

Photo: Project Flytrap, Hohenfels, Germany, June 2025 · U.S. Army, Spc. Elijah Magaña · Public domain · Wikimedia Commons

OSINT · TECHNICAL ASSESSMENT

THREE SENSES ARE BETTER THAN ONE

A microphone struggles in traffic, a camera at night, a radio receiver in a crowded band or against a drone that transmits nothing. Fusion is the answer most researchers now give, but how sensors are combined matters as much as which ones. This second part of Sound and Sight explains the two main ways of fusing them and reads two recent open studies for what fusion actually buys, and what it does not.

REPORT DETAILS

PUBLISHED	SUBJECT	SOURCE BASE
November 2026	Fusing acoustic, optical and radio-frequency sensing for small-drone detection	Inria tri-modal study (2025), Peraton Labs audio-visual study (2025), Halmstad University multi-sensor study (2021), AIP Advances review (2025), U.S. Army Project Flytrap
METHOD	**VERDICT**	**CONFIDENCE**
Compare single-sensor, paired and three-sensor results on the same data, clean and degraded	Ask for single-sensor baselines, degraded-condition results and false-alarm rates; fused accuracy alone hides what each sensor does	Moderate
AUTHOR		
Ghulam Mujtaba, Ph.D.		

BOTTOM LINE

No single drone sensor works everywhere, and the research now shows how badly each one can fail on its own. In a 2025 study from Inria, France's national research institute for digital science, models using sound, camera images or radio signals alone were all close to 97% accurate on clean recordings. On the same recordings deliberately degraded with noise, blur and interference, each fell to 64–67%, about what a system would score by answering "drone" every time, and each flagged most drone-free clips as drones. Combining all three senses raised accuracy to 83% and cut false alarms to 3%. A second 2025 study, from Peraton Labs in the United States, found that adding a microphone's audio to a camera detector improved its detection of small, distant drones by 11–15%. Fusion did not fix everything, since the fused system still missed about one drone clip in four in the worst conditions. But sensors that fail in different ways fail far less often when they are combined well, and acoustic sensing brings something the others lack.

64%

BEST SINGLE SENSOR, HEAVILY DEGRADED DATA

◆ ABOUT WHAT "ALWAYS DRONE" SCORES

83%

SOUND, CAMERA AND RADIO COMBINED

◆ FALSE ALARMS DOWN TO 3%

+11–15%

SMALL DRONES FOUND WHEN SOUND JOINS A CAMERA

◆ MEAN AVERAGE PRECISION, ONE STUDY

6 ms

PER FUSED DECISION ON A SMALL EDGE COMPUTER

◆ ABOUT 75 MILLIJOULES EACH

SERIES · SOUND AND SIGHT · PART 2 OF 3

PART 1	The Ear Points the Eye	✓ READ
PART 2	Three Senses Are Better Than One	YOU ARE HERE
PART 3	The Layered Picture	✓ READ

KEY QUESTIONS

What is multi-sensor fusion in drone detection?

Combining evidence from different kinds of sensor, typically microphones, cameras, radio-frequency receivers and radar, so that one sensor's weakness is covered by another's strength. It can be done by combining each sensor's verdict (decision-level fusion), by combining what the sensors perceive before a verdict (feature-level fusion), or by merging position reports over time in a tracker.

Does combining sensors improve drone detection?

In a 2025 Inria study on heavily degraded recordings, the best single sensor (sound, camera or radio) scored 64–67% accuracy, little better than always answering 'drone', and flagged most drone-free clips. Combining all three raised accuracy to 83% and cut false alarms to 3%, though it still missed about one drone clip in four.

Can sound help a camera detect small drones?

Yes, in research so far. A 2025 Peraton Labs study fused audio from a smartphone microphone into a camera-based drone detector and improved detection of small, distant drones by 11–15% (mean average precision), with smaller gains for larger drones.

THE RULER — THE WHOLE ARGUMENT

Which sense fails when

Each sensing method has conditions that blind it. Fusion works because those conditions rarely coincide.

CONDITION — WHAT CHANGES	SOUND	CAMERA	RADIO RECEIVER
DARKNESS, GLARE	Unaffected	Degraded	Unaffected
LOUD BACKGROUND: TRAFFIC, WIND	Degraded	Unaffected	Unaffected
CROWDED RADIO BAND: WI-FI, OTHER LINKS	Unaffected	Unaffected	Degraded
DRONE TRANSMITS NOTHING NEW	Unaffected	Unaffected	Blind

● FIBRE-OPTIC AND AUTONOMOUS DRONES: SEE OUR REPORT ON FIBRE-OPTIC DRONES

FROM THE INRIA STUDY'S REVIEW OF SINGLE-SENSOR WEAKNESSES AND THE REPORTS LINKED BELOW

FIELD 01 — WHY ONE IS NOT ENOUGH

Every sensor has a bad day

The Inria authors sum up the problem plainly. Acoustic systems struggle in noisy urban settings; camera-based detectors become unreliable in low light or when the drone is partly hidden; radio-frequency detectors, which listen for the link between drone and pilot, are confused in congested areas where other signals overlap, and can be beaten by drones that change frequency (When Jamming Stops Working). [arXiv↗] Radio detection has a further blind spot: drones steered through a fibre-optic cable, or flying autonomously, emit no control signal to find. Our report When the Detector Stays Silent covers that case.

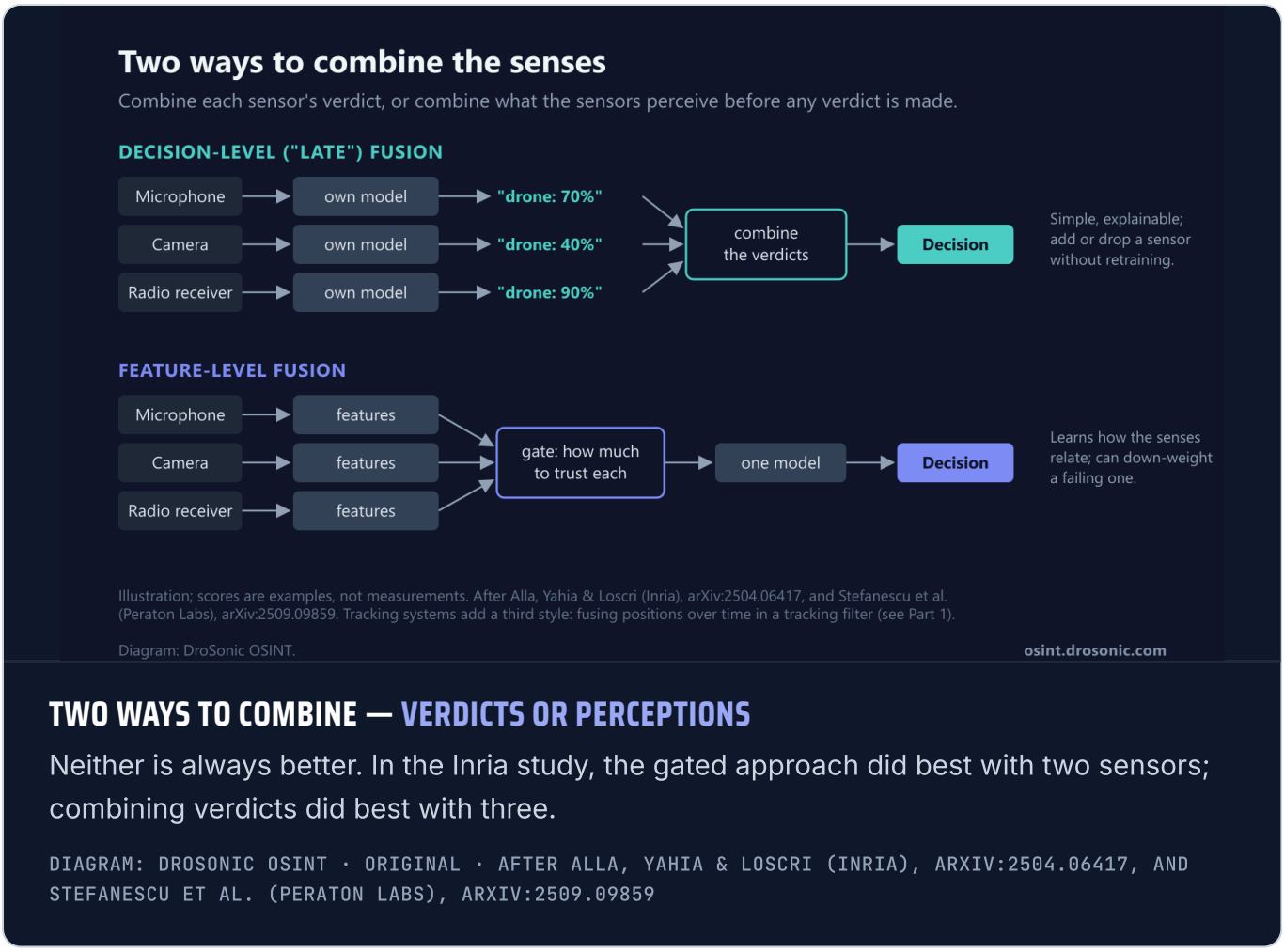
Earlier work pointed the same way. A 2021 study from Sweden's Halmstad University combined visible and thermal cameras with a microphone and found that fusion made the system more robust than any single sensor and reduced false detections; the thermal camera did about as well as the visible one despite its lower resolution. [arXiv↗] A December 2025 review of acoustic drone detection in *AIP Advances* lists ten studies combining sound with cameras, radar or radio, and expects future work to concentrate on synchronising data and smarter ways of fusing it. [AIP Advances↗] Fielded systems are moving the same way. A 2024 study of the war in Ukraine by ETH Zurich's Center for Security Studies notes that newer counter-drone systems use AI to analyse signals from radar, cameras, acoustic sensors and video together, while continually updating a threat library of known drone profiles. [CSS, ETH Zurich↗]

FIELD 02 — TWO WAYS TO FUSE

Combine the verdicts, or combine what they perceive

There are two broad approaches. In **decision-level** ("late") fusion, each sensor has its own model that reaches its own verdict, and the verdicts are then combined. It is simple and explainable, and a sensor can be added or removed without retraining the others. In **feature-level** fusion, what each sensor perceives is merged before any verdict, inside a single model. The Inria study used a "gated" version that learns how

much to trust each sensor, which can down-weight one that is failing. [arXiv↗] Tracking systems add a third style, merging positions over time, as in the camera-and-microphone system in Part 1.



TWO WAYS TO COMBINE — VERDICTS OR PERCEPTIONS

Neither is always better. In the Inria study, the gated approach did best with two sensors; combining verdicts did best with three.

DIAGRAM: DROSONIC OSINT · ORIGINAL · AFTER ALLA, YAHIA & LOSCRI (INRIA), ARXIV:2504.06417, AND STEFANESCU ET AL. (PERATON LABS), ARXIV:2509.09859

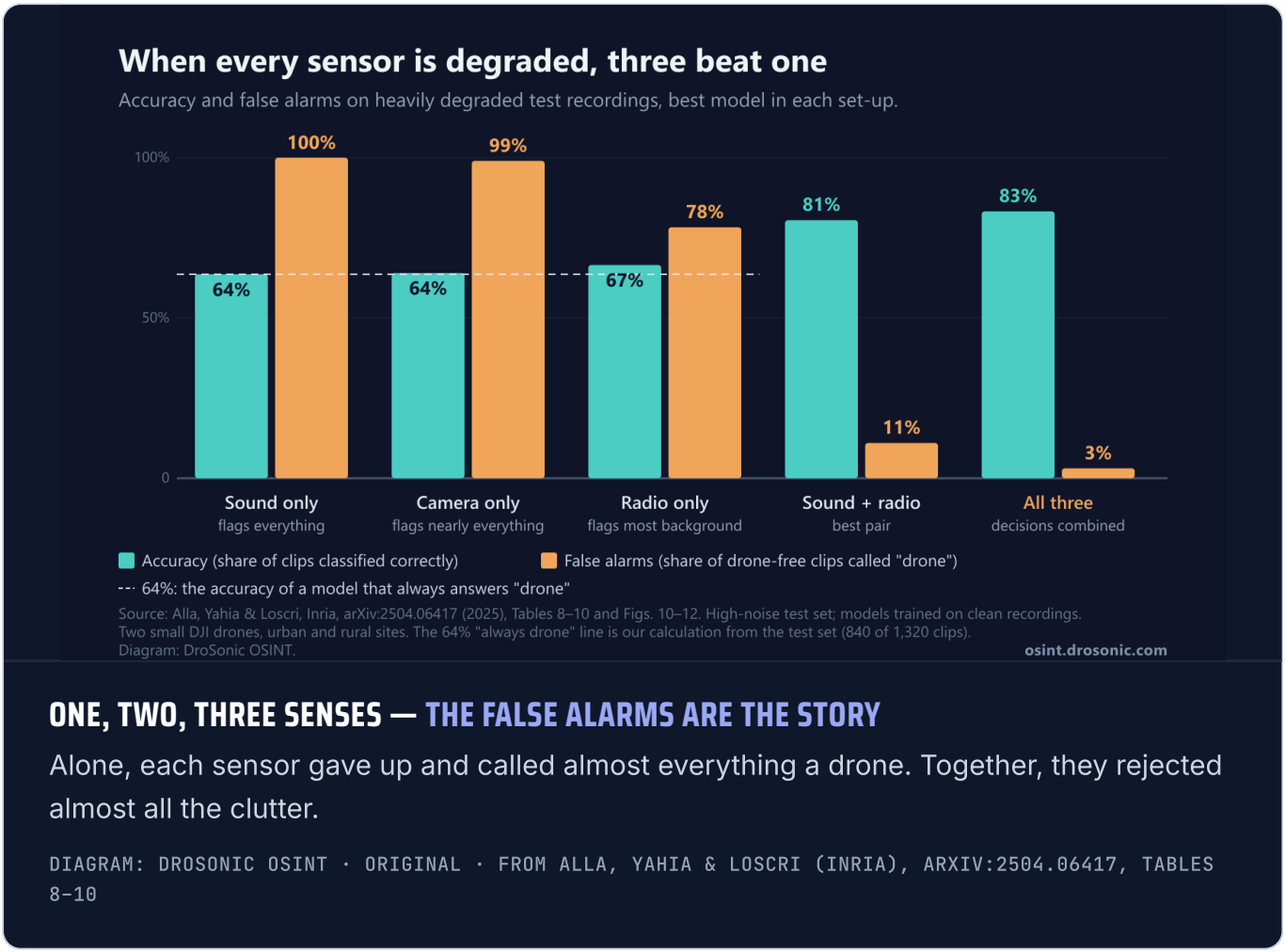
FIELD 03 — THREE SENSES, MEASURED

An open test of sound, sight and radio

The Inria team recorded two small DJI drones, a Mini 2 and a Mini 3 Pro, with an off-the-shelf USB microphone, a compact video camera and a software-defined radio listening on the 2.4 GHz band, in urban and rural places, by day and at sunset. Every sample pairs a quarter-second of sound, seven video frames and one radio snapshot taken at the same moment. They trained on clean recordings, then tested twice: on clean recordings, and on recordings degraded across all three senses at once with added noise, blur, colour shifts and radio jamming, to mimic a bad day in the field. The dataset and code are public. [arXiv↗]

On clean recordings almost everything worked: the best single sensors scored 90–98% accuracy and the fused systems about 97%. On the degraded recordings, the single sensors collapsed. The best sound model, the best camera model and the best radio model scored 64%, 64% and 67%. Because 64% of the test clips contained a drone, a model that always answered "drone" would also score 64%, and that is close to what the sound and camera models did: they flagged 100% and 99% of drone-free clips. The best radio model flagged 78%. [arXiv↗]

Pairs did better. The best, sound plus radio fused at the feature level, reached 81% accuracy with 11% false alarms. All three senses, combined at the decision level, reached 83% with 3% false alarms. [arXiv↗]



ONE, TWO, THREE SENSES — THE FALSE ALARMS ARE THE STORY

Alone, each sensor gave up and called almost everything a drone. Together, they rejected almost all the clutter.

DIAGRAM: DROSONIC OSINT · ORIGINAL · FROM ALLA, YAHIA & LOSCRI (INRIA), ARXIV:2504.06417, TABLES 8–10

The honest caveat is in the same table. In those worst-case conditions the three-sense system caught about three drone clips in four (76%), so it traded misses for far fewer false alarms. Conditions mattered too: under degradation it scored 77% by day, when

shifting light and shadow hurt the camera, against 90% at sunset, and 79% in town against 85% in the countryside. Fusion was cheap to run: about 6 milliseconds and 75 millijoules per decision on a small NVIDIA Jetson Orin Nano edge computer. [arXiv↗]

> **Alone, each sensor gave up and called almost everything a drone. Together, they rejected almost all the clutter.**

FIELD 04 — SOUND HELPS THE CAMERA

Small drones, found by ear

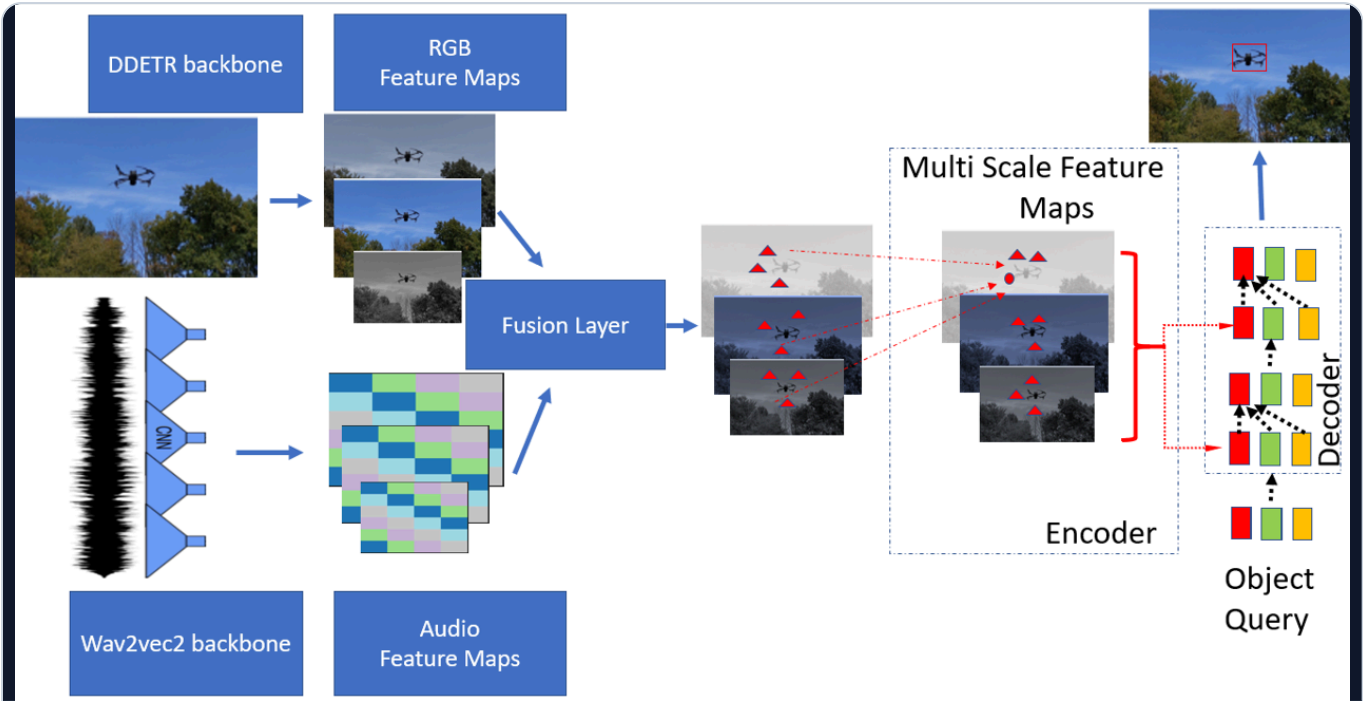
The second study asked a narrower question: can sound make a camera-based detector better at drawing a box around a drone in the picture? Peraton Labs recorded a DJI Mavic 3 Pro with nothing more than a smartphone's camera and microphone on a tripod, among trees, buildings, cars and water, by day and in the evening, and built a set of more than 7,500 matched image and one-second audio pairs. They then fed features from a speech-recognition-style audio model into a transformer-based image detector, testing four ways of mixing them. [arXiv↗]



THE DATA — A DRONE IN MANY SETTINGS

Frames from the Peraton Labs dataset: near and far, over grass and water, by day and at night.

IMAGE: STEFANESCU ET AL., PERATON LABS, ARXIV:2509.09859 (2025), FIG. 2 · CC BY 4.0 · ARXIV.ORG/ABS/2509.09859



THE METHOD — SOUND MIXED INTO SIGHT

Audio features join the image features before the detector decides where the drone is.

IMAGE: STEFANESCU ET AL., PERATON LABS, ARXIV:2509.09859 (2025), FIG. 6 · CC BY 4.0 · ARXIV.ORG/ABS/2509.09859

The gated method again did best. Its biggest gains came where cameras are weakest, on small drones occupying few pixels: detection scores (mean average precision) rose 11.1% on scenes like the training ones and 15.3% on scenes with different backgrounds, weather and light. Across all drone sizes the gain was smaller, 3.3–5.8%. [arXiv↗] That is the complement of Part 1: there, the ear pointed the eye; here, the ear helps the eye recognise what it is already looking at.

FIELD 05 — IN THE FIELD

Acoustic panels on the test range

Armies are testing the same idea outside the laboratory. In June 2025, at the U.S. Army's Project Flytrap trials in Hohenfels, Germany, soldiers of the 2nd Cavalry Regiment set up Squarehead acoustic panels alongside other counter-drone systems, and, in the Army's words, tested "the ability of the systems to work both independently and cooperatively with each other." [U.S. Army↗] The same trials appear in Watching the Bases, on drone incursions over military installations.



SQUAREHEAD ACOUSTIC PANELS — MOVED BY HAND, SET UP IN MINUTES

Soldiers relocate the panels during Project Flytrap, June 2025. The Army describes them as a counter-drone system that detects the sound of nearby drones.

PHOTO: U.S. ARMY, PFC. BRENT LEE · PUBLIC DOMAIN · WIKIMEDIA COMMONS

FIELD 06 — THE LIMITS

What the studies do not show

Both studies are honest about their limits, and a reader should be too. Each covers one or two small DJI quadcopters; neither includes fixed-wing drones, larger military types or swarms. The Inria team's worst-case conditions were simulated by degrading recordings, not observed in the field, and they note the need to add larger drones with different sound and radio signatures. [arXiv↗] Results on research data are usually higher than results in deployment, because field conditions differ from the ones a model was trained and tested on.

For anyone buying or integrating a multi-sensor system, the questions follow:

1. **What does each sensor score on its own, in bad conditions?** A fused figure without single-sensor baselines hides which sensor is doing the work.
2. **What is the false-alarm rate, not just the accuracy?** In the Inria results, the single sensors' accuracy looked middling while their false alarms were near 100%.
3. **Which way are the sensors combined?** Decision-level fusion is easier to audit and extend; feature-level fusion can adapt to a failing sensor but needs retraining.

4. **What happens when one sensor is missing?** A radio-silent drone, a dark night or a loud road should degrade the system, not break it.
5. **Which drones was it tested against?** Two small quadcopters is a start, not a threat library.

CONFIDENCE

How sure we are, and what would change it

High confidence that fusing sensors with different failure modes reduces false alarms and missed detections compared with any one sensor. Independent studies from France, Sweden and the United States agree, and a 2025 review describes the field moving that way.

Moderate confidence in the size of the gains. The figures come from research datasets of one or two small drones, with degraded conditions simulated rather than recorded, and from a single test set each.

What would change it: independent field trials of fused systems against varied drone types, including radio-silent ones, with single-sensor baselines and false-alarm rates published alongside the fused result.

DROSONIC — OUR APPROACH

The sense we build

Of the three senses in this report, DroSonic builds the one that needs neither light nor a radio link: networked passive acoustic sensors that classify drone sound with neural networks on low-cost edge hardware. DroSonic↗ It is designed to be fused, not to stand alone; its U.S. provisional patent application covers detection, classification, localisation and tiered defeat cueing. DroSonic↗

The studies here show what sound adds when the other senses degrade. DroSonic has not published fused-system results of its own; the prototype tested at DIRT Days 26.04 was evaluated as an acoustic sensor against live drones. DroSonic↗



DIRT DAYS 26.04 — THE SOUND LAYER

DroSonic's test position at the field exercise. In a fused system, this is the sense that works in the dark and in radio silence.

PHOTO: DROSONIC · FIELD-TEST WRITE-UP

CLOSE — WHAT IT MEANS

The sense that does not depend on light or radio

Fusion is how a defence keeps working when one sensor is having a bad day, which matters more than whether it appears on a datasheet. Of the three senses studied here, sound is the one that needs neither light nor a radio link, which is why it keeps appearing as the complement to cameras and radio receivers, and why Ukraine's acoustic networks, described in Listening to the Sky, were built as a layer beneath other sensors rather than a replacement for them.

DroSonic builds that acoustic layer, which is passive, emits nothing an enemy can find, and works in darkness and radio silence. Part 3 of Sound and Sight turns from algorithms to plumbing. It looks at how sensors from different makers share what they detect in practice, through open standards, and where an acoustic layer plugs in.

ABOUT DROSONIC

DroSonic builds passive acoustic drone detection: networked sensors that listen for drones instead of their radio signals, using acoustic AI and edge computing. DroSonic OSINT is our open-source research into the threat that technology is built to meet.

THE RECORD SO FAR

OCT 2026	Joined the SOF RACER consortium, which connects USSOCOM with commercial technology companies
SEPT 2026	Prototype tested against live drones at DIRT Days 26.04, a field exercise supported by CMI2 with U.S. Central Command and U.S. Army DEVCOM
SEPT 2026	Technical presentation at NDIA's Emerging Technologies for Defense conference on detecting fibre-optic and radio-silent drones
SEPT 2026	Filed U.S. provisional patent 64/148,522 on passive acoustic detection, classification, localisation and defeat cueing
JUL 2026	Pitched to the U.S. Space Force at Space Systems Command's Reverse Industry Engagement Day, Colorado Springs

How DroSonic detection works → · Where it is used → · Talk to the team →

THIS REPORT ONLINE

https://osint.drosonic.com/assessments/multi-sensor-drone-detection-fusion.html

Cite as: Mujtaba, G. (2026, November 19). *Three Senses Are Better Than One: Multi-Sensor Drone Detection: How Sound, Camera and RF Fusion Works*. DroSonic OSINT. https://osint.drosonic.com/assessments/multi-sensor-drone-detection-fusion.html

Reusing this report. Text and original diagrams © DroSonic. You may quote this report and share its link with credit to DroSonic OSINT; to republish it in full, reuse our diagrams or use it commercially, contact us first. Photos and third-party figures keep the licences shown in their credits.

DroSonic OSINT: open-source assessments on drones and counter-UAS from DroSonic, passive acoustic drone detection. drosonic.com · sudo@drosonic.com



ABOUT THE AUTHOR

Ghulam Mujtaba, Ph.D.

FOUNDER AND CEO, DROSONIC

Ghulam Mujtaba, Ph.D., is the founder and CEO of DroSonic and leads its work on passive acoustic drone detection. A computer engineer with 13 years of experience building machine-learning systems for edge hardware, he holds a Ph.D. in computer engineering from Gachon University in South Korea, was an NSF postdoctoral fellow at West Virginia University in the United States, and is an Assistant Professor of AI and Data Science at Regis University in Denver, Colorado. An IEEE Senior Member, he has published more than 20 papers on deep learning, computer vision and multimedia for edge devices.

PROFILE ↗ LINKEDIN ↗ ORCID ↗

ALL SOURCES (9)

1. arXiv — Alla, Yahia & Loscri (Inria) 2025: acoustic struggles in noisy urban settings; visual unreliable in low-light or occluded settings; RF prone to interference in congested areas
2. arXiv — Svanström, Englund & Alonso-Fernandez, ICPR 2021: thermal IR comparable to visible; fusion more robust than individual sensors and reduces false detections
3. AIP Advances — Kang et al. 2025 review, Table V and Sec. VI.A: multimodal studies; emphasis on data synchronisation, sensor coordination, intelligent fusion algorithms
4. CSS, ETH Zurich — Kunertova 2024, Learning from the Ukrainian Battlefield, Sec. 2.3: AI-enhanced analytics across radar, electro-optics, acoustics and video; constantly updated threat library
5. arXiv — Stefanescu et al. (Peraton Labs) 2025: DJI Mavic 3 Pro; Google Pixel 2 camera and microphone; ARDrone dataset >7,500 synchronized pairs; Wav2Vec2 audio features fused into Deformable DETR; linear, MLP, gated and cross-attention fusion
6. U.S. Army — Project Flytrap, 10-11 June 2025: Skyview, Titan and Squarehead acoustic panel systems tested independently and cooperatively
7. DroSonic — Passive acoustic detection: neural networks on low-cost edge hardware, no emissions
8. DroSonic — U.S. provisional patent application 64/148,522, September 2026
9. DroSonic — DIRT Days 26.04 field test, September 2026