

BIG SOUND ZOOM

NTT's new intelligent microphone array technology captures sounds at a distance with consummate clarity—and can focus on individual voices and noises as well.

BIFUE USHIJIMA



Kenta Niwa
NTT researcher

WHAT is required to capture clear sound from a specific location at a distance? Certain technologies for this purpose do exist, including parabola microphones used in birdwatching and shotgun mikes used for live sports broadcasts. Tabletop mikes used for long-distance telephone calls, as well as those in smartphones, also capitalize on array technology that uses multiple microphones to process audio signals.

Microphone arrays are well suited to picking up certain sounds clearly, but only around two to four microphones are actually used on arrays that are in practical use at present. Unfortunately, such arrays are poor at collecting distant sounds.

The Acoustic Information Group at NTT Media Intelligence Laboratories wondered what would happen if they increased the number of mikes on an array to one hundred or so. This notion led the lab to develop the “zoom-in mike” sound collection system.

The team established the fundamental principles for the properties of sound the

microphone needed to pick up in order to accurately isolate distant sounds. The result was a special array composed of 12 parabola-shaped reflector plates and 96 specialized microphones. Each mike is calibrated to pick up a certain range of sounds from within the wide range of audio the array collects, allowing it to accurately gather audio even at long distances. Using traditional technology, differentiating sounds picked up at a distance of just five meters was tough, but NTT's zoom-in mike can clearly isolate one person's voice out of a group conversation at a distance of around twenty meters.

Traditional microphone arrays try to pick out desired sounds and remove ambient noise, but their ability to pick up sounds degrades rapidly after five meters, leading to lower sound quality than is typically desired. In contrast, zoom-in mikes collect every sound at high quality and offer enhanced noise suppression and speech recognition. In addition, NTT's microarray processing software, developed in-house, allows specific sound sources to be selected or eliminated at will.



NTT's new "zoom-in" mike array precisely picks up and differentiates sounds at a distance

“By using this system together with a telephoto lens camera, it might even be possible to zoom in the video feed on a certain player during a soccer match in a huge stadium while also amplifying only the sound of that player’s voice,” explains NTT researcher Kenta Niwa.

NTT Media Intelligence Laboratories has also built a “target mike” that uses zoom-in mike technology to capture specified sounds collected during sporting events. While it looks like a regular shotgun mike, the target mike utilizes multiple-microphone array technology to pick up the diverse sounds of sporting events and more. The lab is now running experiments in cooperation with NHK to try amplifying sound that brings broadcasts to life, such as the sound of soccer players kicking a ball or yelling in excitement, the sound of sumo wrestlers unleashing their famous *harite* slaps and their collisions, and other electrifying sounds.

While NTT’s target mike uses the same method as its zoom-in mike to accentuate sounds, the current technology is better at doing this for certain sounds than for others. Overcoming this weakness is the next step for the technology.

“Our technology easily picks up quick bursts of sound, such as when a ball is kicked or a sumo

wrestler is smacked, but has trouble extracting flatter sounds, such as the sound made when a swimmer strokes through the water,” Niwa explains. “We hope to improve on that until it captures all sounds to provide a richer, more realistic experience.”

While one goal of development is clearly to finish in time for use at the 2020 Tokyo Olympics, the scope of other potential uses is also quite broad. The technology could be used during the question-and-answer sessions of international conferences held at large venues—removing the need for individual mikes—as well as at other places where large numbers of people gather. **7**



The “target mike” NTT developed captures specified sounds at sports matches and other events