MCP SERVER | NO CODE | CLOUD HOSTED

NVIDIA Audio MCP

Turn any sound recording into structured data.

NVIDIA Audio provides professional-grade tools for handling complex audio files. You can transcribe spoken words, generate realistic voices from text, translate entire conversations across languages, and isolate different speakers in recordings. This MCP lets your AI client handle everything from raw meeting transcripts to polished, multilingual content.

A+ Quality Score 100/100

speech-to-text | text-to-speech | audio-processing | speaker-diarization | voice-cloning | transcription

MAINTAINED BY VINKIUS CLOUD FOR AI AGENTS · JUNE 2026
vinkius.com/mcp/nvidia-audio

SCAN TO VIEW

The connectivity layer between AI and the world's software.

Vinkius sits between AI and every application. All communication passes through Vinkius Cloud via the Model Context Protocol (MCP) — with governance, observability, and security at every layer.

Your AI Connections Run Through Vinkius Cloud

The world's largest managed MCP catalog

Vinkius is the connectivity layer where AI connects to the software your business already runs. We handle the hosting, the security, the credentials, the uptime — you get agents that actually do things.

We operate the world's largest managed MCP catalog. Major SaaS platforms, CRMs, databases, and cloud providers — running, monitored, production-ready. This MCP server is hosted and maintained by the Vinkius Cloud for AI Agents.

> *The agent doesn't manage credentials, doesn't manage uptime, doesn't manage security. Vinkius does.*
>
> — Architecture principle

Four Pillars of the Vinkius Runtime

01 — Security by design

Credentials stay encrypted at rest via AES-256. The AI agent never touches raw keys — they're injected into a sandboxed V8 isolate at runtime. Actions are logged, and connections have an emergency kill switch.

02 — Built on MCP Fusion

This MCP server was built with **MCP Fusion**, the open-source framework (Apache 2.0) that powers the entire Vinkius catalog. Schema-as-firewall strips undeclared fields, compiled PII redaction runs at zero overhead, and cryptographic lockfiles produce git-diffable audit trails.

03 — Deterministic observability

Eight immutable metrics per endpoint: request volume, p95 latency, error rate, active connections, cost attribution. A live payload feed logs every tool call with mutation detection.

04 — Autonomous operations

Servers are deployed, monitored, and patched autonomously. New capabilities and security patches ship weekly. Zero-downtime deployments ensure continuous availability across all managed MCP servers.

AES-256	Ed25519	24h TTL	V8 Isolate
Encryption at rest	PKI vault signatures	Ephemeral session keys	Sandboxed execution

One Token. Instant Access.

Every MCP server on Vinkius is accessed through a **Connection Token**. Tokens are generated in the cloud dashboard and produce a unique MCP endpoint URL. Paste this URL into any MCP-compatible client — no SDK required.

A single token can serve **multiple AI clients simultaneously**, or you can issue separate tokens per client for granular access control. Each token tracks its own request count, last activity timestamp, and can be individually enabled or revoked.

MCP ENDPOINT

```
https://edge.vinkius.com/{token}/mcp
```

Claude · Cursor · VS Code · Windsurf · Grok · Gemini

Security Is the Architecture

Security in Vinkius is not a feature — it's the foundation of the runtime. The gateway enforces multiple independent protection layers between AI agents and third-party APIs.

01 — Ed25519 PKI Vault

Every workspace has an Ed25519 Master Key. Session keys are generated ephemerally (24h TTL) and signed by the Master Key. Credentials never leave the vault boundary.

02 — V8 Isolate Sandboxing

Tool code runs inside isolated-vm V8 isolates with 64 MB memory caps and per-request timeouts. No filesystem access, no network access except through the SSRF-guarded fetch bridge.

03 — SSRF Guard

All outbound HTTP requests are DNS-resolved and validated before execution. Private IP ranges (10.x, 172.16-31.x, 192.168.x, AWS metadata 169.254.x) are blocked at the network layer.

04 — DLP & PII Redaction

A ResponseGuard pipeline intercepts every tool response. Configurable redaction patterns strip sensitive fields (emails, SSNs, card numbers) before data reaches the AI agent.

05 — Cryptographic Audit Trail

Every request is signed into a SHA-256 hash chain with Ed25519 signatures. Events form a tamper-proof, SIEM-exportable forensic record.

06 — Honeytoken Trap System

Phantom credentials are injected into isolated environments. If a honeytoken is used outside Vinkius infrastructure, the server is quarantined instantly.

Emergency Kill Switch

EU AI Act Art. 14(1) Compliant

The kill switch is an **emergency halt** mechanism — not a simple toggle. When triggered, it executes three actions atomically:

01 — Server deactivated

The MCP server is immediately taken offline across the entire cluster.

02 — All tokens revoked

Every connection token is invalidated. Total lockout — reconnection blocked until new tokens are issued.

03 — WebSocket connections killed

Active connections terminated via Redis pubsub broadcast. Propagates to every runtime node in the cluster.

Full Visibility. Zero Guesswork.

The Vinkius cloud dashboard includes a full MCP Governance suite — real-time analytics and security controls for production AI operations.

Control Plane

KPI dashboard with request volume, latency, success rate, token consumption, and AI-generated operational briefings.

Agent Activity

Which AI clients are connecting, how often, and what they're doing — real-time session tracking.

FinOps

Cost tracking per tool, payload compression savings, budget optimization signals, and consumption trends.

Tool Health

Slowest and most error-prone tools, with actionable root-cause insights and performance baselines.

Firewall & DLP

PII redaction activity, sensitive data protection counters, and security event timeline.

Incident Log

Error trends, failure rates, status-code breakdowns, and forensic audit trail access.

> Get started at **cloud.vinkius.com** — connect your AI agent in under 60 seconds.

NVIDIA Audio MCP

10 tools available
Cloud-hosted on Vinkius

This MCP connects advanced audio processing directly into your agent's workflow. Instead of manually feeding long audio files through multiple services—one for transcription, another for cleaning noise, and a third for translation—you pass the file once. Your AI client handles the whole chain: it transcribes speech to text using high-accuracy models, cleans up background noise, identifies who spoke when, and then can summarize that entire conversation into actionable bullet points. You'll find this MCP available in the Vinkius catalog alongside other powerful connectors. If you need to create content for multiple regions or languages, you can convert simple written text into natural speech using various voices, or even clone a voice from a short sample to generate entirely new audio segments. This ability to manage and polish every aspect of spoken word—from classification to punctuation restoration—turns raw recording data into perfectly structured, usable information.

Core Capabilities

01 — Transcribe speech to text

Turns any recorded audio file into accurate written text for immediate use.

02 — Identify different speakers

Separates and labels every voice in a recording so you know exactly who said what and when.

03 — Translate spoken audio

Converts spoken words from one language into another, maintaining natural flow.

04 — Generate realistic speech

Creates high-quality audio files from any text input, using customizable voices.

05 — Clean and improve recordings

Removes distracting background noises or adds proper punctuation to raw transcripts.

One Click on Vinkius — From Prompt to Execution

Available at **vinkius.com/mcp/nvidia-audio** — connect your AI agent in three steps.

01 First, subscribe to this MCP and provide your NVIDIA API Key within your agent's configuration.

02 Next, pass the audio file (or text you want spoken) from your AI client. The agent decides which sequence of tools is needed—like translating or cleaning up.

03 Finally, the MCP returns the processed output: clean transcripts, translated audio files, or new voice recordings ready for the next step in your workflow.

The bottom line is that you get a unified pipeline to turn raw sound into perfectly polished digital assets.

Built For

Content creators, multilingual support teams, and research analysts need this MCP. If your job involves listening to recordings—whether they're podcast interviews or customer calls—and turning that audio into organized text or translated media, this is for you.

Content Creator

Generates voiceovers in multiple languages and clones voices from existing material to quickly scale multilingual video content.

Customer Support Analyst

Processes call recordings, using speaker diarization to map out who spoke what, and transcribing the conversation for easy review and quality assurance.

Journalist/Researcher

Transcribes long-form interviews or public speeches, then uses tools like summarize_audio to pull key themes or topics efficiently.

What Changes When You Connect

01 Stop cleaning audio in multiple steps. Use cancel_noise to remove background buzz or traffic noise instantly, giving you clean source material right away.

02 Don't just transcribe; understand the speakers. speaker_diarization identifies who spoke when across a long call, making meeting minutes infinitely more accurate than simple word-for-word scripts.

03 Scale content globally without hiring translators. Feed text into audio_translation and generate polished voiceovers in dozens of languages using your agent's workflow.

04 Turn notes into media. If you have raw transcripts that lack commas or periods, run them through punctuate_text to make the writing look professionally edited before publishing.

05 Create endless content variations. Use clone_voice to replicate a speaker’s tone and pitch, letting your agent generate new material without needing the original person in the studio.

Real-World Applications

Analyzing multi-party calls

A customer support manager uploads 20 hours of call recordings. The agent uses speaker_diarization and speech_to_text to separate every conversation segment, creating a searchable database that shows who said what across all agents.

Creating global podcast episodes

A content creator records an interview in English. They pass the audio through audio_translation and then use text_to_speech to generate fully polished, localized voice tracks for Spanish and French audiences.

Meeting summary automation

After a 90-minute product planning meeting, the team runs the recording. The agent uses summarize_audio on the transcript to pull out only three key action items and responsible parties, saving hours of manual note-taking.

Cleaning up old field recordings

A researcher has raw audio from a remote location full of wind noise. They first run cancel_noise to clean the file, then use speech_to_text and punctuate_text to get a highly readable transcript.

Patterns to Avoid

Treating audio as just text

✗ AVOID

A user transcribes an interview using only speech_to_text, resulting in one massive block of unpunctuated text that is impossible to read or cite.

✓ INSTEAD

After running speech_to_text, always pass the output through punctuate_text. This ensures proper grammar and structure before you try to summarize it.

Ignoring speaker separation

✗ AVOID

A user sends a group discussion recording to an agent without identifying speakers, resulting in confusing transcripts where Speaker 1 and Speaker 2's comments get mixed together.

✓ INSTEAD

Always use the speaker_diarization tool first. This separates conversations by person, giving your AI client clean data for each contributor.

Assuming native language output

✗ AVOID

A user uploads a Spanish audio file and asks the agent to summarize it without specifying translation needs, resulting in an English summary of non-English content.

✓ INSTEAD

If the source language is not your working language, always run audio_translation first. Specify the target language so you get actionable results.

The Right Fit

Use this MCP if your core problem revolves around transforming *sound* into usable data (text, structured summaries, or new media). You need it when you are dealing with multi-speaker calls, foreign languages, noisy field recordings, or complex voice branding. If your task is purely text manipulation—like reformatting a document or writing an email—this MCP adds unnecessary complexity. Don't use this if you just need simple transcription; while speech_to_text works, remember that using speaker_diarization first gives you far more structural value. You might also not need the full power of this MCP if all you want is to read a file aloud; in that case, a simpler text-to-speech service will suffice.

The tedious mess of reviewing recorded conversations.

Imagine spending hours after every client meeting. You're stuck opening the recording, pausing it constantly to write notes in one window while transcribing what was said in another. Then, if the call involved three different people speaking, you have to manually track who brought up which point—all before you even start writing the final report.

With this MCP, your agent handles the entire process automatically. You feed it the raw audio file, and it returns a single, structured document: accurate transcripts with punctuation restored, clear labels for each speaker, and an instant summary of key decisions made.

NVIDIA Audio MCP delivers professional voice cloning.

Before this tool, creating multilingual marketing materials meant hiring a voice actor, paying for studio time, and dealing with inconsistent tones across different languages. If you needed to update content quickly, the cycle was slow, expensive, and dependent on availability.

Now, you provide one short sample recording of the desired voice and let your agent use clone_voice. You can then generate entire segments in a new language or context, giving you perfect consistency at scale. The time savings are massive.

NVIDIA Audio: 10 Powerful Tools

These tools let your agent handle every facet of audio processing, from simple transcription to advanced speaker identification and multilingual translation.

#	TOOL	DESCRIPTION
01	`list_audio_models`	Shows you a list of all available audio models the API can use.
02	`classify_audio`	Determines what type of sound is in an audio file and gives confidence scores for that classification.
03	`clone_voice`	Creates a digital replica of a voice using a small sample recording, allowing you to generate new speech later.
04	`cancel_noise`	Removes unwanted background sounds and static from the recorded audio file.
05	`speaker_diarization`	Analyzes an audio file to pinpoint and separate different speakers, noting when each person started and stopped talking.
06	`punctuate_text`	Adds correct punctuation and capitalization to raw text transcripts that might be missing these elements.
07	`speech_to_text`	Transcribes audio from multiple languages, taking a public URL for the MP3 or WAV file as input.
08	`summarize_audio`	Takes an existing audio transcript and boils it down to a concise summary.
09	`text_to_speech`	Converts written text into natural-sounding speech, letting you select different voices for the output.
10	`audio_translation`	Translates spoken audio directly from one language to a specified target language.

See It in Action

Real prompts you can use once this MCP is connected to your AI agent through Vinkius Cloud.

U Transcribe this meeting recording: https://example.com/meeting.mp3

Transcription: 'Welcome everyone to the Q2 review. Our revenue grew 15% compared to last quarter...'

U Convert this text to speech: 'Welcome to our presentation today.'

Speech generated successfully! Audio data available for playback.

U Identify different speakers in this call: https://example.com/call.wav

Detected 3 speakers: Speaker 1 (0:00-2:30), Speaker 2 (2:31-5:45), Speaker 1 (5:46-8:20).

Frequently Asked Questions

01 **Does NVIDIA Audio MCP support multiple languages?**

Yes, it supports numerous languages for both transcription and translation. You simply specify the source and target language when using audio_translation or speech_to_text.

02 **Can I clean noise from a recording before transcribing it with NVIDIA Audio?**

Absolutely. Before running the transcript through speech_to_text, you should first run cancel_noise on the audio file to remove background static or hums, ensuring cleaner results.

03 **How does speaker_diarization work with NVIDIA Audio?**

speaker_diarization analyzes an audio recording and outputs a time-stamped log that identifies different speakers by assigning them unique labels throughout the file's duration.

04 **What is the difference between summarize_audio and transcribing with NVIDIA Audio?**

Transcribing (speech_to_text) gives you every word spoken. Summarizing (summarize_audio) takes that full transcript and condenses it into key takeaways, saving you reading time.

05 **Is voice cloning in NVIDIA Audio restricted to one language?**

No, the clone_voice tool allows you to establish a unique audio fingerprint. You can then generate new speech using that cloned voice across multiple languages for consistent branding.

Go Live in 60 Seconds

Get your connection token from **cloud.vinkius.com**, then paste the endpoint URL into any MCP-compatible client.

YOUR MCP ENDPOINT

```
https://edge.vinkius.com/[TOKEN]/mcp
```

CLIENT	WHERE TO CONFIGURE
Claude AI	Profile → Customize → Connectors → "+" → Add custom connector → Paste endpoint
Cursor	Settings → Features → MCP Servers → "+ Add New MCP Server" → Type: SSE → Paste endpoint
VS Code	Ctrl/Cmd+Shift+P → "MCP: Add Server" → add `"nvidia-audio": { "url": "…" }`
Windsurf	MCP Settings → `mcp_settings.json` → Add endpoint URL
ChatGPT	Settings → Tools & plugins → Add MCP server → Paste endpoint
Gemini	Extensions → Add MCP Server → Paste endpoint URL

ASK AN AI ABOUT THIS

Let your preferred AI explain this MCP server

- Ask ChatGPT
- Ask Claude
- Ask Perplexity
- Ask Gemini
- Ask Grok

READY TO CONNECT

NVIDIA Audio is live on Vinkius Cloud.

Get your connection token, paste it into your AI agent, and start building. No SDK. No deployment. Just results.

Start at cloud.vinkius.com →

vinkius.com · support@vinkius.com

INDEPENDENT PLATFORM DISCLAIMER

Vinkius is an independent platform and is not affiliated with, endorsed by, sponsored by, verified by, or otherwise authorized by NVIDIA Audio. All third-party trademarks, logos, and brand names are the property of their respective owners. Their use in this document is strictly for informational purposes to identify service compatibility and interoperability.

DOCUMENT INFORMATION

Generated	June 2026
MCP Server	NVIDIA Audio MCP
Server ID	`019d75e1-0b2d-7243-a216-b8229bc7ffca`
Platform	Vinkius Cloud for AI Agents
Endpoint	`https://edge.vinkius.com/{token}/mcp`

LICENSE & USAGE

This document is generated automatically by the Vinkius PDF Engine. Content reflects the MCP server configuration at the time of generation and may change as updates are deployed. For the most current information, visit vinkius.com/mcp/nvidia-audio.