SEERVISION
BEST PRACTICES

Index

Introduction

This 'best practices' guide provides information and recommendations for designing and setting up a training room, boardroom, or multipurpose space with Seervision audio-based speaker tracking together with Shure's MXA920 microphones. The guide covers topics such as room design considerations, components, and FAQs to help you optimize the implementation of your Seervision system.

Overview of Seervision

The Seervision Suite, at its core, is a software solution to automate cameras powered by computer vision. It analyzes the full-body characteristics of the people in the frame and predicts what will happen in the future for the person being tracked. The Seervision Server then translates this information into multiple commands per second for the camera and lens trajectories. The system is built specifically for a wide range of third-party PTZ cameras, so that you can run autonomous presenter tracking with 'trigger zones' around the space that activate certain shots or actions fully automatically, based on where the presenter stands.

When we talk about Seervision for meeting spaces, we typically also leverage 3D audio data from specific ceiling array microphones such as the Shure MXA920. By processing the positional information from the active speaking participant(s), with Seervision, you can enable dynamic single or multi-camera workflows that point and switch adaptively to participants in the room – *without having to depend on presets*. This means that the seating configuration or room layout can change and the cameras adapt seamlessly.

Through the Seervision API, you can also enable the Seervision system to work alongside other API-enabled AV devices. Connected devices such as a mixer allow you to dynamically switch between camera feeds, based on custom logic and Seervision's scene understanding. Other API-enabled devices, such as a touchpanel, can enable users to switch between different automation workflows, for example, to enable or disable group framing, or simply turn the speaker tracking on/off.

Room Design Considerations

Size and Layout of the Room

The size and layout of the room are important considerations when planning for a Seervision-equipped space.

Seervision's audio tracking solution is meant for larger spaces (typically for 20+ people), not huddle rooms, or small meeting spaces where the needs can be covered with standard out-of-the-box solutions.

Where Seervision's solution shines is in large spaces with a need for multiple cameras, especially in multi-purpose or "high-impact" rooms with U-shape, group-based, or row-based seating configurations; those that are divisible, have flexible layouts and use-cases or changing seating positions.

The room can also have a presenter area that leverages Seervision's video-based auto-tracking. When designing the room, always ensure that there are clear lines of sight, so that the cameras can see above participants' heads to cover those at the back.

Camera Placement

When planning for camera placement, it's important to consider the number of cameras needed, as well as their location and angle. The number of cameras you need will depend on the size of the room, the seating configuration, the number of participants, and your hybrid meeting preferences. A general rule of thumb is to have one camera dedicated to each section of the room, or two cameras covering the space diagonally, as shown below.
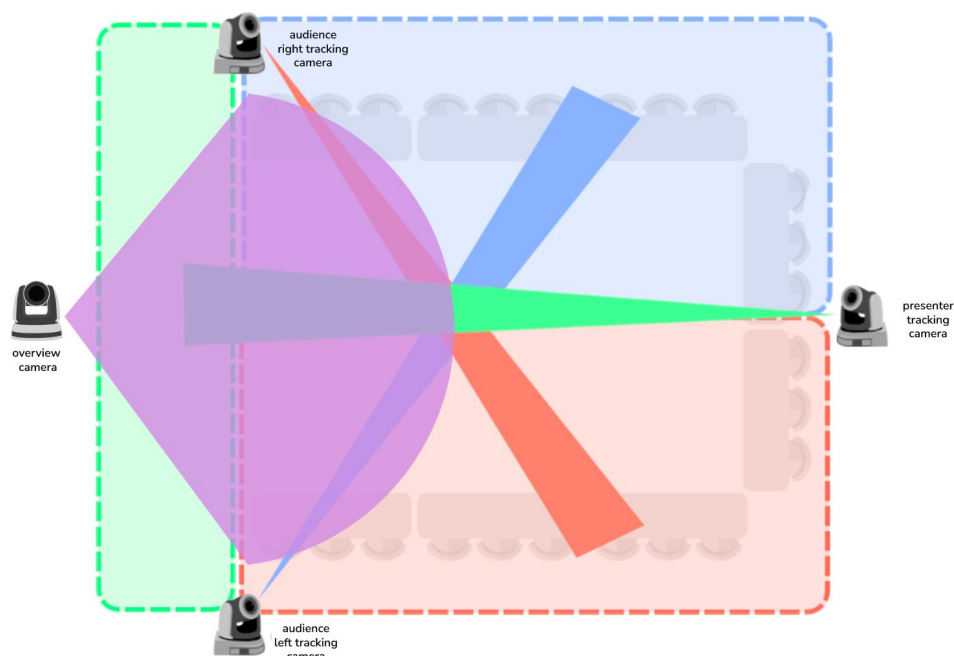
For example, for a large U-shaped desk layout with three sides, we recommend having three audience cameras, one pointing at each side of the U. If you also have a presenter area, a fourth camera can be used to point at the front of the room where the screen or projector area might be.
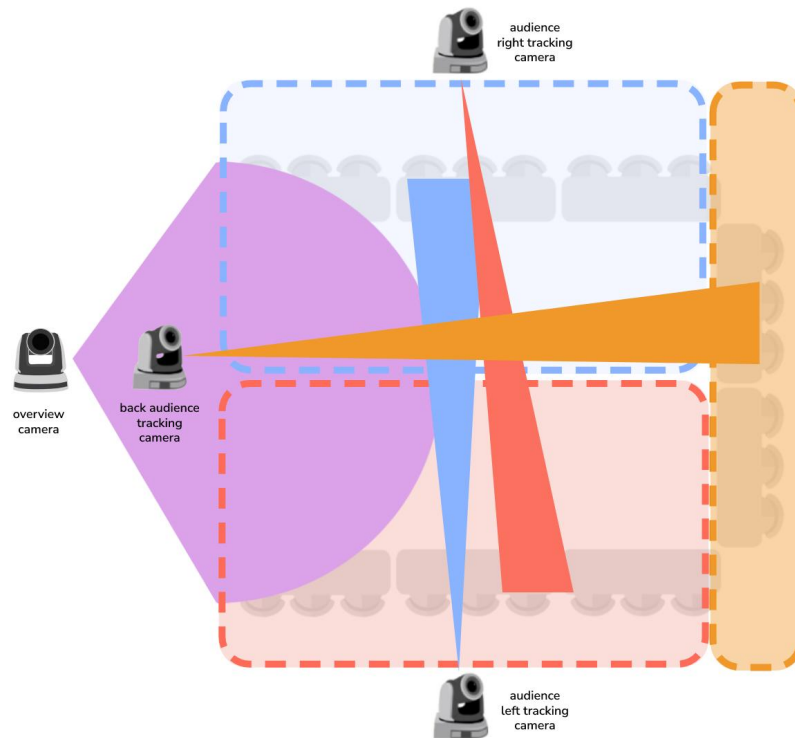
TIP: We typically also recommend that you have an overview camera to show all the participants when everyone is quiet. This camera is especially useful if you also want to 'hide' the movement of the speaker tracking cameras when they are transitioning between speakers. We can send commands via the API about when the camera's movement has begun and ended. In doing so, you can switch to the overview camera while the speaker camera is moving, and once it finishes pointing at the active speaker, you can safely switch back to it. It can also serve as the "chaos backup" for when many people speak at the same time and the audio is indistinguishable.

NOTE: An overview camera doesn't necessarily need to be automated by Seervision; it can be just a static feed that can be switched to as a safety shot. Most of the time, this can be a lower-end wide angle USB camera that is fed directly into the mixer.

As for camera compatibility, you should always check the latest version of the Seervision integrations page. The typical rule of thumb is: the more expensive the camera, the better the level of integration will be (due to the motors, drivers, and frequency of commands it can receive). Those cameras with FreeD also perform better than their non-FreeD counterparts, as we receive more frequent positional feedback.

Keep in mind that for Panasonic PTZ cameras, Seervision has full control over the preset motion and can tune for the trajectory and velocity – these are the smoothest for speaker tracking if you want to show the camera's movement 'live'. For the rest of VISCA over IP cameras, Seervision has less control over the internal presets of the camera, and therefore, most of the time they are not smooth enough to show live. This is why we recommend an overview camera to 'hide' the camera transitions whilst the speaker tracking cameras are in motion.

When choosing a camera, you should also keep in mind its optical zoom level. For covering audience members at the back of a long room, we recommend using a camera with at least 20x optical zoom.

Microphones

In addition to cameras, the number (and placement) of ceiling array microphones is also important to consider. You'll want to let your Shure representative assist you with the number of ceiling array microphones needed and their placement. We recommend having at least one microphone covering every 4.5 x 4.5 meters of space. Coverage areas larger than this typically lead to more inaccuracies in pointing the camera directly at the active speaker, leading to off-centered framing around the edges of the microphone's coverage area. Additionally, for the positioning, verify together with your Shure representative that the microphone panels' orientation in the room is parallel (or perpendicular) to the mounting direction of your PTZ cameras.

TIP: Follow Shure's advice for better results when integrating the MXA920 with camera tracking:

- Talk toward the microphone
    - When possible, ask talkers to speak in the direction of the microphone. The microphone may take longer to send talker positions if they are facing 90 to 180 degrees off-axis from the microphone.

- Speak up
    - Ask quieter talkers to speak up. The MXA920 can sometimes take longer to send positional information from quieter talkers.

- Move closer for single-talker framing
    - If you need to frame 1 talker at a time with the camera, set things up so that the microphone is about 10 feet (3 meters) from the different talker positions. The MXA920 performs well with talkers that are up to 16 feet (4.9 meters) from the microphone, but moving closer helps cameras with single-talker framing.
    - Position talkers so they are no more than 90 degrees off-axis from the microphone.

- Enable Automatic Coverage
    - While Seervision's audio tracking works without Automatic Coverage enabled, it disables tracking outside of the lobes, so we recommended it is enabled by default.

It is worth noting that this preset-less audio tracking with Seervision is currently only possible with the Shure MXA920. This is due to the [command strings](#) available, the built-in echo and active-noise cancellation that separate voices from other noise, and most importantly, the accuracy and frequency of the speaker location data which enables the dynamic speaker pointing without needing presets.

Currently, for any other model apart from the Shure MXA920, such as those from Sennheiser, Biamp, Audio-Technica, Yamaha, or even the MXA910, the Seervision Suite can only do preset-based speaker tracking, as opposed to the *preset-less* dynamic speaker tracking with the MXA920. Further ceiling array microphones will be integrated by Seervision throughout 2023 with the intention of achieving the same functionality as with the MXA920.

Acoustics

Room acoustics play a critical role in ensuring the audio location information is as accurate as possible. The better the quality of the positional data, the more precise the camera can be when 'landing' on the active speaker. That is why it is important to minimize echo and 'audio reflections' by including sound-absorbing elements like carpeting, acoustic panels, or fabric-covered walls, whilst avoiding hard surfaces like concrete, glass, or metal that can bounce sound around. Static noise, such as the one produced by a server rack, can also affect the quality of the audio and therefore also the positional data Seervision receives.

Components

#	Item	Notes
1	Seervision Server(s)	Dedicated server(s) supplied by Seervision. Each SVs4-2U rack mount can automate up to two cameras simultaneously. These can be placed in the server room. No connection to the internet is required except for updates. The servers need either SDI, NDI, or NDI|HX as an input from the cameras to analyze the scene. Seervision communicates with the cameras and microphones via IP. The Seervision Server does not manage the switching nor processes the video, it is only the automation platform for commanding the cameras and (if desired) any other AV device through Node-RED.
2	PTZ Camera(s)	Panasonic PTZs are recommended for smoothest results. The full list of integrations is [here](#). Most camera models have the ability to be powered through POE+.
3	PTZ camera power supply	Only needed if POE+ switch isn't used.
4	Shure MXA920(s)	Must be on the same network as the Seervision Server and cameras. The communication is not Dante, it's a direct TCP/IP connection.
5	CAT5e network cables	Mutliple needed. Length depends on room dimensions and design.
6	Network Switch	POE+ if the camera needs to be powered through that. In case the set-up is done in an already existing network, there is no need for a new switch.
7	Laptop/Desktop client computer for the Seervision User Interface	The user computer (can be a laptop or desktop) on Windows/Mac/Linux OS, and the only requirement is to have Google Chrome installed. This is needed at least for the setup and configuration and needs to be on the same network as the camera(s) and server(s).
9	3G SDI cables	To connect the camera to the server. If the set-up is NDI-based, there is no need for SDI cables.
10	Tape Measure or Laser Distance Meter	Used for the calibration process to measure the microphone's position coordinates accurately.

Additional Components

#	Items (optional)	Notes
1	Video Mixer	Seervision does not provide a video switching feature. For Seervision to control the switching logic, a video mixer with an API that Seervision can interface with is needed.
2	Overview camera	Optional camera (highly recommended!) that can be switched to during periods of silence or to hide the PTZ camera transitioning from one speaking participant to another. This camera doesn't need to be controlled by Seervision, but it would need to feed into the mixer.
3	SDI/NDI to USB Converter	In the case that you're feeding the final output into a meeting computer like a Zoom Room / Teams Room PC.

Other items such as a DSP, Touchpanel, Meeting Room Computer, etc, are typically also needed but not covered here as they are out of scope and aren't critical in making the Seervision solution functional.

Example Diagram
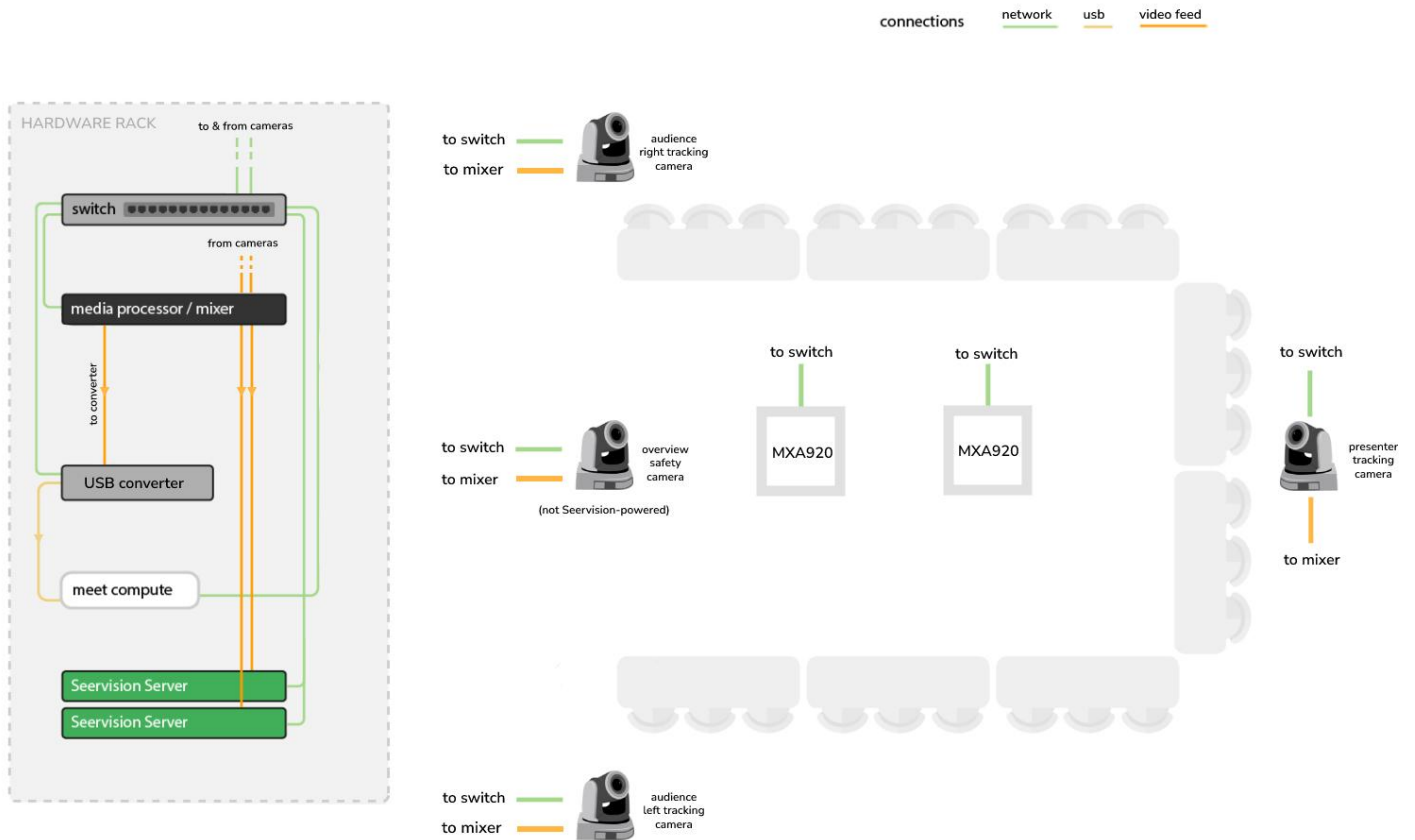
Note: The Seervision Server's video input should be either in SDI, NDI or NDI|Hx. For QSC-based setups, we use RTSP as a video input into the server.

Installation

For further details on setting up and connecting a Seervision Server to a camera and microphone, you can read this guide on our support page.

FAQs

What inputs and outputs do I need?

For Seervision to work, we need the video feed from the camera to run the computer vision analysis. This can be via SDI, NDI, NDI|Hx, or RTSP. We also need a network connection between the server and the camera. The other thing needed is a power cable, it comes in the box with the Seervision Server.

Any other device that you wish to control and automate via Seervision's Node-RED instance (microphones, mixer) should also be on the same network or AV subnet as the server.

All other ports (HDMI, display ports) are disabled and can't be used for anything. The interface is accessed via IP through Google Chrome.

Apart from just controlling the camera's movement, Seervision doesn't do any other processing. This means no additional delay is added, and the image quality isn't worsened.

| 1 | Network | 2 | SDI Inputs | 3 | Power |

How do I feed the video into my Zoom Room PC / Meeting Computer?

You will need to convert the video signal into a USB input for the meeting room computer. If you're using NDI, you'll need an NDI to USB Webcam converter. If you're using SDI, you'll need an SDI to USB converter like the AJA U-TAP SDI or Inogeni's Toggle USB 3.0.

It is worth noting that in cases where you'll be mixing the feeds with a switcher and want one single output, we recommend using the 'Master Out' from the mixer and feeding it into the USB converter directly.

How does Seervision communicate with the microphone's audio?

In the case of the Shure MXA920, we communicate with the microphone via the network. The audio does not need to be processed by a DSP for this to work, Seervision receives the raw positional audio data directly and filters it.

How does the automated switching work?

Seervision doesn't provide a switching functionality. However, you can configure your Seervision Server to communicate via our API with third-party mixers/switchers like a Blackmagic ATEM, or similar, to configure a logic for when they should switch according to your meeting preferences. Integrators can create a simple workflow with the help of Seervision that tells the mixer when it has to switch based on when Seervision receives a signal from the microphone, or when a camera movement has begun/ended. This switching logic can be set up on Seervision's [Node-RED](#) instance, accessed on port 1880. This is where the computation for the ceiling microphone data is already done, and the switching logic can be built on top of the existing flow diagrams.

Creating a switching logic is especially useful if you have multiple cameras and want to create a single output instead of feeding each camera individually into your meeting room computer. We recommend choosing a broadcast mixer that can "clean switch" the feeds (i.e. the Blackmagic ATEM line) between frames so that no black frames are shown when switching between cameras.

Alternatively, integrators can create the switching logic on any other platform of their choice, like Crestron, AMX, Q-SYS, etc. Keep in mind that we only provide the template code for communicating with the microphones via Node-RED, so a simple bridge will have to be written.

NOTE: When using a Blackmagic Mixer like the ATEM, Seervision will need to communicate with it via its [Bitfocus Companion](#) plug-in. This can be loaded on the Seervision Server itself, however, it must be communicated to Seervision in advance of the installation.

What types of speaker framing can Seervision do?

Seervision can do two types of speaker framing:

Individual speaker framing: whereby the camera automatically and dynamically points toward the active speaker. This is done based on the raw 3D positional data of the MXA920, meaning that no presets need to be created for where you expect people to sit, instead, the system can adapt to where people are even if the seating layout changes.

If two people are having a conversation, the camera will pan between the speakers quickly, although this can become annoying after some time. Alternatively, if you have two cameras covering people who are at opposite ends of the table, Seervision can send commands to the mixer about when it has to switch between the cameras based on who is speaking.

Group framing: This is the alternative to not having to quickly pan between active speakers. In using group framing, Seervision takes a 'bubble' of the active speakers and frames them all in the same shot (assuming the FOV of the camera is wide enough). Each speaker has their own bubble, and if multiple people speak, their bubbles are combined to frame them all at once. As a speaker stops talking and another speaker takes over, the bubble dies down, and the system adjusts the framing to cover the active speakers. If there is silence for a custom amount of time, the system can then be configured to show an overview shot of the room. Additionally, with the "active speaker focus" feature, you can have group framing as default, but configure the system to zoom-in on the active speaker after 6 seconds of them talking, so that you can see them close up.

Can Seervision individually crop participants like Zoom's gallery view?

This is not something currently available as the system only controls the camera's movement and the Seervision Servers don't output video. Based on our computer vision, we can however detect how many people there are, and where they are in the frame. We can already send out that information via our private API for other systems to do cropping or any other layout with that information. In the meantime, we are exploring ways to individually crop participants via the Seervision Server itself.

Is there a limit to the number of cameras or microphones you can use together?

No, we haven't found a limit. As long as for every two cameras that you want to automate, you have a Seervision Server with the right specifications and an active license for each of the cameras, Seervision will be able to automate them. As for microphones, there is also no limit. In fact, the more microphones, the better the coverage will be, and therefore the positional data of the active speakers will be more accurate.

Further Reading

Supplementary resources:

- [Seervision Audio Tracking Setup Guide](#)
- [Seervision for Corporate - One Pager](#)
- [Seervision System Overview](#)
- [Seervision Manual](#)

Videos:

- [Seervision 9 Person U-Shape Demo](#)
- [Seervision + Shure Webinar](#)
- [Seervision 'Frame on Landing'](#)

DISCLAIMER: This guide is not definitive and shouldn't be relied on as the only point of information for designing your room. Always double-check specifications and designs with Seervision, third-party manufacturers of your other AV devices, and of course, your integrator or designer.