AVer CAM570 Voice Tracking Setup Guide

V1 May 2026

AVer

CAM570 Voice Tracking Setup Guide

Smarter speaker tracking for more natural meetings

The **CAM570** is a 4K dual-lens PTZ camera designed for modern meeting spaces. With **Single-Camera Voice Tracking** enabled, it automatically detects the active speaker and adjusts pan, tilt, zoom, and framing to keep that person in view.

This guide explains how to configure CAM570 for reliable voice tracking and a better meeting experience.

To apply these settings, use the latest version of Room Management or the device web interface.

Basic Installation

Voice tracking relies on both the **AI reference lens** and the **main lens**. For the best results:

- Mount the camera at approximately **59 -62 inches (150-160cm) above the floor**
- Keep all participants within the field of view of the **AI reference lens**
- Make sure the **AI reference lens and voice sensors are not blocked** by people laptops, bottles, or other table items

Lens Specifications

- **Main lens HFOV:** 78°
- **AI reference lens HFOV:** 82°

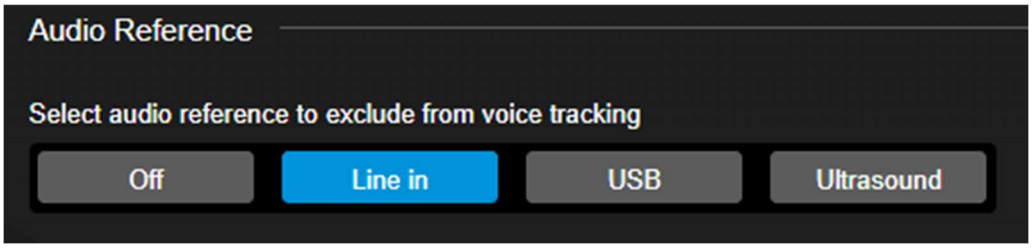
Best Practice
Position the camera so all participants remain visible to the AI reference lens. This improves speaker detection and framing accuracy.

Audio Integration

If the CAM570 is connected to an external audio system, configure the audio reference correctly so the camera does not respond to room audio playback when no local speaker is present.

CAM570 includes a **3.5 mm TRS input** for integration with external audio systems.

When using this connection, select **Line In** as the audio reference.



When paired with a **FONE540**, CAM570 can use ultrasound to pause tracking while audio is being played through the speakerphone.

Important
If the audio reference is not configured correctly, the camera may respond to audio playing through a speaker instead of an in-room talker.

How Voice Tracking Works

CAM570 voice tracking works in two steps:

1. It detects the direction of a **human voice** using the built-in voice sensors
2. It moves the PTZ camera toward the active speaker and applies **smart framing**

The system is designed to ignore non-human sounds for more accurate tracking.

Tip
For the most natural meeting experience, combine correct camera placement with tighter framing settings.

Voice Tracking Modes

CAM570 supports two voice tracking modes:

- **AI Lens Tracking**
- **Main Lens (PTZ) Tracking**

Main Lens (PTZ) Tracking

When **Main Lens Tracking** is enabled:

- All three sets of voice sensors are active
- This mode is best for wider speaking areas

AI Lens Tracking

When **AI Lens Tracking** is enabled:

- Only the center voice sensors are active
- Tracking distance is reduced to up to **23 ft (7m)**

Important
When using AI Lens Tracking, the talker must remain within the AI camera's field of view. If voice is detected outside the field of view, tracking will not activate.

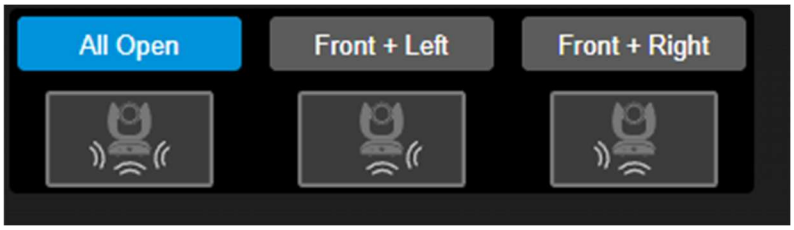
Voice Sensor Coverage

The CAM570 includes three sets of voice sensors across the **left, center, and right front** of the device. Together, they provide **180-degree horizontal coverage** in front of the camera.

These sensors are intended to support **voice tracking only** and are **not used for conferencing audio**.

For the best tracking performance, Install the CAM570 at least **3 ft (1 m)** away from side walls

If that spacing is not possible and **Main Lens Tracking** is enabled, selected voice sensors can be disabled to help reduce echo interference in corner installations.



Best Practice
Avoid mounting the camera too close to side walls. Extra space helps reduce reflections and improves tracking consistency.

Tip
Rooms with better acoustical design typically deliver more reliable tracking and clearer conference audio.

Conversation Mode

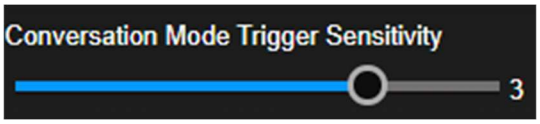
Conversation Mode activates when two or more speakers are detected within a short time. The camera widens

the shot to include multiple participants, reducing frequent reframing and creating a more comfortable view for remote attendees.

Sensitivity Settings

Sensitivity controls how quickly the camera switches from an individual speaker's view to a wider group view during back-and-forth discussion.

- **Lower sensitivity** makes the camera less reactive, so speakers must alternate more times before the view widens.
- **Higher sensitivity** makes the camera widen sooner.
- **Setting 0** turns Conversation Mode off.



Sensitivity level	Speaker alternations before group view	Example sequence
0 —Mode off	N/A	None
1 — Least sensitive	3	A → B → A
2	5	A → B → A → B → A
3	7	A → B → A → B → A → B → A
4 — Most sensitive	9	A → B → A → B → A → B → A → B → A

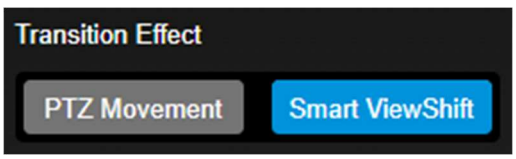
Tip
Use a lower sensitivity setting if the camera switches to group view too often. Increase sensitivity if the room has frequent back-and-forth discussion.

Transition Effects

CAM570 includes two transition options that control how the camera moves between speakers:

Smart ViewShift

Creates a smoother transition by briefly holding a freeze frame before moving to the next position.



PTZ Movement

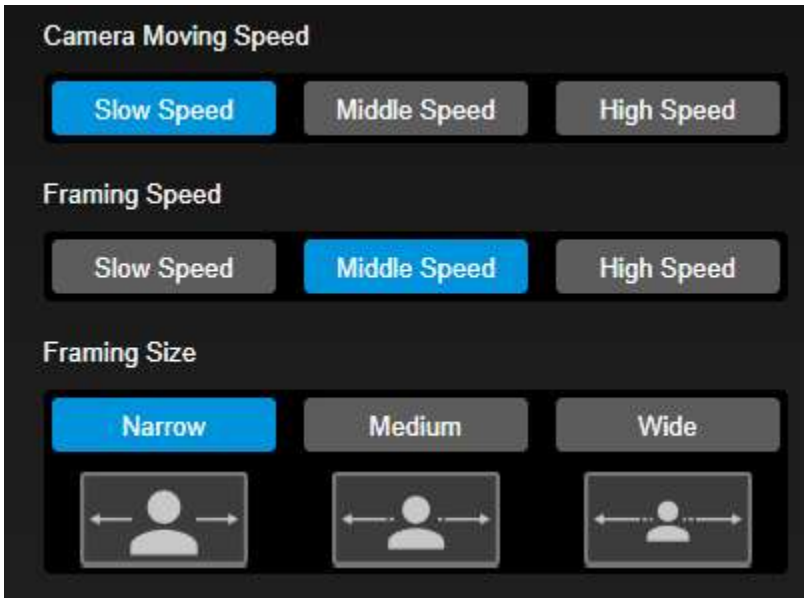
Shows live video while the camera physically moves between speaker positions.

Adjustable Settings

For either transition mode, you can adjust:

- **Camera moving speed** – how fast the PTZ head moves
- **Framing speed** – how quickly the camera responds and reframes
- **Frame size** – the size of the area shown around the active participant(s)

A **narrower frame size** typically improves tracking accuracy.



Best Practice

Start with a narrower frame size to improve tracking precision, then widen only if more room context is needed.

Audio Tracking Idle Interval

If no human voice is detected for a specified period, the camera returns to auto framing.

- **Default idle interval:** 15 seconds
- **Adjustable range:** 7 to 25 seconds

This setting can be adjusted in Room Management.

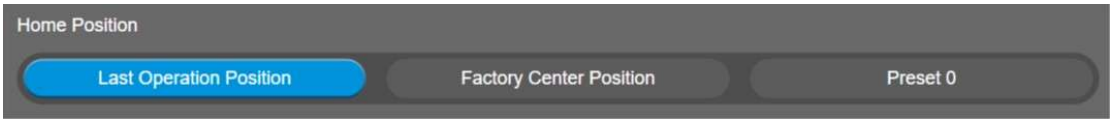
Behavior by AI Detection Source

- If **AI Lens** is selected, CAM570 returns to **auto frame** after the idle interval
- If **Main Lens** is selected, CAM570 first returns to the **Home Position**, then switches to **auto frame**

Important
The selected AI detection source changes how the camera resets after inactivity. Confirm this setting during deployment to ensure the expected room behavior.

Home Position

Home Position defines where the camera returns when no active speaker is detected after the idle period.



There are three options:

1. **Last Operation Position**
   Returns to the most recently used PTZ position
2. **Factory Center Position**
   Returns to the straight-ahead default position
3. **Preset 0**
   Returns to a position manually saved as Preset 0

Tip
Use **Preset 0** when you want the camera to return to a specific room view after periods of inactivity.

Presentation Mode

Presentation Mode is a variation of audio tracking that uses audio detected within a predefined zone to trigger preset camera positions. This is useful for spaces such as:

- Podiums
- Whiteboards
- Interactive displays
- Presenter positions
- Participant seating areas

How to Configure Presentation Mode

1. Define the preset zones based on the room layout
2. Assign preset points for target locations such as the podium, whiteboard, touchscreen, or participant seats
3. When Presentation Mode is active, the camera moves to the assigned preset whenever human voice is detected within that zone

For best accuracy, frame each preset zone as tightly as possible to avoid including multiple people in the same scene.

Best Practice
Keep each preset zone focused on a single activity area. Tighter zones improve accuracy and reduce unintended camera switching.

Quick Setup Checklist

Use this checklist before finalizing installation:

- Camera mounted at 59-62 inches (150-160cm) above the floor
- Participants visible to the AI reference lens
- Voice sensors free from obstruction; and at least 3 ft (1 m) from side walls when possible
- Audio reference configured correctly

- Tracking mode selected for the room layout
- Idle interval and Home Position configured
- Presentation presets adjusted as needed

Summary

CAM570 voice tracking helps create a more natural and engaging meeting experience by automatically focusing on the active speaker. With the right setup and configuration, it can deliver smooth transitions, better room coverage, and more consistent framing across a wide range of meeting environments.