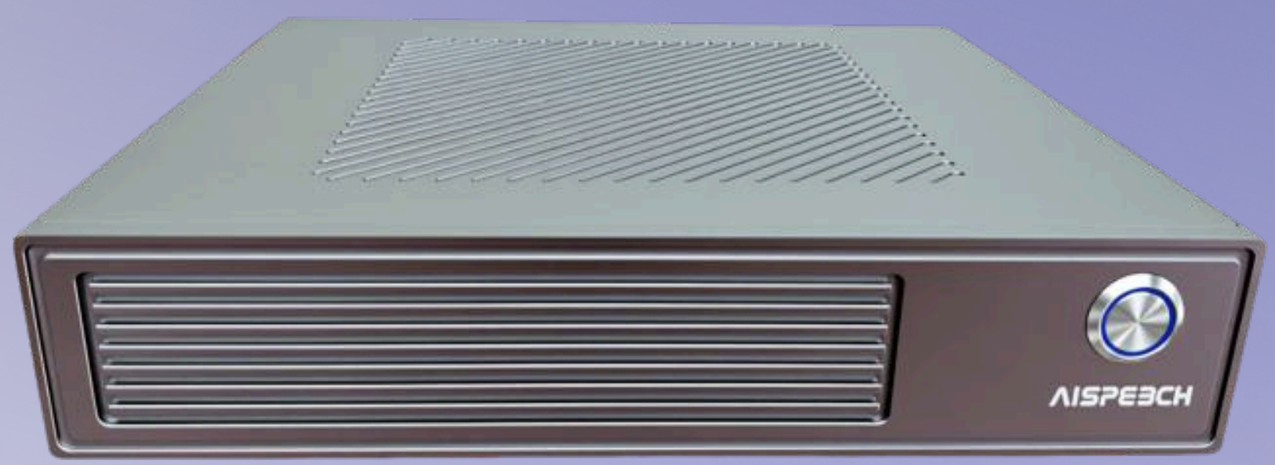
Multimodal AI Audio-Visual Tracking Box MT200



© 2026 AISPEECH and/or its affiliates. All rights reserved

CATALOG

1 Product Introduction

2 Main Features

3 Product Specifications

4 External Dimension

5 Component Instructions

© 2026 AISPEECH and/or its affiliates. All rights reserved

Product Introduction

AIVIS-MT200 is a high-performance multimodal AI audio-visual tracking box from AISpeech, designed for corporate conference rooms, training rooms, university lecture halls, and other medium-to-large spaces. It works with AISpeech PTZ cameras and AISpeech flagship ceiling microphone arrays. By deeply fusing acoustic and visual spatial coordinates, MT200 delivers a natural "camera follows the voice" production experience throughout a meeting.

Without preset camera positions or manual switching, MT200 accurately locks onto the active speaker, tracks them smoothly, and intelligently combines multiple video feeds. It adapts to complex scenarios such as multi-person discussions, ad-hoc speaking, standing conversations, and movement across the room, helping remote participants quickly and clearly identify the current speaker. It delivers a more polished meeting view, smoother remote collaboration, and professional-quality streaming and recording for medium-to-large spaces.

© 2026 AISPEECH and/or its affiliates. All rights reserved

Main Features

Precise audio-visual spatial alignment

MT200 fuses ceiling-microphone source localization with PTZ camera visual perception to build a unified audio-video coordinate system. Real-time DOA data lets the system identify who is speaking and where they are for accurate voice-to-camera matching.

Multi-scenario visual algorithm scheduling

Works with AISpeech C40T, C60, and C80 cameras for large conference rooms, training rooms, and seminar rooms. Camera roles can be configured for view response, presenter tracking, whiteboard close-ups, switching, composition, and scene management.

Automatic Re-Centering for Active Speakers

Using real-time DOA data, MT200 precisely controls PTZ pan, tilt, and zoom. Even when a speaker changes posture, the camera automatically keeps the speaker centered in frame.

Fine-grained tracking strategy configuration

The C/S visual configuration interface enables fine-grained management of microphones, PTZ cameras, and sound-source zones, with real-time status monitoring, spatial calibration, camera preview, and video switching strategy control.

Preset-free multi-camera linkage and video composition

MT200 supports cascading ceiling microphones and C80 binocular 4K tracking cameras for preset-free audio-video linkage. It automatically dispatches camera positions, selects the best close-up across multiple feeds, and applies transition strategies for natural remote-meeting video.

© 2026 AISPEECH and/or its affiliates. All rights reserved

Product Specifications

Function	Specification
Resolution	• Up to 4K @ 30 fps output
Audio Interfaces	• 1 x Line In (3.5 mm, reserved) • 1 x Line Out (3.5 mm, reserved) • 1 x Phone In (3.5 mm, reserved)
Video Input	• Up to 4Kp30 input
Video Output	• Up to 4Kp30 output
Video Switching	• Built-in AI algorithm with automatic video switching
HDMI Ports	• 1 x HDMI IN • 2 x HDMI OUT, up to 4K video input/output
USB Ports	• 1 x USB 3.0 Host Type-C OUT • 4 x USB 3.0 Device IN
VGA Port	• 1 x VGA OUT (reserved), up to 1080p output

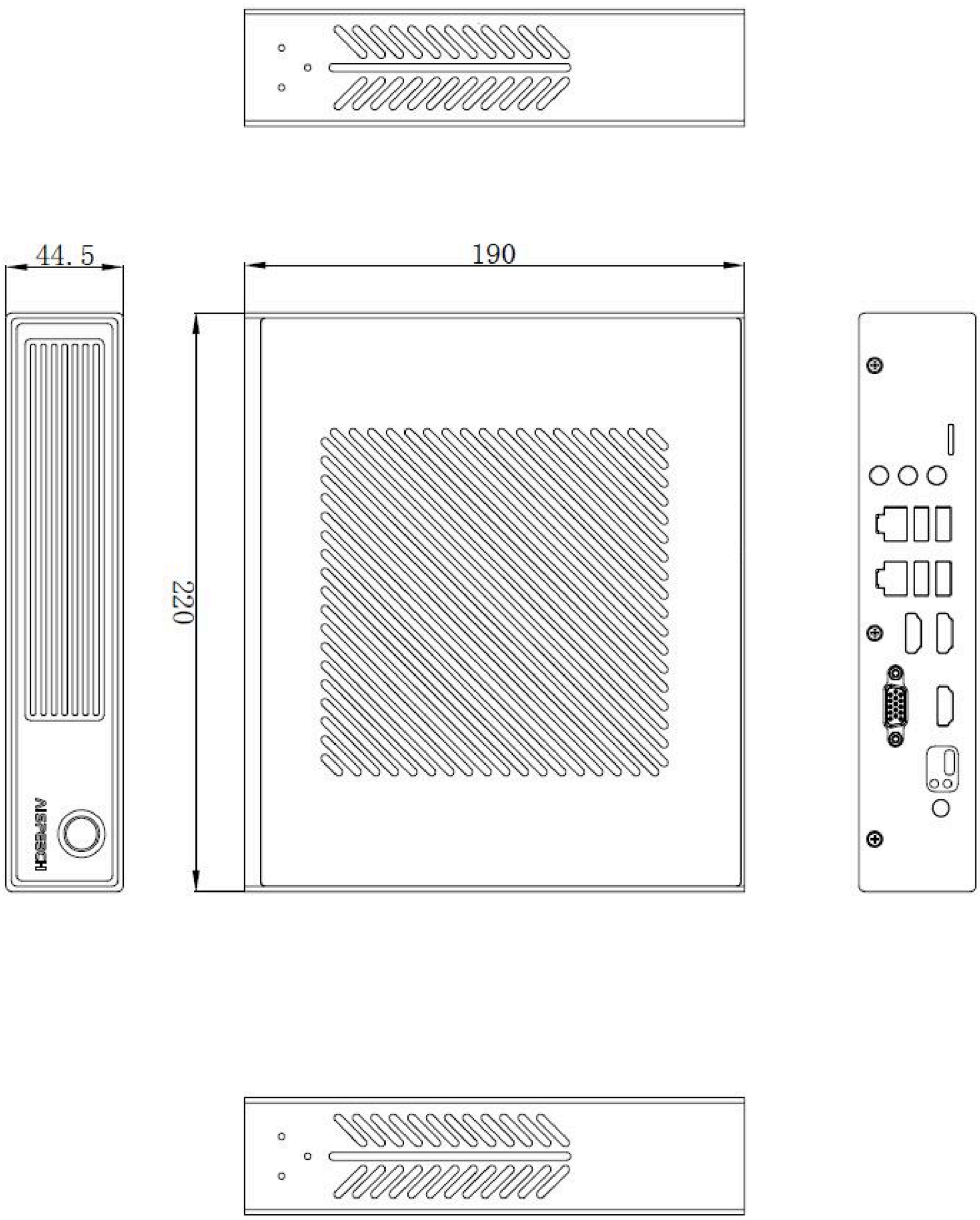
© 2026 AISPEECH and/or its affiliates. All rights reserved

Product Specifications

Function	Specification
Network Ports	• 2 x RJ-45 100/1000 M Ethernet ports, including 1 PoE port • Supports multiple 4Kp30 video inputs and single-channel external output up to 4K30
TF Card	• 1 x external Micro SD card (reserved)
NPU Computing Power	• 6 TOPS
Power	• DC 12 V
Power Supply	• PoE + DC power
Operating Temperature	• 0°C to +40°C
Storage Temperature	• -20°C to +60°C
Dimensions	• 220 mm x 190 mm x 44.5 mm
Weight	• < 1.5 kg

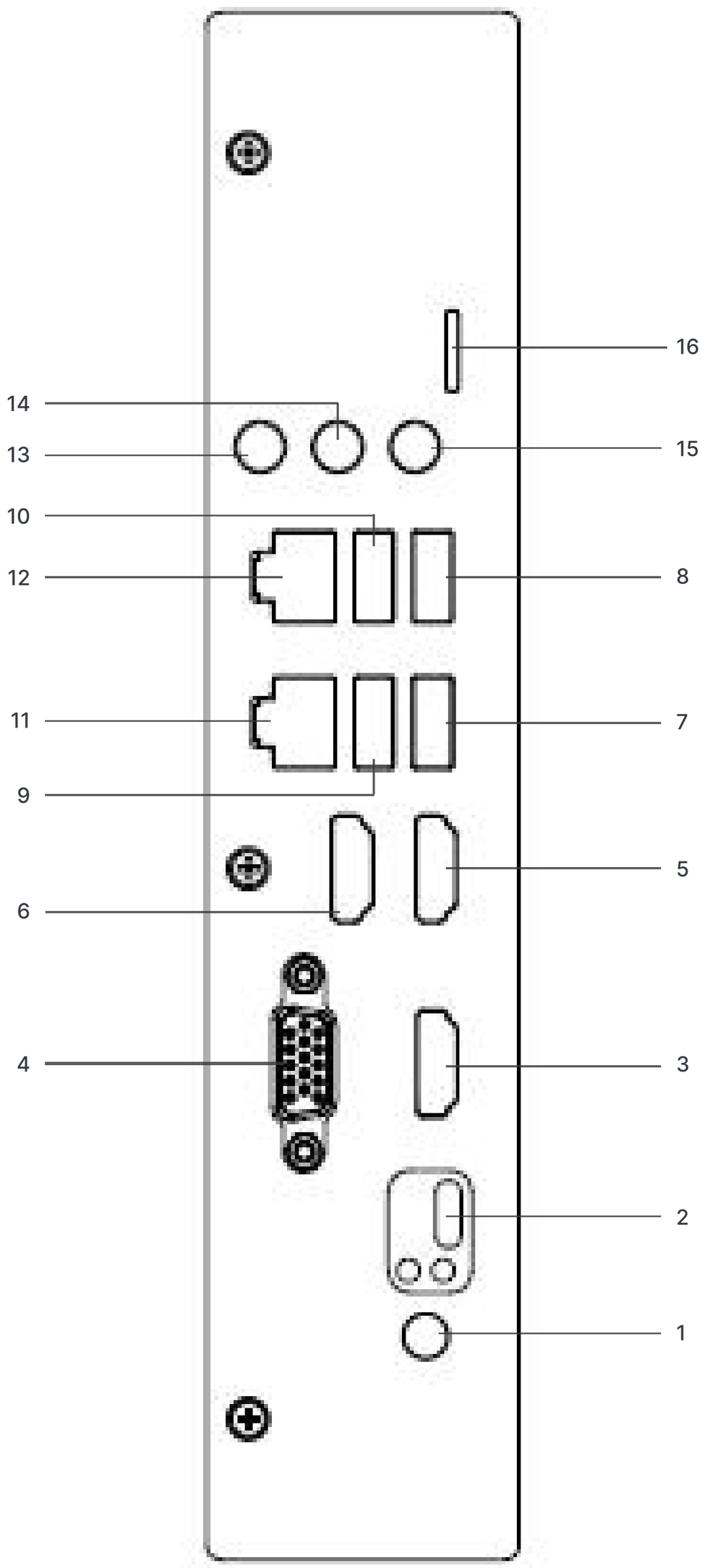
© 2026 AISPEECH and/or its affiliates. All rights reserved

External Dimension



© 2026 AISPEECH and/or its affiliates. All rights reserved

Component Instructions



1	DC IN 12V
2	Type-C OUT
3	HDMI IN
4	VGA OUT
5	HDMI OUT
6	HDMI OUT
7	USB 3.0
8	USB 3.0
9	USB 3.0
10	USB 3.0
11	LAN AV/POE
12	LAN MGMT
13	Line-In
14	Line-Out
15	Phone-In
16	TF Card Slot

© 2026 AISPEECH and/or its affiliates. All rights reserved