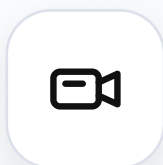




Multimodal AI Audio-Visual Tracking Box AIVIS-MT200



Cross-Modal
Coordinate Alignment



Multi-Camera
Control



Speaker
Tracking



Preset-Free
Tracking



Scenario-Adaptive
Algorithms



Visual
Configuration Tool

AIVIS-MT200 is a powerful, high-quality multimodal AI audio-visual tracking box developed by AISpeech. It is designed for spaces such as corporate meeting rooms, training rooms, and university lecture halls. MT200 can be paired with AISpeech PTZ cameras and work together with AISpeech's flagship ceiling microphone system. By fusing cross-modal audio-visual spatial coordinates, MT200 ensures that wherever the voice goes, the camera follows. Without preset camera positions or manual switching, it accurately locks onto the speaker, tracks smoothly, and outputs intelligent multi-view video layouts. MT200 adapts to complex scenarios, from multi-person discussions and impromptu speaking to presenters moving around the room. Remote attendees can instantly identify who is speaking, improving visual storytelling quality, remote communication efficiency, and the professionalism of live streaming. Designed for medium-to-large spaces, MT200 elevates your remote collaboration, live streams, and meeting archives with polished, broadcast-grade production quality.

Product Features



High-Precision Audio-Visual Spatial Alignment

Combines ceiling-microphone sound-source localization with PTZ camera vision to build a unified audio-visual coordinate system in real spaces, quickly identifying who is speaking and where.



Intelligent Speaker-Tracking Camera Control

Uses real-time DOA data from ceiling microphones to precisely control PTZ pan, tilt, and zoom, keeping the speaker centered even when posture changes.



Scenario-Adaptive Vision Algorithms

Supports C40T, C60, and C80 cameras for large meeting rooms, training rooms, seminar rooms, and other typical scenarios.



Preset-Free Full-Space Tracking

No preset positions are required. With acoustic-visual cross-modal coordinate alignment, the system enables free speaker tracking across the entire space.



Automatic Re-Centering for Active Speakers

The system monitors speaker posture shifts and position changes in real time, dynamically adjusting the camera to keep the person visually centered.



Intelligent Multi-Channel Video Integration

The system supports automatic intake of multiple video streams and uses AI vision algorithms to intelligently merge multi-camera views into a combined output.



Fine-Grained Tracking Strategy Configuration

Built on a C/S architecture, the visual configuration tool enables one-stop setup and management of ceiling microphones, PTZ cameras, sound-source zones, spatial calibration, video preview, and speaker-tracking rules.

Hardware Specifications

General Specifications		Hardware Interfaces	
Color:	Starlight Brown	Audio:	Line In / Line Out / Phone In (reserved)
Dimensions:	220mm × 190mm × 44.5mm	HDMI:	1 x HDMI-IN, 2 x HDMI-OUT, up to 4K@30fps
Power Supply:	PoE + DC power	VGA:	1 x VGA-OUT, up to 1080p
Net Weight:	< 1.5kg	USB:	Type-C OUT x1, USB Device IN x4
Operating Temperature:	0°C ~ +40°C	Network:	RJ-45 ×2, 100M/1000M adaptive, 1 x PoE
Speaker Tracking		TF Card:	External MicroSD card (reserved)
Video Input:	Supports input up to 4K@30fps		
Video Output:	Supports output up to 4K@30fps		
Video Switching:	Built-in AI algorithms; automatic switching		
Speaker Tracking:	Preset-free tracking with intelligent multi-camera production		