

# Enabling Higher Quality Audio Capture for Smartphones

The rise of user-generated content is placing more and more demands on the capture capabilities of today's smartphones. Yet while imaging and camera technologies have improved significantly over the past few years, audio recording features have been lagging behind. Until Nokia's OZO Audio.

By  
**Jyri Huopaniemi and  
Kai Havukainen**

(Nokia Technologies)

User-generated content is big. Over half of all consumers globally share their videos through mobile. Millennials spend 30% of daily media use viewing user-generated content (5.4 hours)—almost as much as print, radio, and television combined, which comes in at 33%. Yet while imaging and camera technologies have improved significantly over the past few years, audio recording features have been lagging behind until recent advances in spatial audio processing.

## Introducing Immersive Audio Capture

Recent advances in spatial audio processing have seen the audio capabilities of today's smartphones improve significantly. For smartphones with two or more microphones, we can now capture and process spatial sound, enabling faithful reproduction of the spatial sound field and salient audio cues that are present in the recording.

But audio innovation doesn't stop there. The combination of accurate spatial audio analysis with smart multi-microphone audio processing algorithms enables an entirely new set of spatial audio features to be introduced, including advanced audio focusing and audio zoom.

Here, we look at the key factors and technologies behind making modern smartphones into true audio recording devices and showcase the latest application examples.

## Innovating Audio Capture through Algorithms

Mobile and consumer devices come in many different shapes and sizes, from large smartphones to dedicated audio and visual capture devices, such as cameras and video recorders. The technology is incredibly advanced. Today, we practically have home recording studios that can fit in our pockets. But these capabilities have yet to be unlocked or—at best—are largely unusable.

How do we unlock these new audio capture experiences? First, microphones must be placed in such a way that ensures optimal audio capture. There must be at least two microphones present in the device to enable these capabilities—but three or more is better. Put simply, the more microphones, the more use-cases and audio fidelity accuracy you can achieve.

But simply slotting in more microphones by themselves is not enough. Intelligent placement of microphones on devices can overcome the most challenging form factors (e.g., slim, modern smartphones). Embedding high-quality spatial audio does not mean we have to re-invent the wheel.

The design of advanced audio capture systems requires a coordinated implementation, combining the design of product acoustics and spatial audio algorithms. Product design involves many requirements that manufacturers need to consider

and seldom can the number of microphones or their placement be selected freely. Even though a modern smartphone design raises a very challenging form factor for spatial audio capture, the design can be optimized for excellent performance in all conditions, without a need to rely on external microphones.

If optimizing hardware is one side of the coin, then the software is the other. Creation of true-to-life immersive audio recordings can be delivered through OZO Audio smart algorithms, which are carefully tuned to provide the best performance for any hardware design.

This approach means OEMs and ODMs don't need to go back to the drawing board when it comes to the form factor. The algorithm does the heavy lifting and the design can be adapted to the desired form factor and microphone locations through a specialized algorithm tuning process (see **Figure 1**).

### True-To-Life Immersive Spatial Audio

For the Nokia OZO Audio portfolio, the basis of all our algorithms is in the analysis of the multiple microphone signals. These allow us to simultaneously enable high-quality spatial reproduction, as well as being able to provide audio focus and zoom capabilities.

To maximize the performance of the audio capture technology for a device of any shape and size, we can utilize several alternative design methods. The methods include rapid prototyping techniques enabling acoustic measurements or modern virtual prototyping techniques modeling the product acoustics. The key variables addressed in the design are related to the device's dimensions as well as the microphone locations, along with other acoustic details. This workflow enables early evaluation of the target performance and efficient comparison of alternative designs without a need to delay the performance evaluation until mature target product hardware becomes available.

So, what are the audio capture capabilities we can expect? There are three primary capabilities that radically transform user-generated content—spatial audio capture, audio focusing, and audiovisual

zoom. These features and the processing flow are described in Figure 1.

The aim of spatial capture is to faithfully reproduce the full spatial extent of the captured sound scene. Human hearing is omnidirectional—we are able to hear sounds from all directions—and it is able to distinguish sounds coming from front and back, left or right, up and down. To deliver true-to-life experiences, this must be replicated.

Traditionally, spatial recording is achieved by using expensive and dedicated multi-microphone systems and using audio formats that enable modifying the captured audio into desired use case, such as Ambisonics. By utilizing two or more microphones, as mentioned above, we can capture and analyze the content's spatial features and replicate human hearing capabilities.

This is Audio 3D—superior spatial sound capture. Audio 3D captures and delivers a natural sound experience within one degree of accuracy, similar to how the human ear works. With Audio 3D, consumers capture the full richness, depth, direction and detail—without missing a moment, even if it's out of frame.

### Focusing on the Sounds that Matter

But spatial audio capture and playback is just one application. By combining with artificial intelligence machines already present in a device, we can unlock further audio innovations that radically transform the content experience.

Audio focus is one of these—allowing users to focus the audio in a direction that matters when capturing content, similarly as we do for video. An audio focus feature enables users to capture the direction of sound during recording. This gives them the power to focus on only the desired sound from any direction and attenuate all other sounds and background noises.

Imagine you are in a crowd—a very noisy one—in a busy tourist destination. You are recording what you see and are narrating the video you capture to later share on social media. Usually, your voice is

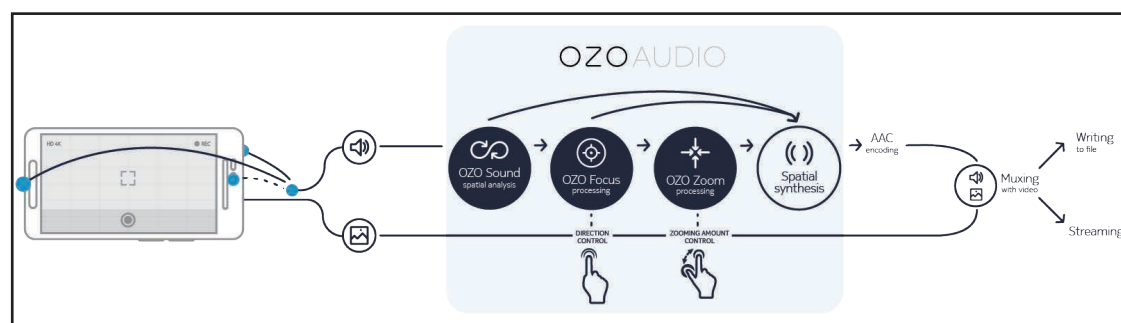


Figure 1. This diagram explains how OZO Audio processes captured audio.

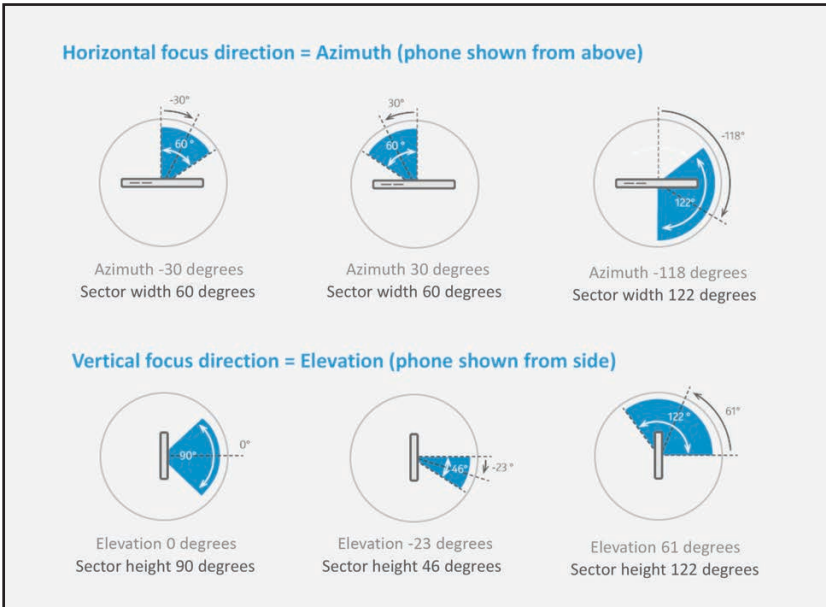


Figure 2. Audio Focus capabilities include audio directional controls. Parameters can be changed dynamically while recording.

competing against all the distracting surrounding noises—it is drowned out. But with audio focus technology in your device, you can focus on your voice and eliminate the background noises you don't want to feature. Your voice comes in crisp and clear, matching the high-definition video you capture.

We can achieve this in consumer devices through advanced multi-microphone audio capture algorithms that analyze and process the audio signals within the device software. For the user, this is delivered through a user interface in native or third-party video recording applications, allowing them to control the audio focus capability, either manually or automatically.

This thinking led us to develop Audio Focus, which minimizes distracting background noise and captures sound from a specific direction, whether the user is shooting a selfie or a street musician.

Audio Focus makes switching from rear-camera video recording to selfie recording simple and

effective. It enables the user to adjust audio to a specific part of the screen and select what matters, and—with audio focus parameters—the direction and sector size, too. By integrating with artificial intelligence algorithms in the device, Audio Focus also enables users to maintain audio focus on moving people or objects and automatically follow a sound source with audio focus parameters controlled by object tracking (see **Figure 2**).

## Realizing Audio-Visual Zoom

The ability to zoom video is now a given for any image capture device or app. We are all used to being able to zoom through pocket cameras, digital single-lens reflex (DSLR) cameras, and through our smartphones. It is an essential capability. But the one thing that this function is missing is being able to match the audio to the zoomed area of the video.

Audio-visual zoom realizes this capability and resolves the disparity between what we see and what we hear. When you zoom the area of video during recording an object of interest, the audio zooms in line with this. It allows you to reduce unnecessary background noises and retain the high fidelity of the visual and audio content in view.

Imagine a busker playing on a busy street. You want to capture this moment, but are a little distance away, so you zoom the video. Before, this would result in you capturing the subject of the video, but still hearing everything else around you.

But with audio-visual zooming, when you zoom the video, you hear what you capture—delivering a superior content experience. It gives the user even greater control, allowing them to hear what they want to hear. This is why we integrated the Audio Zoom functionality into the OZO Audio portfolio. Audio Zoom features intelligent audio zooming that enables the user to dynamically adjust audio to the area of zoomed video. As the picture zooms in on the subject, so can the audio (see **Figure 3**).

## The Importance of Universal Playback and Sharing

There is a rapidly increasing amount of user-generated content being created and consumed on smartphones. This means there is a new requirement for OEMs and ODMs to ensure that all the above spatial audio functionalities can be universally played back and shared across any device or platform. If one user can capture great audio but is not be able to share it with anyone else, then there is no point to the technology.

Universal playback and sharing means that a user can send the video or audio to anyone else and they can also experience true-to-life, immersive

### About the Authors

Jyri Huopaniemi leads product and technology strategy for the Media Technologies organization in Nokia Technologies, the licensing arm of Nokia. Jyri joined Nokia in 1998 and has 25 years of experience in developing and productizing breakthrough audiovisual and software technologies. He has actively contributed to multiple fields of digital media, including audio and imaging technologies, mobile platforms and virtual reality solutions. Jyri received his doctoral degree in 1999 in electrical and communications engineering on the topic of 3-D audio and virtual acoustics, and he has published more than 60 peer-reviewed technical articles in international journals and conferences.

Kai Havukainen is a senior product manager for audio solutions in Nokia Technologies. Kai holds a master's degree in signal processing (2003) and has nearly 20 years of experience at Nokia. He has been working in various roles, including audio engineering, business development, technology licensing, and sales.

experiences—even if they do not have the same spatial audio software running on their device. It means that these experiences can be shared across any social media platform and heard through any device—regardless if it is a laptop, tablet, or smartphone.

When designing OZO Audio, we ensured this was a key design requirement. This is why OZO Audio also utilizes common audio standards (e.g., AAC), so content can be played on a range of popular devices and shared on social media.

### Re-Defining Audio on Consumer Devices


People are telling their stories with technology, and they want devices that allow them to capture the full richness of their lives—both in pictures and sounds. OZO is an evolution in audio, imaging and video technologies. It is born from our proven track record in innovation, using industry-leading technology to enable anyone to seamlessly capture and share the reality of the moment with unprecedented accuracy and ease. OZO enhances the way we see, hear, and feel our human experience.

The OZO Audio portfolio transforms how people



Figure 3. This is a UI demonstration of how Audio Zoom automatically and dynamically adjusts audio focus to the area of the zoomed video.

capture and share their moments through intelligent audio, imaging, and video technologies. Sharing our lives and fostering interaction and community has become more important than ever, and we need the ability to do it in a way that helps us connect more naturally—and in a more authentically human way—with others.

The Nokia Media Technologies group offers licensing of innovative audio and video technologies to manufacturers of consumer devices, providing a key opportunity for device differentiation for its customers. It is time we unlock and let consumers enjoy amazing sound capabilities with their mobile devices, bringing audio up to the same quality we expect from video. 

**HEAD acoustics**

# BEST TESTING

Experience cutting-edge technology that enables true-to-reality and comprehensive tests for your headphones, headsets and hearables.

HEAD acoustics, Inc. • [info@headacoustics.com](mailto:info@headacoustics.com) • [www.head-acoustics.com](http://www.head-acoustics.com)