ViewNoise Acoustic Camera

Technical Overview & User Guide

MA64.UB530.R6 | 64-channel synchronized microphone array

Document revision: August 2026

TELENET LTD.

About this document

This document provides a technical overview of the ViewNoise acoustic camera, its MA64.UB530.R6 microphone array, the ViewNoise acquisition and visualization software, and the signal-processing principles behind acoustic imaging, virtual directional listening, and ICA-based sound-source separation.

The emphasis is on explaining what the system does, how the main processing blocks are related, and which parameters control the trade-off between spatial, temporal and frequency resolution. It is not intended to disclose implementation source code or proprietary optimization details.

Terminology

In this guide, “acoustic image” means the spatial map of sound-related response calculated from the synchronized microphone channels and superimposed on the optical camera image. “Virtual HDR microphone” is a ViewNoise product term for an ICA-based separated virtual audio source; it is not a conventional microphone hardware type.

Contents

1. System overview

ViewNoise is an integrated acoustic-imaging system consisting of the MA64.UB530.R6 64-microphone array, the ViewNoise_K60H_C64 data-acquisition controller, an integrated video camera, environmental sensors, and the ViewNoise software. The hardware and software are designed as a synchronized system for real-time acquisition, recording and offline analysis.

- 64 synchronized MEMS microphone channels, 24-bit at 48 kHz.
- Real-time and offline acoustic imaging with the acoustic map overlaid on the optical video image.
- GPU acceleration through CUDA on NVIDIA GPUs or OpenCL on supported NVIDIA, AMD and Intel GPUs.
- Virtual directional microphone controlled from the acoustic image.
- ICA-based source separation through ViewNoise Virtual HDR microphones.
- Recording of all 64 audio channels together with synchronized video.
- Spectrum, spectrogram and time-domain waveform visualization.
- Directional SPL measurement and real-time source suppression (“Acoustic Eraser”).

Design principle

The raw multichannel recording is preserved. Acoustic images can therefore be recalculated during offline playback using a different image resolution, FFT frame length, overlap or other analysis settings. The recording is not limited to the visualization settings that were selected during acquisition.

2. MA64.UB530.R6 microphone array

The MA64.UB530.R6 is a 520 mm diameter microphone array using 64 MEMS microphones in a non-periodic UB Spiral arrangement. The irregular spatial distribution avoids the strong periodicity of a rectangular or equally spaced array and is suitable for wideband acoustic imaging.

Parameter	Specification
PC interface	USB 2.0 / USB 3.0
Power	USB powered, approximately 2.5 W (5 V / 500 mA)
Array diameter	520 mm
Microphones	64 MEMS microphones, UB Spiral geometry
Audio resolution	24-bit
Sampling rate	48 kHz, synchronized across all channels
Microphone SNR	65 dBA
Sensitivity	-26 dBFS
Sensitivity tolerance	±1 dB
Frequency response	50 Hz to 20 kHz (±3 dB)
Acoustic overload point	117 dB SPL
Weight	Approximately 6 kg without tripod
Acquisition controller	ViewNoise_K60H_C64, integrated in the array
Environmental sensors	Relative humidity, atmospheric pressure and air temperature

2.1 Integrated video camera

The optical camera provides the visual reference on which the acoustic image is overlaid. The following capture formats are supported by the current hardware configuration:

Parameter	Specification
16:9	3840×2160 @ 30 Hz; 1920×1080 @ 30 Hz; 1280×720 @ 30 Hz
4:3	2592×1944 @ 30 Hz; 2048×1536 @ 30 Hz; 1600×1200 @ 30 Hz; 1280×960 @ 30 Hz; 1024×768 @ 30 Hz; 800×600 @ 30 Hz; 640×480 @ 30 Hz

2.2 Array geometry and spatial response

For a microphone array, source direction is inferred from the small differences in arrival time between microphones. A regular array can produce repeated spatial responses at frequencies where the element spacing becomes large relative to the acoustic wavelength. A non-periodic array distributes these ambiguities more irregularly and helps avoid a small number of dominant grating-lobe directions.

ViewNoise processing is not restricted to a regular grid of microphone positions. The steering calculation can use the actual coordinates of the microphones, which allows arbitrary 2D and 3D array geometries to be represented in the processing model.

3. ViewNoise software

ViewNoise combines acquisition, synchronization, visualization, recording, playback and acoustic signal processing in a single Windows application. Virtually all analysis functions are available in real time and during offline playback.

3.1 Real-time and offline processing

During acquisition, the software receives the synchronized 64-channel audio stream and video stream, calculates the selected acoustic visualization and optionally records the raw multichannel data. During offline playback, the acoustic image is recalculated from the recorded microphone channels rather than stored only as a rendered video overlay.

This architecture is important because it allows an engineer to return to the same measurement and change the FFT size, overlap, spatial image resolution, display dynamic range and other analysis settings without repeating the physical measurement.

3.2 Processing acceleration

Acoustic imaging over tens or hundreds of thousands of pixels is computationally intensive. ViewNoise can offload highly parallel operations to a GPU using CUDA or OpenCL while using the CPU for control, multithreaded source separation and other tasks. The exact achievable real-time resolution depends on the selected processing parameters and the available CPU/GPU performance.

4. Acoustic imaging and ViewNoise beamforming

The ViewNoise acoustic image is based on the propagation-time differences measured across the microphone array. The processing is correlation-based: synchronized microphone frames are transformed to the frequency domain, pair-wise correlation information is formed, and the response expected for each candidate image point is evaluated from the microphone geometry.

With 64 microphones there are $64 \times 63/2 = 2016$ unique microphone pairs. Every pair contributes independent spatial timing information. Combining all pairs provides much more spatial evidence than steering from a single reference microphone.

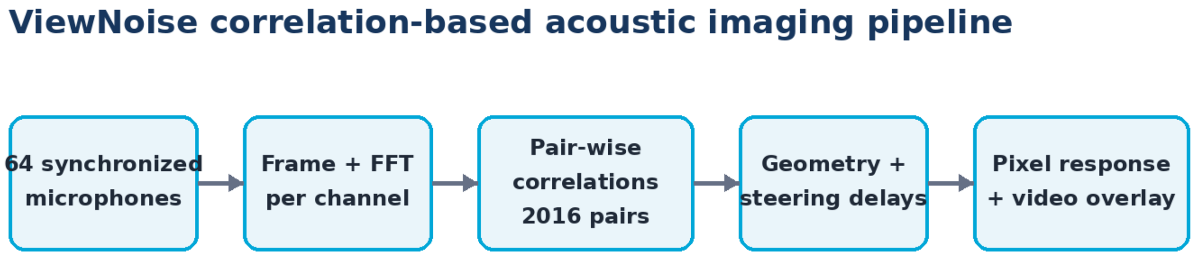


Figure 1. Conceptual ViewNoise correlation-based acoustic-imaging pipeline.

4.1 Propagation delay for a candidate point

Let r_i and r_j be the positions of microphones i and j, and let p be a candidate point in the acoustic image. The propagation distance from p to microphone i is $d_i(p)$. For sound speed c, the expected time-difference of arrival (TDOA) for the pair is:

$$\tau_{ij}(p) = [\, d_i(p) - d_j(p) \,] / c$$

For a distant source, the spherical-wave geometry approaches the familiar far-field plane-wave delay determined by the source direction and the microphone baseline. For finite-distance focusing, using the actual distances retains the near-field geometry.

4.2 Pair-wise correlation

For each microphone frame $x_i[n]$, an FFT produces the complex spectrum $X_i[k]$. A pair-wise cross-spectrum can be formed as:

$$G_{ij}[k] = X_i[k] \cdot \mathrm{conj}(\, X_j[k] \,)$$

The inverse transform of the cross-spectrum provides a correlation function versus delay. Generalized cross-correlation methods can additionally apply a frequency weighting before the inverse transform. PHAT is one well-known phase-only weighting; it normalizes the cross-spectrum magnitude and emphasizes timing consistency. ViewNoise acoustic imaging is not conceptually dependent on PHAT weighting: the essential operation is the use of synchronized pair-wise correlation information together with the geometrically predicted delay.

4.3 Forming one acoustic-image pixel

For each candidate point p, ViewNoise obtains the expected delay $\tau_{ij}(p)$ for every microphone pair and samples the corresponding pair-correlation response. The contributions of the available pairs are then accumulated to produce the spatial response for that image point. In conceptual form:

$$B(p) = \sum_{i<j} w_{ij} \cdot R_{ij}(\, \tau_{ij}(p) \,)$$

Here R_{ij} is the pair-correlation function and w_{ij} represents any pair weighting or normalization used by the implementation. A real sound source produces mutually consistent delays across many microphone pairs, so the accumulated response rises at the corresponding image location. Inconsistent noise and unrelated reflections tend to add less coherently.

Why 2016 pairs matter

The method uses spatial information from every unique pair of the 64 microphones. This large set of baselines is one reason the acoustic image can remain useful over a broad frequency range and with an irregular array geometry. The cost is computational load, which is why GPU acceleration is important.

4.4 FFT frame length, frequency resolution and overlap

The acoustic image is calculated from finite time frames. ViewNoise supports FFT frame lengths from 256 to 65,536 samples. At a 48 kHz sampling rate, the nominal frequency-bin spacing is $\Delta f = Fs/N$. A longer frame gives finer frequency resolution but spans a longer time interval; a shorter frame responds more quickly to transient changes but has coarser frequency resolution.

Parameter	Specification
256 samples	5.33 ms frame; 187.5 Hz bin spacing
1,024 samples	21.33 ms frame; 46.875 Hz bin spacing
4,096 samples	85.33 ms frame; 11.719 Hz bin spacing
16,384 samples	341.33 ms frame; 2.930 Hz bin spacing
65,536 samples	1.365 s frame; 0.732 Hz bin spacing

ViewNoise also supports strongly overlapped processing frames. The overlap factor can be selected from 2 to 128, corresponding to a hop of N/2 down to N/128 samples (50% to approximately 99.22% overlap). High overlap increases the number of acoustic images calculated per second without changing the underlying frequency resolution of the FFT frame.

Acoustic slow motion

With N = 256 samples and an overlap factor of 128, the hop is only 2 audio samples. At 48 kHz this corresponds to 24,000 acoustic-analysis positions per second. During offline playback this can be used to inspect rapid acoustic events in “slow motion”. The optical video remains limited by its own frame rate; the acoustic analysis can have much finer temporal sampling.

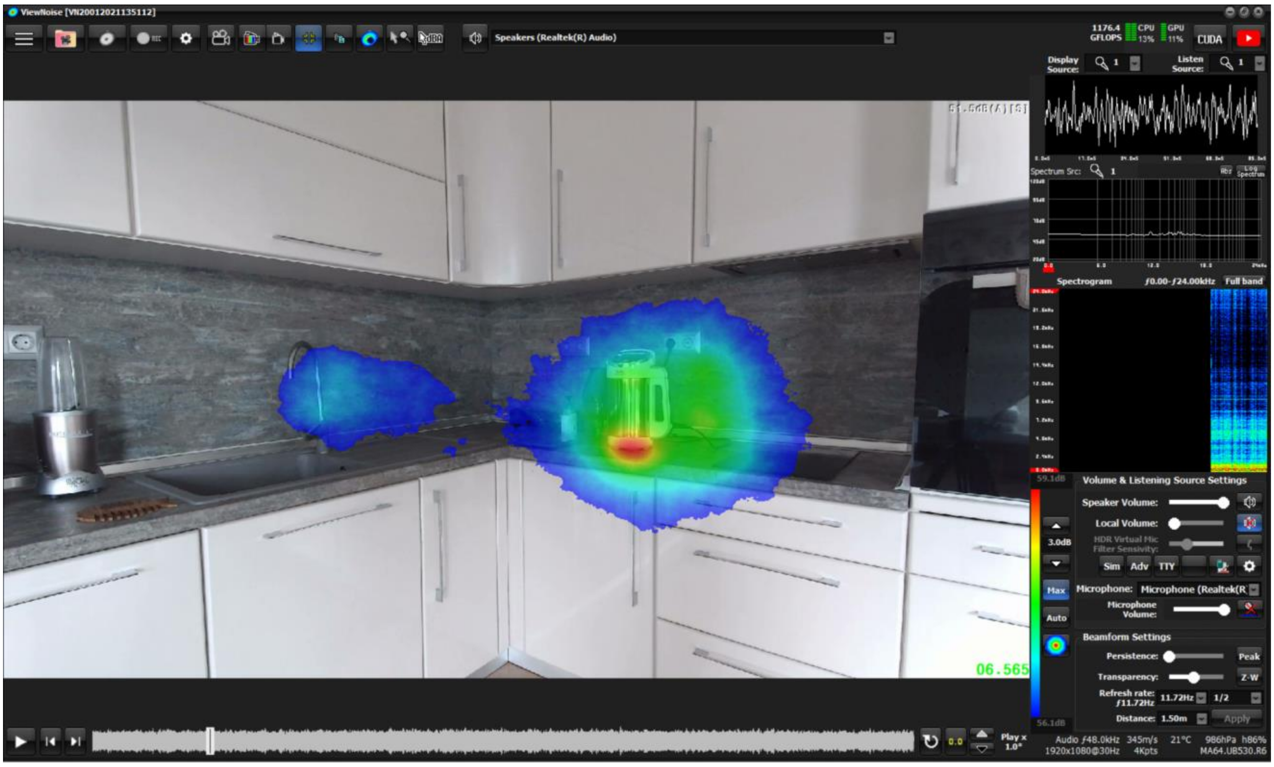
4.5 Acoustic-image resolution

The spatial display grid can be selected from 80×60 up to 640×480 pixels. A 320×240 grid contains 76,800 candidate image points and is a practical default for many measurements. Increasing the grid resolution increases the number of steering evaluations and therefore the processing load, but does not change the physical aperture of the microphone array.

Interpretation

Pixel count is a display/search-grid parameter; it should not be confused with the fundamental angular resolution of the array. Physical resolution is governed primarily by wavelength, array aperture, source distance, array geometry and the signal-processing method.

The screenshot below shows a typical ViewNoise operating screen during localization of a kitchen blender. The acoustic image is superimposed on the optical video, while the right-side panels display the time-domain signal, spectrum, spectrogram and listening controls. In this example the dominant acoustic hotspot is correctly localized near the blender jar/base assembly, while weaker activity is visible in the surrounding kitchen environment.



VN Figure 2. Example of the ViewNoise user interface with acoustic-image overlay during localization of a kitchen blender.

5. Virtual directional microphone

The virtual directional microphone converts a selected position in the acoustic image into an audio signal. The operator points to a location with the mouse, and ViewNoise applies the delays associated with that direction or point so that signals arriving consistently from the selected location add coherently.

The result can be monitored in real time or reproduced during offline playback. This is useful when the acoustic image identifies several possible sources and the operator wants to listen to one spatial region without placing a physical microphone there.

Real-time monitoring

Headphones are recommended when listening to a virtual microphone in real time. Loudspeaker playback in the same room can feed sound back into the microphone array and alter the measurement.

6. ICA-based sound-source separation — Virtual HDR microphones

A conventional steered or delay-and-sum virtual microphone improves sensitivity toward a selected direction, but it cannot always isolate a quiet source located close to a much louder source or separate sources whose spatial responses overlap. ViewNoise addresses this case with Virtual HDR microphones based on Independent Component Analysis (ICA) / Blind Source Separation (BSS).

Conceptual ViewNoise Virtual HDR / ICA source-separation chain

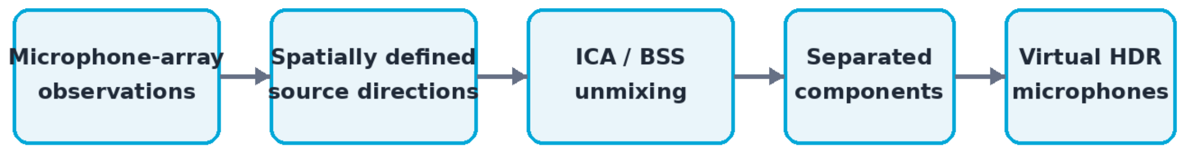


Figure 3. Conceptual relationship between spatial source selection and ICA-based Virtual HDR microphones.

6.1 ICA model

ICA starts from a multichannel observation model in which the measured signals are mixtures of several underlying sources. In the simplest linear instantaneous form:

$$x(t) = A \cdot s(t) + n(t)$$

x(t) is the vector of microphone observations, s(t) contains the unknown source signals, A is an unknown mixing matrix and n(t) represents noise and modelling error. ICA estimates an unmixing matrix W so that:

$$y(t) = W \cdot x(t)$$

The components y(t) are chosen to be as statistically independent as possible. Practical ICA methods exploit the fact that independent real-world sources are typically non-Gaussian and that mixing tends to make observed combinations more Gaussian. Pre-processing commonly includes centering and whitening so that the numerical optimization can focus on statistical independence rather than simple covariance.

6.2 Assumptions and real acoustic environments

ICA is fundamentally a statistical separation method, not a directional filter. Its classical identifiability assumptions include statistical independence of the sources and non-Gaussianity of all but at most one component. Real rooms are more difficult because propagation includes delays, reflections and reverberation, creating a convolutive rather than perfectly instantaneous mixture.

ViewNoise combines the spatial information available from the microphone array with ICA-based separation. The operator places Virtual HDR microphones at the positions of the sound/noise sources of interest, and the system separates the corresponding signals in real time or offline. This spatially guided workflow is especially useful when source direction is already visible in the acoustic image but simple steering does not provide sufficient isolation.

6.3 Why “HDR”

“HDR” is a ViewNoise product term describing the ability to recover useful audio from sources with a large level difference — for example, a weak mechanical tone located close to a much louder source. The analogy is to high dynamic range imaging: the objective is to retain information from sources that would otherwise be masked in a conventional directional output.

6.4 Number of virtual sources and processing load

With the 64-channel array, ViewNoise supports up to 63 Virtual HDR microphones. The practical number that can be processed in real time depends strongly on CPU core count, clock rate, SIMD support (AVX/AVX2/AVX-

512 where available), selected processing parameters and the complexity of the acoustic scene. Offline analysis is not constrained by real-time throughput in the same way.

ICA limitations

Source separation should not be interpreted as perfect physical isolation. Separation quality depends on source independence, source positions, reverberation, signal bandwidth, relative levels and the amount of usable spatial/statistical information in the recording.

7. Acoustic Eraser

Acoustic Eraser is the complementary operation to directional listening: instead of emphasizing a selected spatial source, the operator selects one or more directions to be suppressed. The function can be used in real time, with up to 64 operator-selected sources according to the current system implementation.

This is useful when a dominant source masks weaker acoustic events. Suppression should be interpreted as signal-processing attenuation rather than removal of the physical sound field; residual energy can remain because of reflections, spatial overlap and modelling error.

8. Directional SPL measurement

ViewNoise can calculate sound-pressure-level information associated with a selected spatial direction, providing a measurement concept similar to a directional microphone but using the array processing. The displayed SPL result should be interpreted together with the selected processing bandwidth, calibration and acoustic environment.

9. Recording and video export

9.1 Multichannel recording

A ViewNoise recording contains the synchronized optical video together with all 64 audio channels at 24-bit / 48 kHz. Because the multichannel source data are retained, offline analysis can reproduce a real microphone channel, a virtual directional microphone or a Virtual HDR microphone and can recalculate the acoustic image using different settings.

9.2 Video codecs

ViewNoise supports H.264/AVC and H.265/HEVC recording/export workflows with available hardware acceleration on supported Intel, AMD or NVIDIA platforms, or software encoding on the CPU. H.265 generally provides better compression efficiency, while H.264 remains the recommended choice when broad playback compatibility is more important.

Exported acoustic videos can include text subtitles at selected times/positions and audio comments recorded from a microphone.

10. Spectrum, spectrogram and time-domain views

10.1 Spectrogram

The spectrogram shows how spectral energy changes with time. Its time/frequency trade-off follows the selected FFT frame length and overlap: longer frames improve frequency discrimination, while shorter frames improve temporal localization. Increasing overlap makes the displayed evolution smoother and allows finer temporal stepping between adjacent analyses.

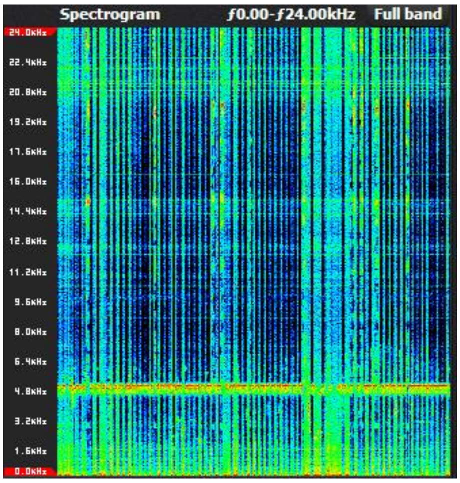


Figure 4. ViewNoise spectrogram display.

10.2 Spectrum

The spectrum view displays the frequency-domain magnitude of a selected real microphone and, when applicable, the currently reproduced virtual directional or Virtual HDR source. Linear/logarithmic and absolute/relative display modes are available.

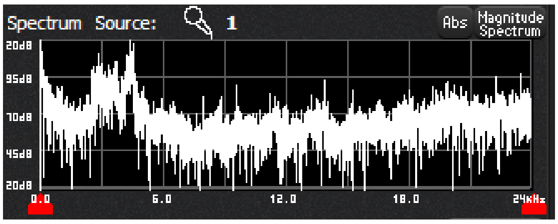


Figure 5. ViewNoise spectrum display.

10.3 Audio waveform

The time-domain view displays the waveform of the selected microphone channel or multichannel selection. It is useful for examining transients, clipping, periodicity and the relationship between audible events and the acoustic/video timeline.

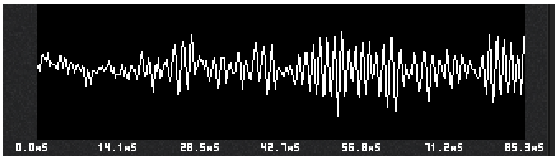


Figure 6. ViewNoise time-domain waveform display.

11. Offline playback and sound reproduction

ViewNoise reproduces recorded audio and video in synchronization. The video display can be switched between color, black-and-white and contour presentation. GPU-assisted video decoding can be used where supported, particularly for H.265/HEVC recordings.

Audio can be reproduced from:

- any one of the 64 physical microphones;
- the virtual directional microphone;
- a Virtual HDR microphone.

Playback volume can be controlled locally and globally, including mute. When playback is slowed or accelerated, audio reproduction follows the selected playback rate. Acoustic-image processing can use much finer temporal steps than the 30 Hz optical video, enabling close inspection of rapid sound events during offline analysis.

12. Computer requirements

The following requirements are based on the current ViewNoise manual and are intended as a practical minimum for the 64-channel system:

Parameter	Specification
Operating system	Windows 10, 64-bit
CPU	4 or more CPU cores; higher core count is recommended for multiple real-time Virtual HDR microphones
GPU	CUDA-capable NVIDIA GPU or OpenCL-capable NVIDIA / AMD / Intel GPU; approximately 1 TFLOPS single-precision performance or higher recommended for GPU acoustic imaging
Memory	16 GB RAM minimum
Storage	SSD recommended; multichannel recordings are data-intensive and approximately 60 seconds can require up to about 1 GB depending on video settings

Performance scaling

There is no single “maximum” real-time setting independent of the computer. Acoustic-image pixel count, FFT size, overlap, number of Virtual HDR microphones, video codec and other simultaneous functions all compete for CPU/GPU/memory bandwidth.

13. Practical interpretation of acoustic images

- Use frequency-selective analysis when a known tonal component is being localized. Broadband maps can combine several physical sources and reflections.
- Do not interpret a small bright spot as proof that the physical source is exactly that size. The spot width is the array point-spread response, not an optical measurement of source dimensions.
- Strong reflections can appear as secondary acoustic sources. Moving the array or changing the frequency band can help distinguish a real source from a reflection.
- A finer pixel grid makes the displayed image smoother but cannot create physical spatial resolution beyond the microphone-array aperture and wavelength limits.
- For transient analysis, reduce the FFT frame length and/or use high overlap. For closely spaced spectral tones, increase the FFT frame length.
- Use Virtual HDR microphones when source level differences or spatial overlap make the ordinary virtual directional microphone insufficient.

14. Technical background references

The ViewNoise processing implementation is proprietary. The following publications are included as background references for the established signal-processing concepts described in this guide:

Hyvärinen, A.; Oja, E. “Independent Component Analysis: Algorithms and Applications.” Neural Networks, 13(4–5), 411–430, 2000.

Knapp, C. H.; Carter, G. C. “The Generalized Correlation Method for Estimation of Time Delay.” IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4), 320–327, 1976.

DiBiase, J. H.; Silverman, H. F.; Brandstein, M. S. “Robust Localization in Reverberant Rooms.” In: Microphone Arrays: Signal Processing Techniques and Applications, Springer, 2001.

15. Notes and nomenclature

Parameter	Specification
BSS	Blind Source Separation
FFT	Fast Fourier Transform
GPU	Graphics Processing Unit
ICA	Independent Component Analysis
PHAT	Phase Transform weighting used in generalized cross-correlation methods
SPL	Sound Pressure Level
TDOA	Time Difference of Arrival
Virtual HDR microphone	ViewNoise term for an ICA-based separated virtual source