

Who is speaking, answered for you.

Automatic active-speaker detection for live production — as a library you embed in your own product.

VVD Detect listens across every microphone in a room and tells your application, in real time, **which person currently holds the floor** — the single decision an automated multi-camera production is built on. It also reads **how the room feels**, and it scales from a handful of microphones to **hundreds**. It is the detection engine proven in **Virtual Video Director**, offered to selected partners to build into broadcast switchers, conferencing systems, PTZ controllers and streaming appliances.

[Overview for prospective partners · share freely](#)

What it does, and why it matters

In any room with more than one microphone, the hard problem is not capturing audio — it is deciding, moment to moment, *who to show*. VVD Detect solves exactly that, so your product can switch cameras, frame a talker or cue a shot automatically, with the timing of a human operator.

You feed it audio from each microphone. It continuously reports:

- **The active speaker** — one stable channel to put on screen, held sensibly through the natural pauses in speech rather than flickering on every breath.
- **Who else is talking, and when two people talk at once** — the cue for a wide shot.
- **Silence**, and a live meter and speaking-likelihood for every channel.
- **The room's mood** — an optional read of the emotional tone (calm, tense, excited) that can drive shot width: a tense or excited room wider, a calm one tighter. It follows whoever is speaking on its own, from the audio you already send — nothing extra to wire up.

1 FILE TO SHIP	5 SAMPLE LANGUAGES	0 RUNTIMES TO INSTALL	100% OFFLINE
-------------------	-----------------------	--------------------------	-----------------

Proven, not a prototype

This is not a lab demo. It is the same engine that ships inside Virtual Video Director, in daily use in broadcast and corporate production. The value is in the **feel** — the difference between a switch that looks like a person made it and one that chases every cough. That judgement is years of tuning, and you get it as a finished component rather than something to rediscover.

Built-in intelligence, no cloud

AI models are built into the library — one that tells speech from paper shuffles, chairs and air-conditioning, and one that reads the room's emotional tone. Everything runs on the machine. **Nothing is sent anywhere** — no audio, no metadata, no phone-home — which is what lets it run on the firewalled and air-gapped networks broadcast facilities insist on.

Scales from one room to a whole rack

The AI runs across every channel in a single batched pass, so the same engine serves a small conference room or an installation with **hundreds of microphones**. For the largest channel counts it is offered in **GPU-accelerated builds** — for NVIDIA, and for Intel/AMD via DirectML — that are drop-in with the standard build: same interface, you simply choose which file to ship.

What we provide

The detection library

A single native library with a simple, language-neutral interface — active-speaker detection, AI voice-activity and room-mood sensing, with the models inside it. The standard CPU build is one file, no runtime to install; GPU builds for the largest installs are drop-in.

Sample applications

The same working app in five languages — C, C++, C#, Python and Rust — so your team starts from code that already runs, in whatever stack you use.

Activation tooling

A command-line activator for your installer, and a single-file USB provisioner for a manufacturing line or a field engineer — no runtime to install on the target machine.

Documentation & support

A full developer guide, an integration walkthrough, and direct engineering support during your integration.

How you implement it

Integration is deliberately small. Most of a first integration is your own audio and UI; the detection itself is a handful of calls.

- **Feed it audio.** Whatever you already have — from a mixer, a capture card, OBS, NDI, Dante or your own pipeline — is what you pass in. No special format or hardware.
- **Read the result.** Your switching or framing logic reads the active speaker and acts on it, at whatever rate suits your product.
- **Ship one file** beside your application. It works the same on every Windows machine, with nothing to install alongside it.

TYPICAL EFFORT

Teams have a live proof-of-concept running against the samples in an afternoon, and a real integration in **days, not weeks**. Your product keeps full control of cameras, framing and look — VVD Detect only answers "who is talking".

What stays yours

VVD Detect is one component, not a platform. Cameras, switchers, PTZ, transports, scene logic, your user interface and your product's identity are entirely yours. We answer one question extremely well; the product around it is yours to build and to brand.

How licensing works

Simple, flexible, and designed around how hardware and software are actually sold.

Evaluate for free. The library runs unlicensed in a limited mode — the complete product, including the AI model — so your engineers can prove it against your own audio before any commitment.

Two commercial models, whichever fits your business:

Model	How it works	Suits
Prepaid	Buy a pool of units up front; each machine you activate draws one.	A known volume, or a fixed fleet.
Royalty (true-up)	Activate freely; we invoice for what you shipped in the period.	Ongoing OEM production at scale.

Each licensed machine gets a credential that is **bound to that machine and verified entirely offline** — no per-use fees, no metering server your customers depend on, and no ongoing connection required in the field. Activation is one step in your installer or on your production line. Re-installing or servicing a machine you have already licensed costs nothing.

Working together

1 Introductory call & evaluation

We share this overview and an evaluation build; you try it against your own setup, at no cost and no commitment.

2 NDA & developer kit

Once there is mutual interest, an NDA opens the full kit — the developer guide, the complete interface, and engineering support.

3 Integration

Your team builds it in, starting from the samples, with our engineers on hand. Days, not weeks.

4 Commercial agreement

We agree the model — prepaid or royalty — and the terms that fit your product.

5 Ship

You activate machines as you build or install them, and ship. We support the integration for the life of the agreement.

Let's talk

If automatic, operator-quality speaker detection would strengthen your product, we would like to show you what it can do against your own audio.

Toby Mills · Noise Productions Ltd · toby@np.co.nz · +64 274 597427 · vvd.nz