemotron-reuse/