VoiceConnect Setting Guide

Revised：16 November 2026

1 System overview

VoiceConnect receives audio data from multiple VC-TR60A cameras and switches to the video source associated with the active speaker. By assigning cameras to different zones, the system extends coverage beyond the typical 8 m range of a single voice tracking camera.

This guide covers equipment requirements, seven room layouts, camera configuration, position calibration, output settings and commissioning checks. Configure the cameras first, register them on the OIP-N, calibrate each tracking zone, then verify switching and audio reference behavior.

Signal or function	System role
Audio detection	VC-TR60A cameras provide audio data for speaker detection.
Video acquisition	The OIP-N receives camera video through RTSP.
Output	OIP-N60D: HDMI and Virtual USB. OIP-N40E: Virtual USB only in VoiceConnect operation.

2 Equipment and firmware requirements

Device	Quantity	Minimum firmware
OIP-N60D	1 OIP-N device	OIB1.38 or later
OIP-N60D Dante AV-H	Alternative OIP-N device	OIC1.38 or later
OIP-N40E	Alternative OIP-N device	OIA1.38 or later
VC-TR60A	2 to 4, according to layout	VRK208 or later
VC-TR60A Dante AV-H or V1	According to layout	VRK300 or later

Note. Some layouts require OIP-N FW v1.39 or later. Apply both the device requirement above and the layout requirement in Section 4; the higher requirement takes precedence.

VoiceConnect requires a license for activation. Request a license before commissioning.

VoiceConnect Activation License Request Form

3 Installation planning

3.1 Network and acoustic coverage

- Connect the OIP-N and all VC-TR60A cameras to the same LAN.
- Install VC-TR60A cameras at a recommended height of 2 to 2.4 m.
- Maintain at least 2 m between each camera and its first row of seats. Additional layout-specific distances are shown in Section 4.
- The maximum voice tracking distance for each camera is 8 m. Multi-camera layouts extend overall room coverage by dividing the room into zones.
- Each camera must receive audio from all participants within its assigned tracking area. Use the OIP-N60D audio reception display to check coverage.
- The panoramic lens field of view must cover the camera's assigned tracking area.

Note. VoiceConnect is not supported in a Voice Lift environment. Only one OIP-N should run VoiceConnect on the same LAN at a time.

3.2 Select a supported layout

Section	Room layout	Cameras	OIP-N firmware
4.1	Standard rectangular room	2	v1.38 or later
4.2	U-shaped room — perimeter layout	3	v1.38 or later
4.3	Standard rectangular room — audio judge layout	3	v1.38 or later
4.4	U-shaped room — central layout	3	v1.39 or later
4.5	Long table — front camera and rear cross coverage	3	v1.39 or later
4.6	Long table — front and rear cross coverage	4	v1.39 or later
4.7	Square room — central layout	4	v1.39 or later

3.3 Configuration sequence

1. Select the layout and verify firmware and licensing requirements.
2. Position the cameras and confirm audio and panoramic coverage.
3. Configure every VC-TR60A as described in Section 5.
4. Enable VoiceConnect, register the cameras and calibrate positions using Section 6.
5. Configure outputs, voice triggers and audio reference routing using Sections 7 to 9.
6. Complete the validation checks in Section 10 before handover.

4 Supported room layouts

4.1 Standard rectangular room with two cameras

Cameras: 2 | Required OIP-N FW: v1.38 or later

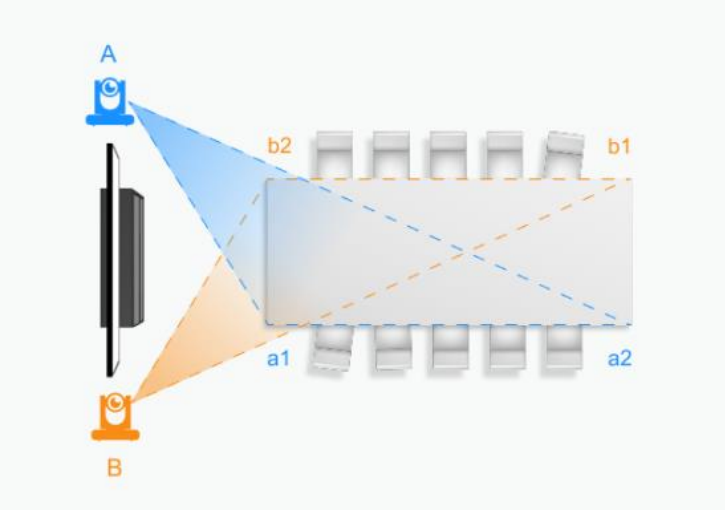


Figure 4.1 Camera positions and calibration boundaries

Install Cameras A and B near the display wall. Aim them diagonally across the table so that each camera captures the opposite side. This reduces side-profile views.

Note. Calibration point a1 must be at least 2 m from Camera B; point b2 must be at least 2 m from Camera A.

4.2 U-shaped room with perimeter cameras

Cameras: 3 | Required OIP-N FW: v1.38 or later

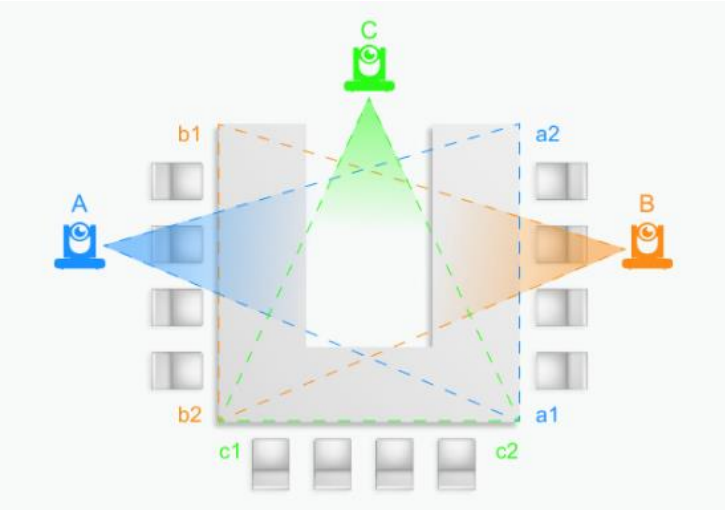


Figure 4.2 Camera positions and calibration boundaries

Use three cameras for the three seating directions. Cameras A and B capture the opposite seating sides; Camera C captures the seats facing the display. VoiceConnect selects the camera facing the active speaker.

4 Supported room layouts continued

4.3 Rectangular room with an audio judge camera

Cameras: 3 | Required OIP-N FW: v1.38 or later

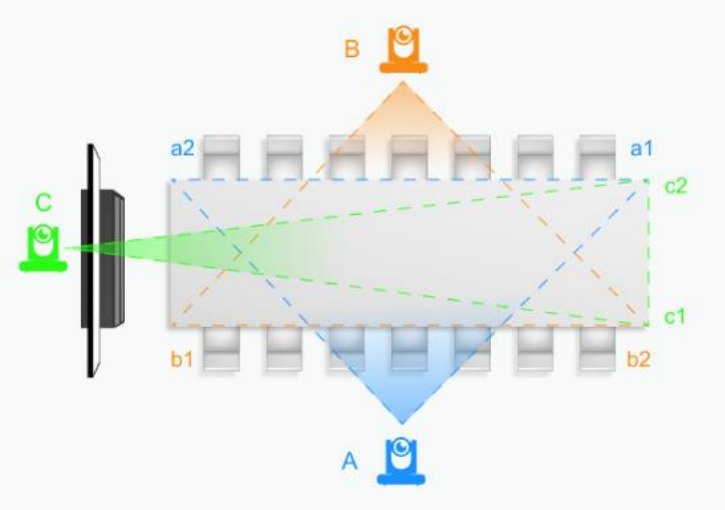


Figure 4.3 Camera positions and calibration boundaries

Mount Cameras A and B symmetrically on opposite sides of the table. Camera C is positioned at the end near the display and acts as the audio judge to identify the speaking side. Camera C does not provide the selected video output in this layout.

4.4 U-shaped room with central cameras

Cameras: 3 | Required OIP-N FW: v1.39 or later

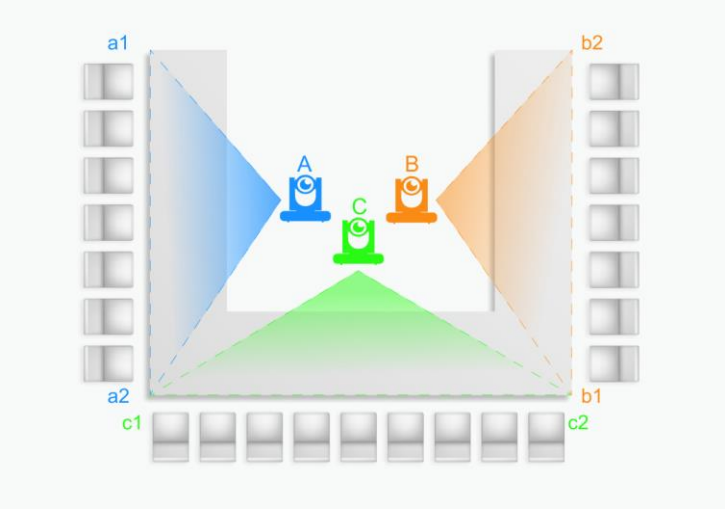


Figure 4.4 Camera positions and calibration boundaries

Install Cameras A, B and C together in the central area, each facing its assigned seating side. VoiceConnect switches to the camera facing the active speaker.

4 Supported room layouts continued

4.5 Long table with three cameras

Cameras: 3 | Required OIP-N FW: v1.39 or later

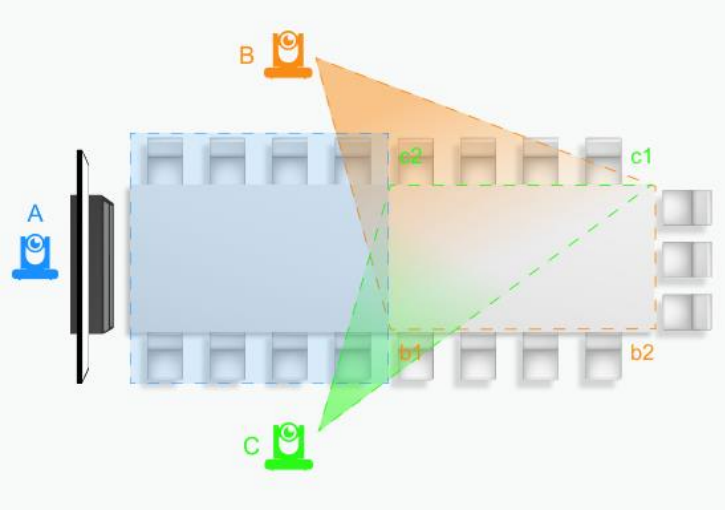


Figure 4.5 Camera positions and calibration boundaries

Camera A covers the front zone. Cameras B and C cross-cover the rear zone. Camera B captures the chairperson side. Dividing coverage between front and rear zones extends room coverage compared with the standard rectangular layout.

Note. Calibration point b1 must be at least 2 m from Camera C; point c2 must be at least 2 m from Camera B. Camera B and C need to face the opposite seat as directly as possible. The reason is to get all the audio in front zone and rear zone. Camera B and C panoramic view no need to shot all front zone area.

4.6 Long table with four cameras

Cameras: 4 | Required OIP-N FW: v1.39 or later

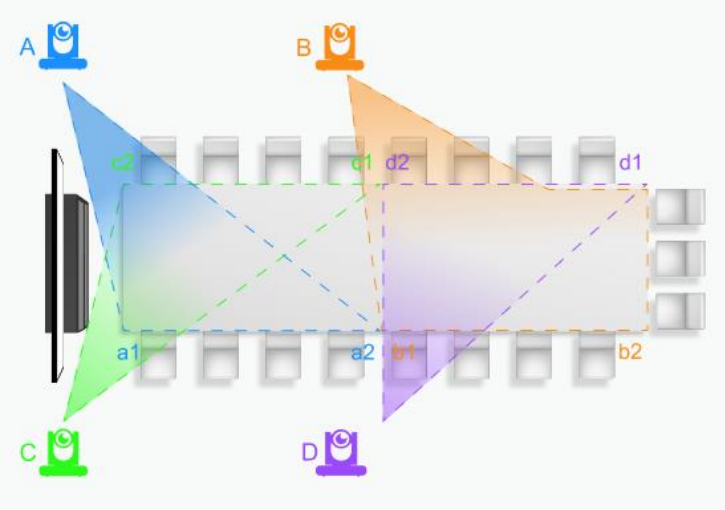


Figure 4.6 Camera positions and calibration boundaries

Cameras A and C cross-cover the front zone; Cameras B and D cross-cover the rear zone, as illustrated. Camera B captures the chairperson side.

Note. Minimum 2 m separation: c2 from Camera A; a1 from Camera C; b1 from Camera D; d2 from Camera B. Camera B and D need to face the opposite seat as directly as possible. The reason is to get all the audio in front zone and rear zone. Camera B and D panoramic view no need to shot all front zone area.

4.7 Square room with central cameras

Cameras: 4 | Required OIP-N FW: v1.39 or later

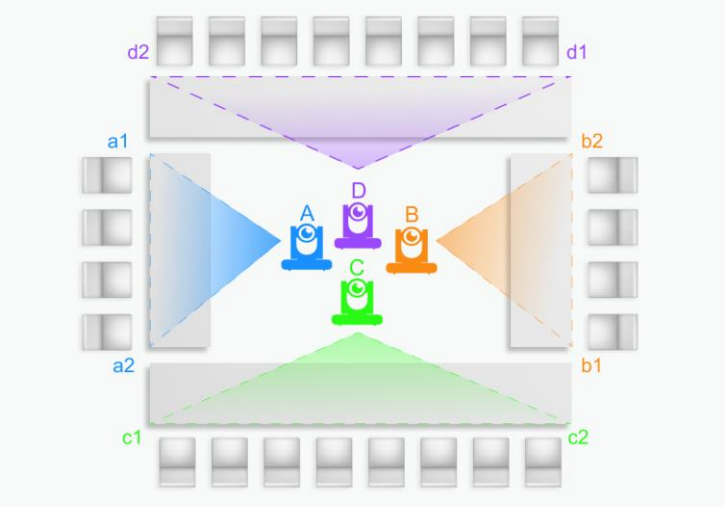


Figure 4.7 Camera positions and calibration boundaries

Install Cameras A, B, C and D together in the central area, each facing one of the four seating sides. VoiceConnect switches to the camera facing the active speaker.

4.8 Confirm the installation before configuration

1. Verify that the physical camera labels match the selected layout diagram.
2. Confirm the OIP-N firmware meets both the model and layout requirements.
3. Check that each panoramic view includes its full tracking zone.
4. Check audio reception at the nearest and farthest seats in each zone.
5. Identify the calibration boundary points shown in the selected layout.

Note. The detailed configuration example in Section 6 uses the three-camera U-shaped perimeter layout. For other layouts, follow the corresponding diagram and on-screen calibration labels.

5 Configure each VC-TR60A

Complete the following procedure on every camera before configuring the OIP-N.

5.1 Disable Multi-Voice Framing

1. Open the VC-TR60A web interface and go to [Tracking].
2. Disable [Multi-Voice Framing].



Figure 5.1 Tracking controls and Multi-Voice Framing

5.2 Define block areas and enable voice tracking

1. Open the tracking advanced settings.
2. Define [Block Area] regions where tracking is not required, then click [Done].
3. Enable [Voice Tracking]. Repeat this procedure for all cameras.

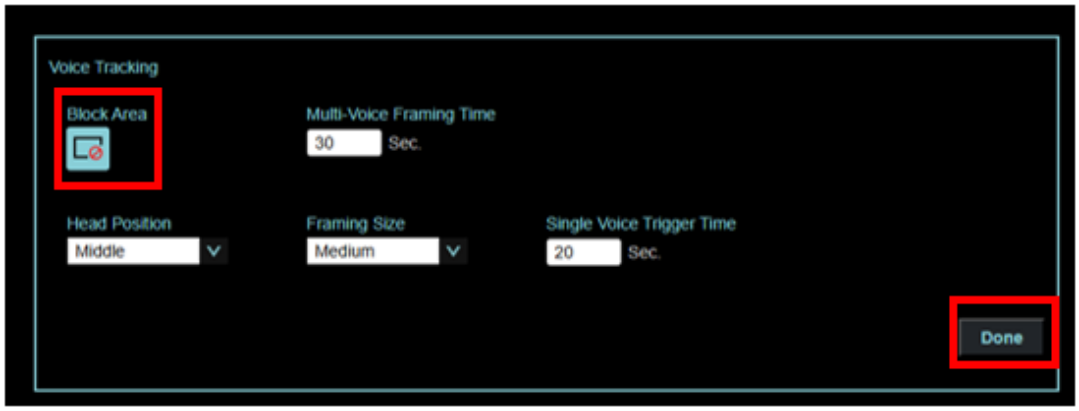


Figure 5.2 Block Area and framing controls

6 Configure VoiceConnect on the OIP-N

6.1 Enable VoiceConnect and select the layout

1. Open the OIP-N web interface and go to [System] > [Settings] > [VoiceConnect].

2. Enable [VoiceConnect Mode]. Select the mode corresponding to the physical layout, then click [Apply].



Figure 6.1 VoiceConnect mode and room layout selection

6.2 Register camera addresses and streams

1. Enter the IP address and RTSP stream URL for each VC-TR60A.

2. Assign camera entries according to the selected layout diagram. In the U-shaped perimeter example, A is on the right of C and B is on the left of C when viewed from Camera C toward the seating area.

3. Click [Apply]. Confirm that all camera status indicators are green. If an indicator is not green, verify the camera connection and RTSP URL.

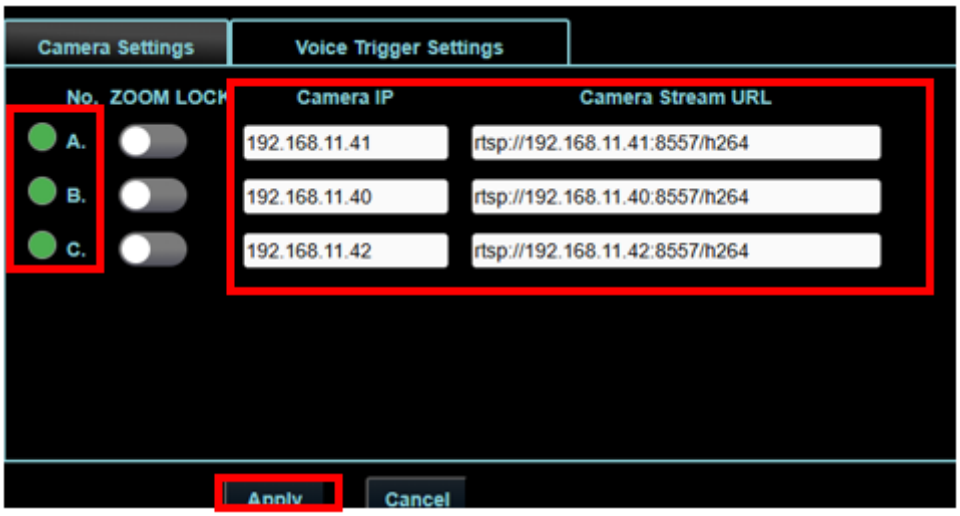


Figure 6.2 Camera registration and status indicators

6 Configure VoiceConnect continued

6.3 Calibrate relative camera positions

For the three-camera U-shaped example, calibrate a1, a2, b1, b2, c1 and c2. Use the boundary labels displayed for the selected mode.

1. At each calibration point, generate a continuous sound for 5 to 10 seconds. Position the point at least 50 cm beyond the actual tracking-area boundary, as illustrated.

2. Wait until the audio level is stable and verify that the detected angle changes. Click [Set] for the corresponding point.

3. Repeat for every calibration point.

4. Verify x1 > x2 for each camera, where x is a, b or c in this example. If the condition is not met, check camera orientation and calibration positions, then recalibrate.

5. After verifying the values, set [Auto Tracking] to [On] and click [Apply].

Note. If the detected angle of nearby cameras remains at 230, the camera has not detected the calibration sound. Do not save that point. Check the sound source and audio reception before continuing.

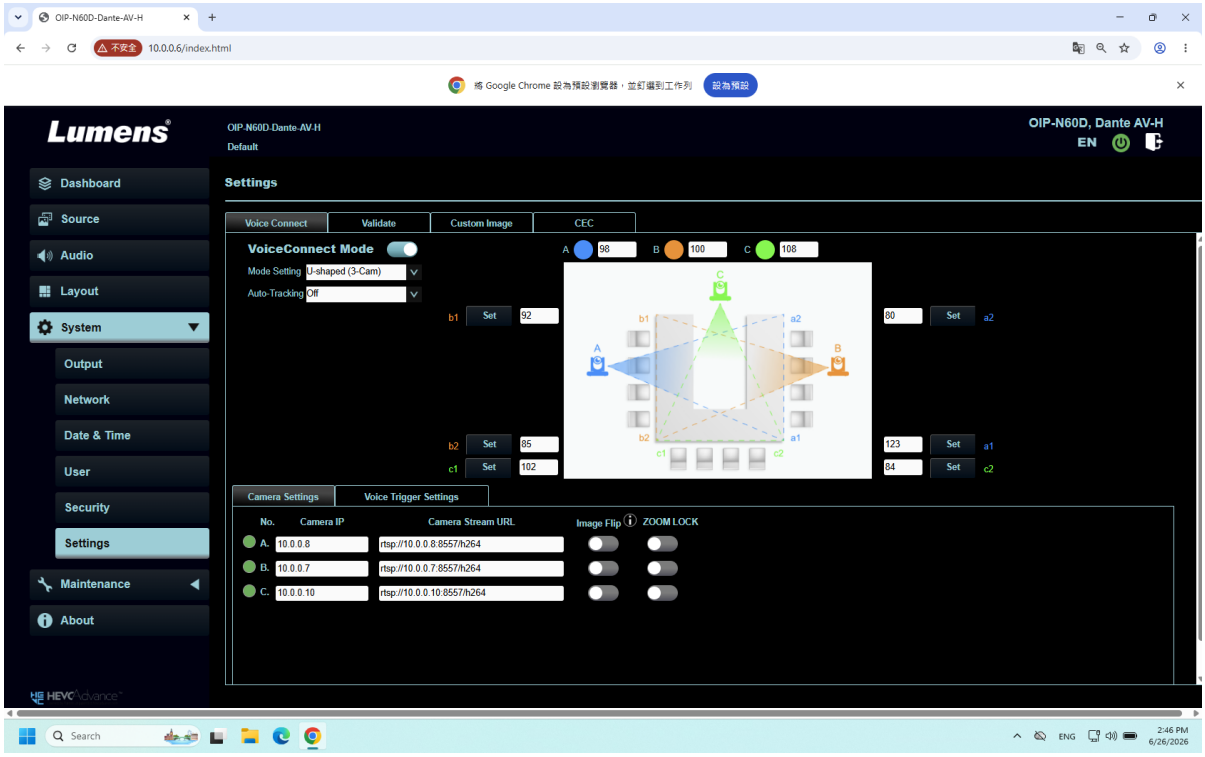


Figure 6.3 Position calibration on the VoiceConnect page

Note. Calibration values are installation-specific. Do not copy the numerical values shown in the example screenshot.

7 Configure video output and switching

7.1 Enable the required video output

Device	VoiceConnect output capability
OIP-N60D	HDMI: up to 4K60. Virtual USB video: up to 1080p30.
OIP-N40E	Virtual USB video: up to 1080p30. HDMI output is not available for this application.

To enable Virtual USB, go to [System] > [Output] and enable [Virtual USB Out].

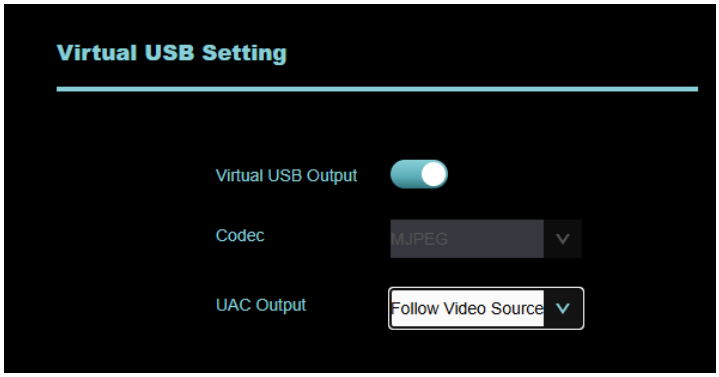


Figure 7.1 Virtual USB output

7.2 Configure voice trigger behavior

Open [Voice Trigger Settings], configure the parameters below, and then click [Apply].

Parameter	Function and setting
Trigger Delay Time	Sets the delay used for camera switching. Default: 1.0 s. A setting of 0.5 s provides faster switching.
No Voice Trigger Action	Select [Specify] to use a designated video source after a period without speech.
No Voice Duration	Sets the no-speech interval before the specified action is triggered.
Screen Stream URL	Sets the source shown when no one is speaking, such as a VC-TR60A panoramic stream or another camera RTSP stream.

Note. The web interface may label the URL field [Screen Switch URL]. Identify it by its associated function: no speech, audio reference or conversation mode.

8 Configure audio reference and conversation mode

8.1 Audio reference behavior

Use an audio reference to reduce incorrect camera switching caused by far-end speech reproduced through the room loudspeakers. Open [Voice Trigger Settings], configure the following parameters and click [Apply]. Complete the physical routing in Section 9.

Parameter	Function
Audio Reference	Select [UAC], [Audio in] or [Dante] to match the reference connection.
Detection Sensitivity	Choose [High], [Medium] or [Low], in descending sensitivity order.
Screen Stream URL	Sets the video source shown while the far-end participant is speaking.

8.2 Conversation mode

Conversation mode selects a specified video source when repeated camera changes indicate an ongoing conversation. Configure the parameters below and click [Apply].

Parameter	Function
Conversation Detection	Enables or disables conversation mode.
Screen Switch URL	Specifies the video source displayed in conversation mode, such as a panoramic camera RTSP stream.
Multi-Voice Frequency / Multi-Voice Time	Enter X switches within Y seconds. Example: 3 camera switches within 30 seconds activate conversation mode.
Multi-Voice Release Time	Conversation mode ends when one person speaks continuously for Z seconds, or when no one speaks for Z seconds. Minimum: 10 seconds.

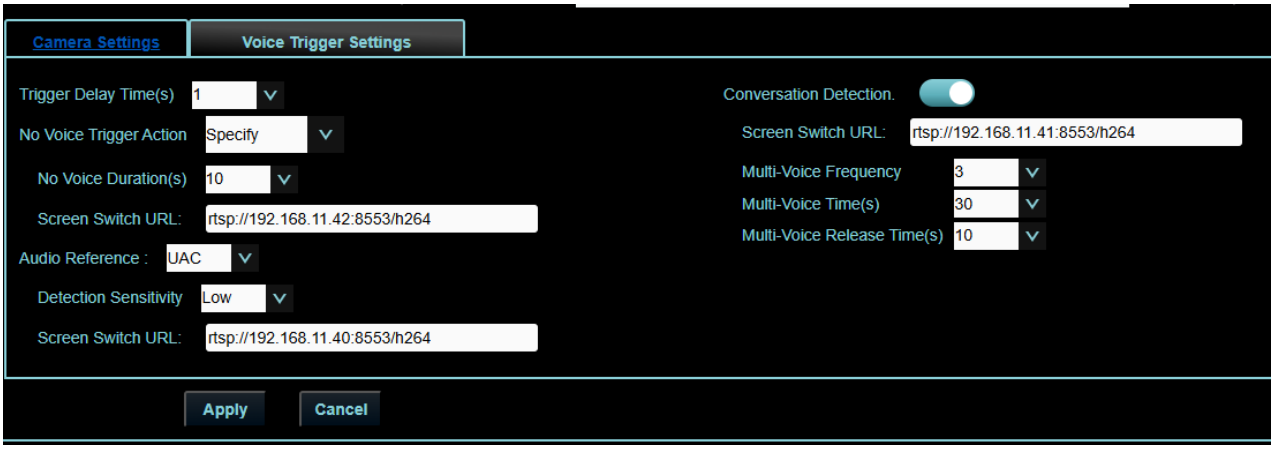


Figure 8.1 Voice trigger, audio reference and conversation parameters

9 Connect the audio reference source

9.1 Dante reference

1. On the OIP-N, go to [System] > [Output].
2. Enable [Virtual USB Output] and set [UAC Output] to [Follow Dante].
3. In Dante Controller, route the Dante DSP transmitter channel to the OIP-N receiver channel.
4. In [Voice Trigger Settings], select [Dante] as the audio reference source.

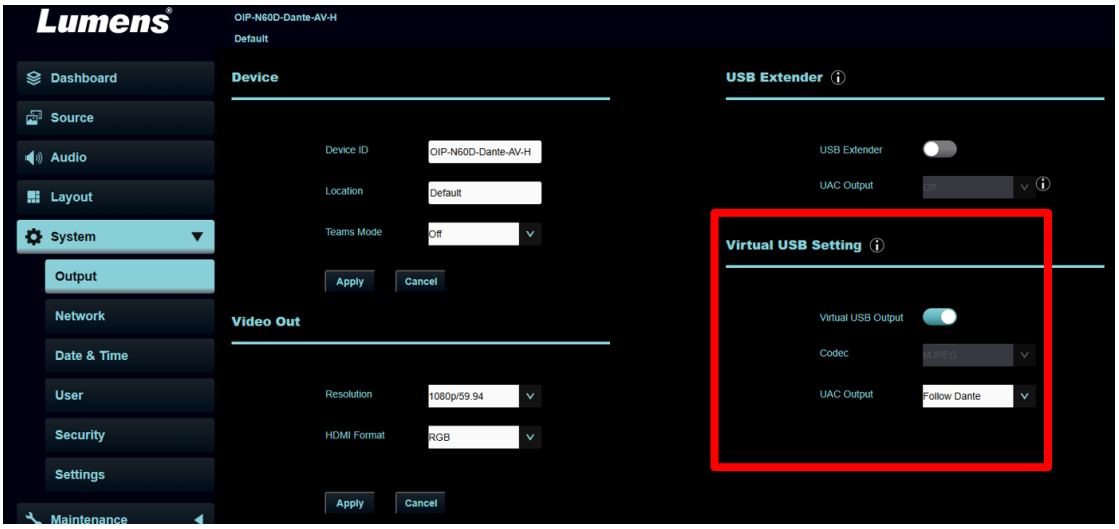


Figure 9.1 OIP-N output settings for a Dante reference

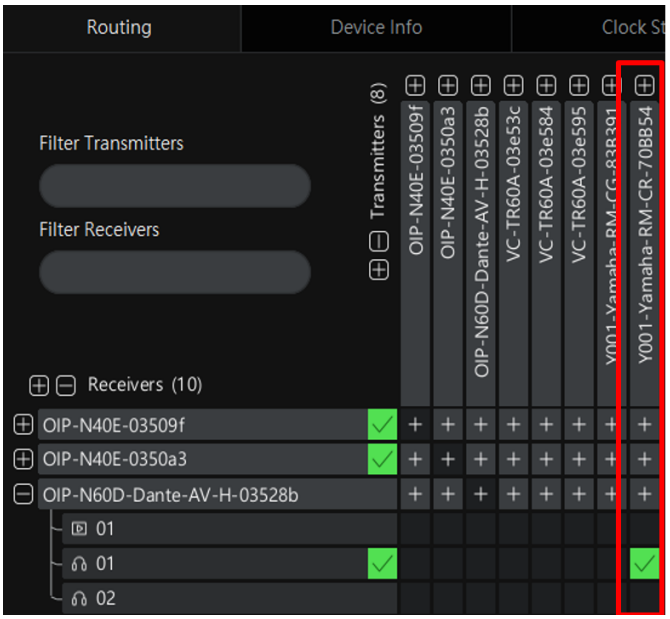


Figure 9.2 Dante Controller subscription from DSP to OIP-N

Note. The Dante audio reference connection is separate from the camera video stream. VoiceConnect camera video uses RTSP; Dante AV-H video streams are not supported in this guide.

9 Connect the audio reference source continued

9.2 USB audio reference

1. On the OIP-N, go to [System] > [Output] and enable [Virtual USB Output].
2. Set [UAC Output] to the required option: [Follow Video Stream], [Follow Dante] or [Follow Routing].
3. Connect the OIP-N to the PC using a USB cable.
4. Select the OIP-N as the PC audio output device.
5. In [Voice Trigger Settings], select [UAC] as the audio reference source.

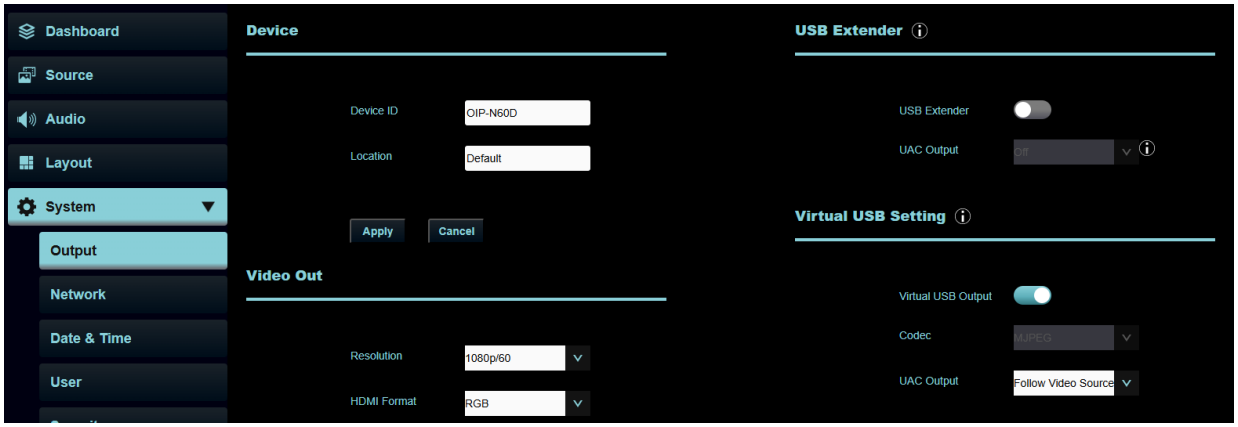


Figure 9.3 Virtual USB and UAC output settings

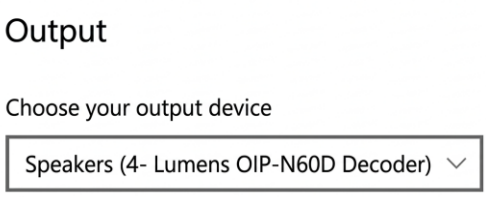


Figure 9.4 Select the OIP-N as the PC audio output device

9.3 Analog audio reference

1. Connect a 3.5 mm audio cable from the DSP output to the OIP-N audio input.
2. In [Voice Trigger Settings], select [Audio in] as the audio reference source and click [Apply].

10 Validate the installation

10.1 Verify the selected source

Open the [Validate] tab. The green camera indicator identifies the currently selected video source. Speak from each tracking zone and verify that the indicator and actual video output correspond to the intended camera.

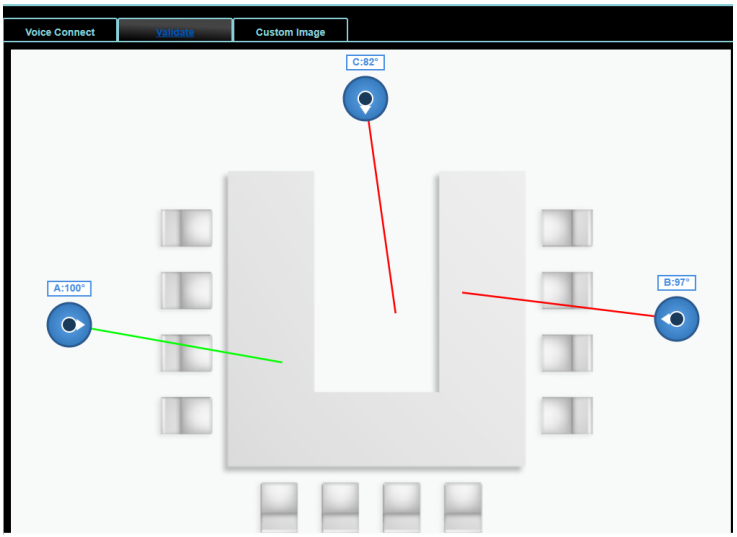


Figure 10.1 Validate view and active camera indication

10.2 Commissioning checks

1. All configured cameras show green connection indicators.

2. Voice Tracking is enabled on every VC-TR60A and VoiceConnect Auto Tracking is On.

3. Calibration angles respond to the sound source; no saved point remains at 230. The x1 > x2 condition has been checked for each camera in the three-camera example.

4. Test speech at representative seats in every zone. Verify the camera choice, framing and output.

5. Stop speaking and verify the configured no-voice action and source.

6. Play far-end speech through the room loudspeakers and verify audio reference behavior.

7. If conversation mode is enabled, test its activation and release using the configured thresholds.

8. Verify HDMI or USB output on the connected display or PC, as applicable.

10.3 Operating limitations

- Camera video input supports RTSP only; Dante AV-H video streams are not supported.
- Only one OIP-N may run VoiceConnect on the same LAN at a time.
- VoiceConnect is not supported in a Voice Lift environment.

11 Troubleshooting and frequently asked questions

11.1 The camera does not track or switch to the speaker

1. Confirm [VoiceConnect Mode] is enabled on the OIP-N.

2. Confirm every configured camera status indicator is green.

3. Confirm [Voice Tracking] is enabled on every VC-TR60A.

4. Set [Last-in Switch] to [Off].

5. Repeat calibration using Section 6.3. Verify changing angle values, reject values fixed at 230 and check the calibration point order.

6. If the problem remains, reboot the OIP-N and repeat the validation checks.

11.2 How do I change the framing size

On the VC-TR60A, go to [Tracking] > advanced settings. Select [Framing Size], choose [Tight], [Medium] or [Wide], then click [Done].

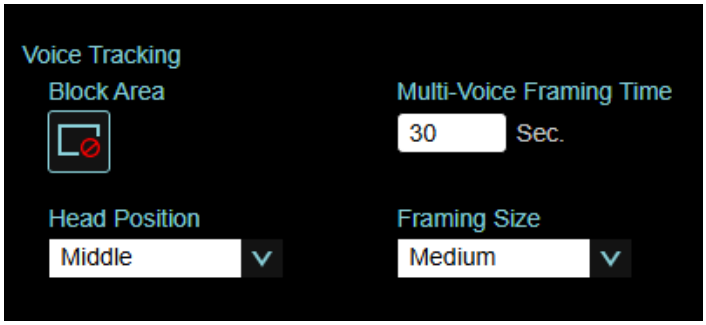


Figure 11.1 Framing Size selection

11.3 How do I reduce switching delay

Adjust [Trigger Delay Time] in the OIP-N web interface. The default is 1.0 s; reduce it to 0.5 s for faster camera switching.

11.4 How do I prevent switching caused by far-end audio

Select and configure the correct audio reference source using Sections 8.1 and 9. Confirm that the reference connection matches the selected input, then test with far-end speech reproduced through the room loudspeakers.

11.5 What is the Validate tab used for

It displays the currently selected video source. Use the green camera indicator together with the actual video output to verify source selection for each tracking zone, as described in Section 10.1.