Principles of Echolocation
in Bats and Machines

Note 7

From Call to Image

The processing pipeline of an eRTIS sonar, step by step

Jan Steckel
Cosys-Lab, University of Antwerp, and Flanders Make, Belgium jan.steckel@uantwerpen.be

Version 2026-11-18 16:28

1 What an eRTIS Measures

An eRTIS sensor sends out a short broadband call, listens with 32 microphones, and turns what it hears into an image of echo energy over range and direction, the energyscape. Between the call and the image lie a handful of processing steps, each with a clear purpose (Fig. 1). In this note, we follow one measurement through all of them. We simulate the sensor in two dimensions (the horizontal plane), with the real layout of the 32 microphones and four point reflectors, so that every step can be shown on the same data.

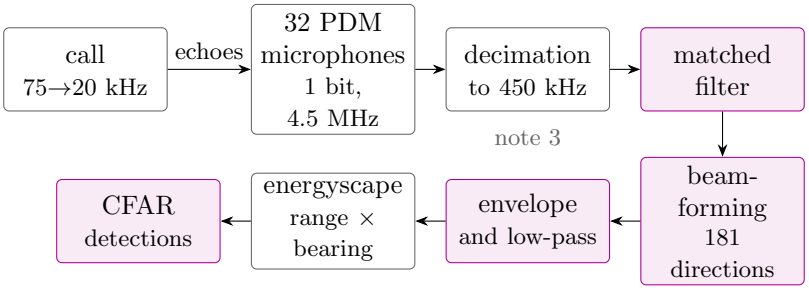


Figure 1: The eRTIS pipeline. This note covers the steps in magenta. The decimation of the 1-bit microphone streams is the subject of note 3.

All positions and directions in this note are given in the frame of the sensor (Fig. 2a). The x-axis points forward, the y-axis to the left and the z-axis up. The 32 microphones lie in the YZ-plane, the front of the sensor, at positions $(0, p_y(i), p_z(i))$ for microphone i (Fig. 2b). A direction ψ is given by its azimuth θ, positive to the left, and its elevation φ, positive up, and points along the unit vector

$$\boldsymbol{u}(\psi) = \begin{bmatrix} \cos\varphi\cos\theta & \cos\varphi\sin\theta & \sin\varphi \end{bmatrix}^T . \tag{1}$$

Our simulation stays in the horizontal plane ($\varphi = 0$), where only the horizontal coordinates $p_y(i)$ of the microphones matter.

The amounts of data are worth keeping in mind. The 32 microphones together deliver $32 \times 4.5 = 144$ Mbit/s of raw bits. After decimation, that becomes $32 \times 450{,}000 = 14.4$ million samples per second. One measurement of the eRTIS software holds 163,840 bits per microphone, about 36 ms, enough for echoes from about 6 m away.

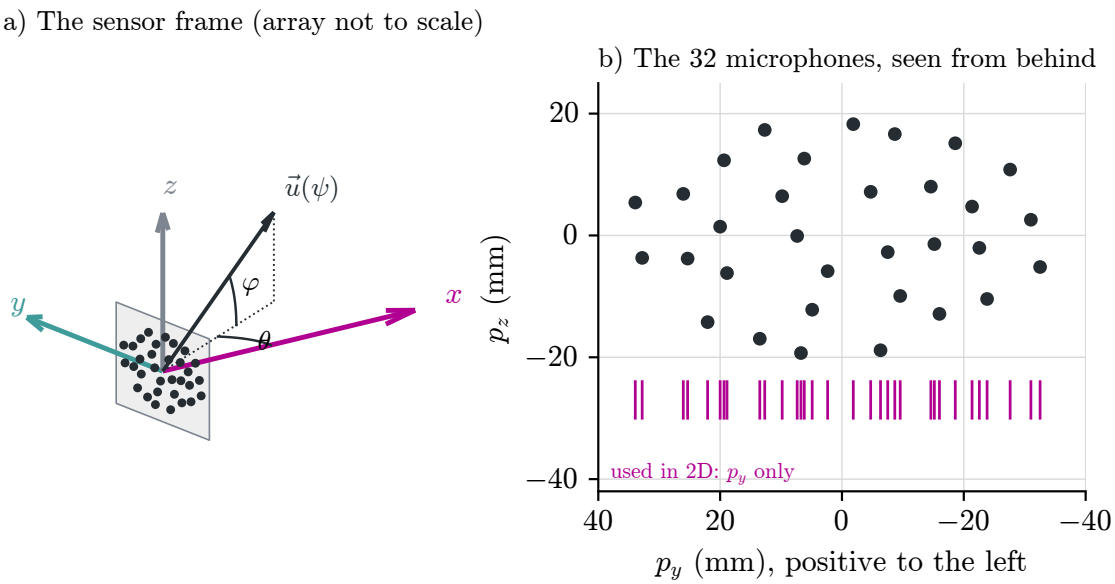


Figure 2: a) The sensor frame: x forward (magenta), y to the left (teal), z up (grey), the array in the YZ-plane (drawn enlarged), and a direction ψ with azimuth θ and elevation φ. b) The real layout of the 32 eRTIS microphones, seen from behind the sensor so that y points to the left (an aperture of about 6.6×3.8 cm). In the horizontal plane, only their coordinate p_y matters (magenta ticks).

2 The Call and the Echoes

The eRTIS emits a hyperbolic frequency sweep of 2.5 ms, from about 75 down to 20 kHz (Fig. 3). In a hyperbolic sweep, the period of the signal, rather than its frequency, changes linearly with time. Such a sweep keeps its shape when it is compressed or stretched by the motion of the sensor, which makes it tolerant to Doppler shifts. (The eRTIS software lets the user choose the sweep, its default is a sweep from 25 to 50 kHz.)

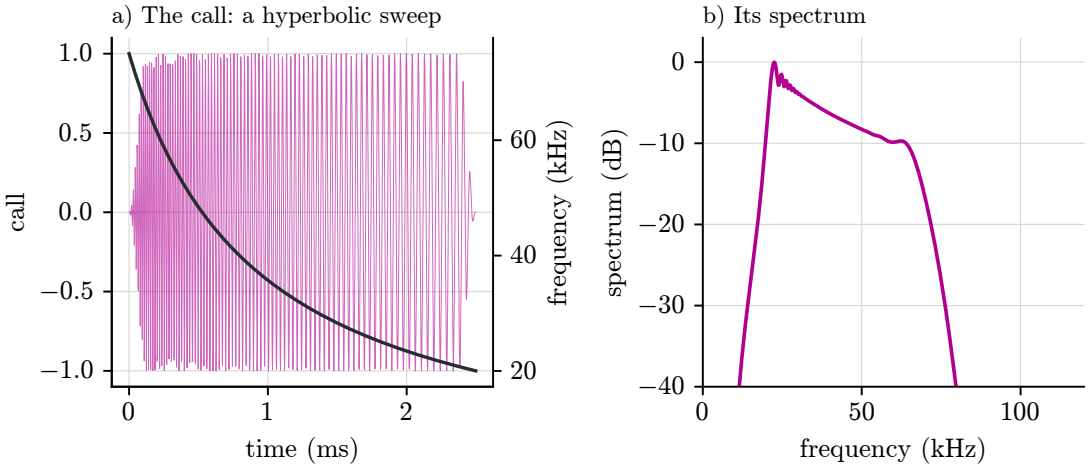


Figure 3: The call. a) Waveform (magenta) and instantaneous frequency (black). b) Spectrum.

Every reflector returns a delayed and attenuated copy of the call to every microphone. In Fig. 4, four points (A to D) lie in front of the sensor, and two of them (B and C) at the same range of 2 m, 10° apart. The microphone signals (Fig. 5a) show the problem. Every echo is as long as the call, 2.5 ms or 43 cm of range, the echoes of B and C overlap completely, and nothing in a single microphone signal tells us where the echoes come from. The two steps that

solve this are the matched filter, for range, and the beamformer, for direction.

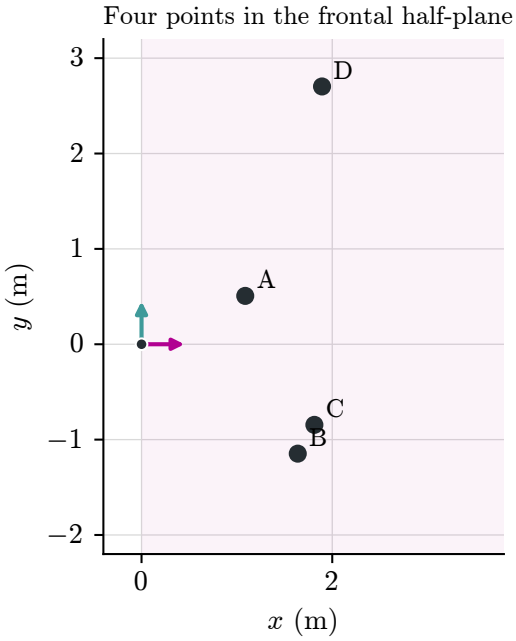


Figure 4: The simulated scene in the horizontal plane, seen from above, with the sensor drawn as a triad (position, pointing direction x in magenta, left y in teal).

3 The Matched Filter: Range

The matched filter correlates every microphone signal with the call. Wherever an echo matches the call, all its frequency components add up in phase, and the 2.5 ms echo collapses into a short pulse at the time of arrival (Fig. 5b). Two things are gained. The first is range resolution: the width of the compressed pulse is set by the bandwidth B of the call, not by its length, and two reflectors can be separated if their ranges differ by about

$$\Delta r = \frac{c}{2B} = \frac{343 \text{ m/s}}{2 \times 55 \text{ kHz}} \approx 3.1 \text{ mm}. \tag{2}$$

The second is sensitivity: the filter collects the energy of the whole call, which raises the echo above the noise by the time-bandwidth product, $TB = 2.5 \text{ ms} \times 55 \text{ kHz} \approx 140$, or about 21 dB. In practice, the matched filter is computed in the frequency domain, as a multiplication of the spectra. If we keep only the positive frequencies, the output is the analytic signal, whose magnitude is the envelope we will need later.

4 Beamforming: Direction

An echo from a reflector a few meters away reaches the small array as a plane wave from direction ψ, travelling along $-\boldsymbol{u}(\psi)$. A microphone at position $\boldsymbol{p}_i = (0, p_y(i), p_z(i))$ lies closer to the reflector than the array center by the projection of its position onto the direction of the echo, $\boldsymbol{u}(\psi)^T \boldsymbol{p}_i = \psi_y(i) + \psi_z(i)$, with

$$\psi_y(i) = \sin\theta \cos\varphi \; p_y(i), \qquad \psi_z(i) = \sin\varphi \; p_z(i). \tag{3}$$

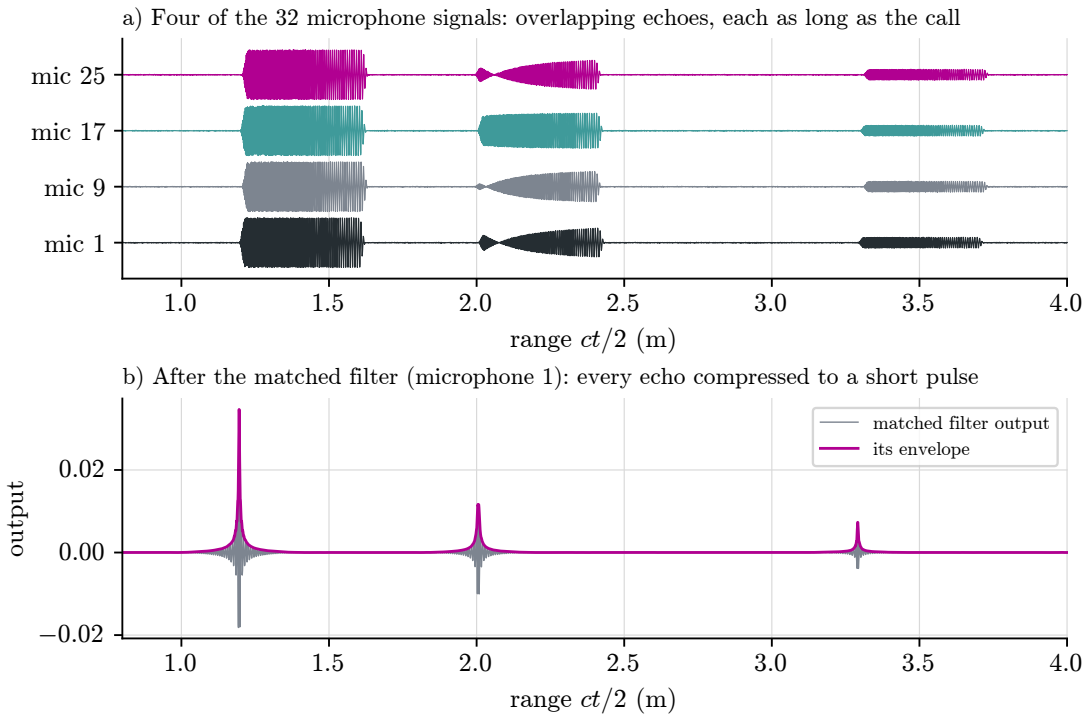


Figure 5: a) Four of the 32 microphone signals (normalized, offset for clarity). b) The output of the matched filter for one microphone, and its envelope.

The echo needs less time to reach microphone i, so that microphone hears it earlier than the center, by the time delay

$$\tau_i(\psi) = \frac{\psi_y(i) + \psi_z(i)}{c}, \qquad \text{or } f_s\,\tau_i(\psi) \text{ samples at the sample rate } f_s. \tag{4}$$

For our array, these delays are at most 99 µs, or 45 samples at 450 kHz.

For a single frequency f, with wavelength $\lambda = c/f$, a time shift is the same as a phase shift. Delaying a signal by τ multiplies its spectrum by $e^{-j2\pi f\tau}$. The phase delays of the 32 microphones for direction ψ together form the steering vector of narrowband array processing,

$$v_i(\psi) = e^{j\frac{2\pi}{\lambda}(\psi_y(i)+\psi_z(i))} = e^{j2\pi f\tau_i(\psi)}. \tag{5}$$

The call, however, covers 20 to 75 kHz, and a phase shift that aligns the microphones at 75 kHz is wrong at 20 kHz. A broadband sonar therefore needs the time delays themselves, or, equivalently, a separate phase for every frequency. Delay-and-sum beamforming undoes the delays for one chosen direction and averages the microphones,

$$b_\psi(t) = \frac{1}{M}\sum_{i=1}^{M} z_i\big(t - \tau_i(\psi)\big), \qquad B_\psi(f) = \frac{1}{M}\sum_{i=1}^{M} e^{-j2\pi f\tau_i(\psi)}\,Z_i(f), \tag{6}$$

in time or in frequency, with z_i the matched-filter output of microphone i, $Z_i(f)$ its spectrum and $M = 32$. The eRTIS software rounds $f_s\tau_i(\psi)$ to whole samples at 450 kHz and stores them as a matrix, one delay per microphone and direction. Our simulation applies the exact delays, as phases per frequency bin of the FFT. In the horizontal plane, $\varphi = 0$ makes $\psi_z(i) = 0$, and the delay reduces to $\tau_i = p_y(i)\sin\theta/c$.

For the direction of an echo, all 32 copies line up and add, while for every other direction they arrive at different times and largely cancel (Fig. 6a). Repeating this for many directions turns the 32 microphone signals into a fan of 181 beams from −90° to +90°, the default of

the eRTIS software. Beamforming gives direction, and it adds sensitivity as well: averaging 32 microphones lowers uncorrelated noise by $10 \log_{10} 32 \approx 15$ dB. How sharp the beams are depends on the size of the array: with an aperture of 6.6 cm, the beam towards A is about 10° wide (where it falls by 3 dB), which is why B and C, 10° apart, are only just separated.

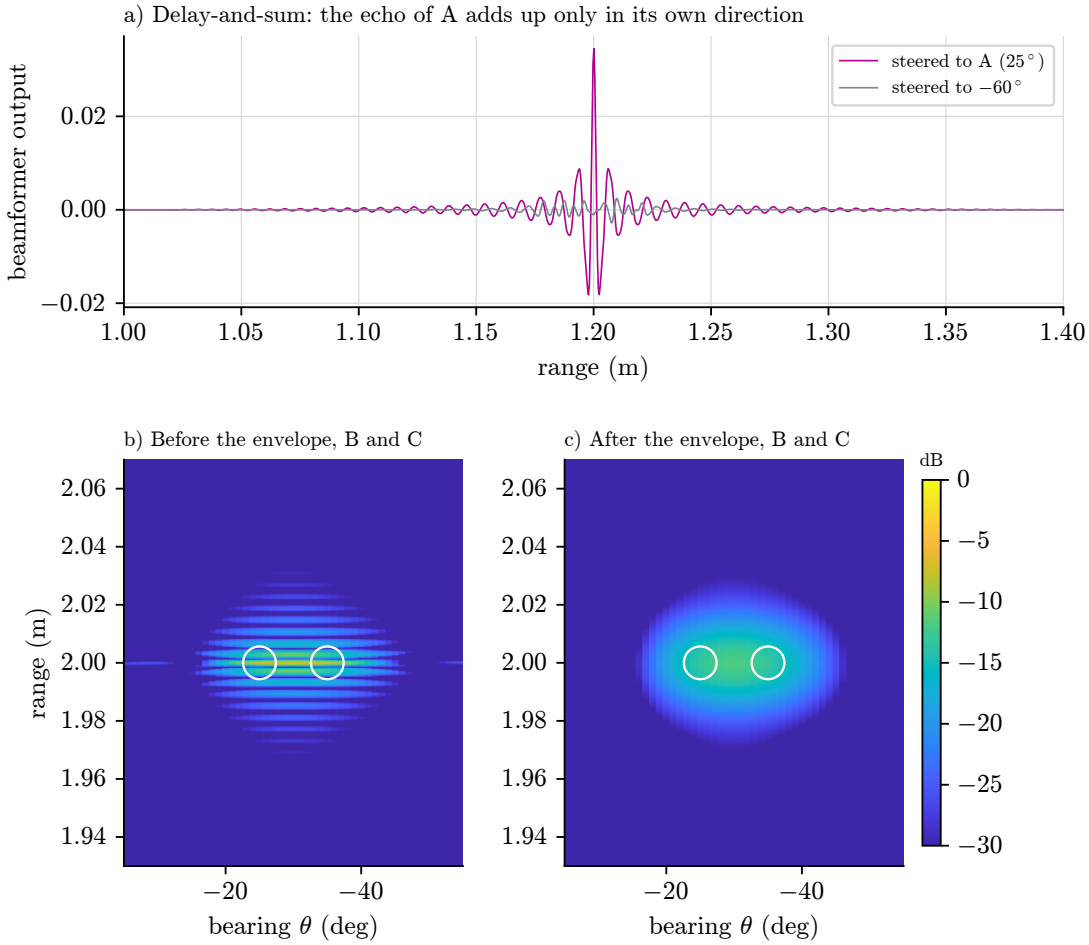


Figure 6: a) The beamformer output around the range of A, steered towards A and towards −60°. b, c) The beamformed image around B and C, before and after the envelope (dB). The circles mark the true positions.

5 The Envelope and the Energyscape

After beamforming, the signal in every direction still oscillates at the frequencies of the call (Fig. 6b). For an image, we only want to know how much energy arrives from each range and direction, so we take the magnitude of the analytic signal, the envelope, and smooth it with a low-pass filter (Fig. 6c). The low-pass sets a trade-off: a lower cutoff gives a smoother image that can be stored with fewer samples, but it also blurs the range.

On real eRTIS data, we use a soft low-pass at 5 kHz, a second-order Butterworth filter that is applied once forward and once backward in time. The reason is its phase (Fig. 7). An ordinary filter is causal, it only uses samples from the past, and every causal filter has a phase that falls with frequency (grey dashed line in Fig. 7a). A phase that falls with frequency delays the signal, by the slope of the phase, the group delay. For the 5 kHz Butterworth, the group delay at low frequencies is $\sqrt{2}/(2\pi \times 5\ \text{kHz}) \approx 45$ µs, or 7.7 mm of range. A causal low-pass therefore moves every echo about 7 mm further away and gives it a tail on the far side (grey line in Fig. 7b).

We do not need a causal filter, because the whole measurement is in memory. We filter the envelope once, reverse it in time, filter it again and reverse it back. Reversing a signal in time conjugates its spectrum, so the second pass has exactly the opposite phase of the first, and the two phases cancel. The combined response is

$$H(f)\,H^*(f) = |H(f)|^2, \tag{7}$$

which is real and positive at every frequency. Its phase is zero (magenta dashed line in Fig. 7a), so it delays nothing, and its impulse response is symmetric around zero. The filtered envelope peaks exactly at the range of A (magenta line in Fig. 7b). The price is that the gain is applied twice, so the filter is 6 dB down at 5 kHz instead of 3 dB. The softness of the Butterworth matters as well. A brick-wall low-pass, which passes everything below 5 kHz and nothing above, also has zero phase, but its impulse response rings. Every strong echo then gets a train of weak copies before and after it, the largest one −16 dB below the echo (teal line), and these show up in the image as regular range rings. With the forward-backward Butterworth, a point is 23 mm wide in range (where its envelope falls to half), against the 3.1 mm the bandwidth allows.

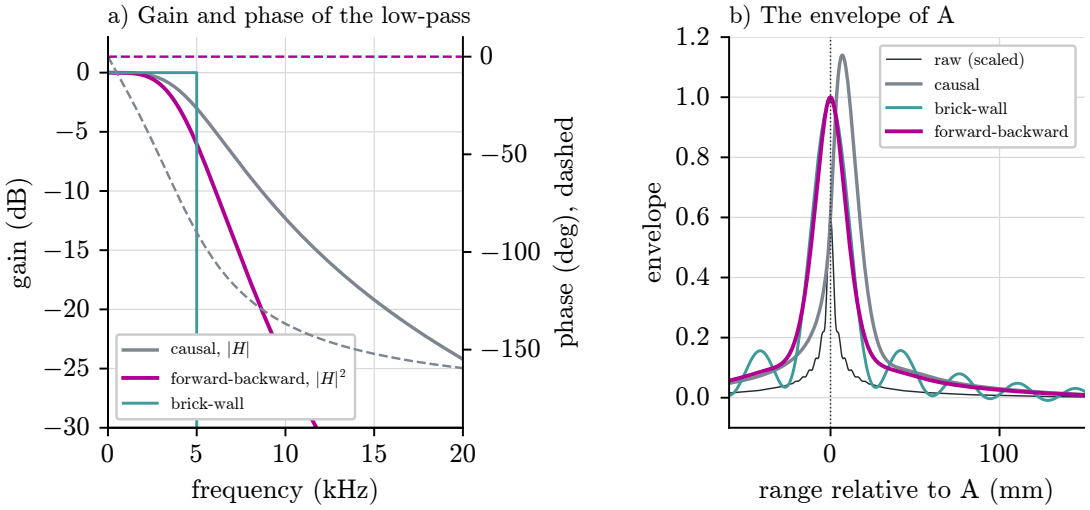


Figure 7: Why the envelope low-pass runs forward and backward. a) Gain (solid) and phase (dashed) of the causal second-order Butterworth at 5 kHz (grey), of the same filter applied forward and backward (magenta, zero phase), and of a brick-wall low-pass (teal). b) The envelope of A in the beam towards A, before the low-pass and after each filter. The causal filter shifts the peak, the brick-wall filter rings.

The eRTIS software itself uses a 1 kHz low-pass by default and keeps every tenth sample, which gives an even smoother image with a resolution of several centimeters. Finally, the energy is averaged in range cells of 2 cm.

The result is the energyscape (Fig. 8a): echo energy over range (here from 0.5 to 4 m, the eRTIS software uses 0.5 to 5 m by default) and bearing. Every reflector appears as a spot at its range and bearing, with the shape of the point-spread function of the sensor (see note 2). The faint horizontal streaks at the range of every echo are the sidelobes of the array: a strong echo leaks a little into all directions.

6 Detection: CFAR

The energyscape still contains everything: the reflectors, the sidelobes of the array, noise and, in real environments, diffuse clutter. To decide where the reflectors are, the image is compared with a threshold that adapts to its surroundings, a constant false-alarm rate (CFAR) detector. The one we use in the lab is the range-wise CFAR (RW-CFAR). It takes four steps, which Fig. 8a–d

shows one by one on the energyscape $P(r, \psi)$ of our scene. In the first step, it computes for every range cell r a reference level, the mean power over all D directions ψ,

$$T(r) = \frac{1}{D} \sum_{\psi} P(r, \psi). \tag{8}$$

This is one number per range, the grey line in Fig. 8b. It peaks at the ranges of the reflectors, because their main lobes and sidelobes add to the mean. In the second step, the reference is smoothed along range with a Gaussian of 3 cells (magenta line), so that it does not jump from one range cell to the next and also covers the flanks of every echo in range. In the third step, the reference is multiplied by a gain $\alpha \approx 1.5$ and every cell of the energyscape is compared with it. Figure 8c shows the ratio $P(r, \psi)/\alpha T(r)$, and only the cells above 0 dB, outlined in white, will survive. In the fourth step, the threshold is subtracted and everything below it is set to zero,

$$P^{+}(r, \psi) = \max\big(P(r, \psi) - \alpha\, T(r),\ 0\big), \tag{9}$$

which leaves the image of Fig. 8d.

It works because of the way the sidelobes behave. A strong reflector spreads its sidelobes over all directions at its own range, and raises exactly the mean $T(r)$ of that range. The sidelobes therefore stay below $\alpha T(r)$, while the main lobe of the reflector rises far above it (Fig. 8e). In our scene, RW-CFAR keeps all four reflectors, and the strongest structure more than 20° away from A at its range drops from −14 dB to −20 dB. Diffuse clutter, which also covers many directions at once, is suppressed in the same way. In range cells that hold only noise, the mean is set by the noise itself, and 10% of those cells also rise above $\alpha T(r)$. They are at most −68 dB below the strongest echo and invisible in Fig. 8d. In practice they are removed by keeping only a fixed dynamic range below the strongest CFAR output (25 dB in the lab's RTISSLAM 3D).

RW-CFAR has a known weakness, inherited from all cell-averaging detectors: the mean is pulled up by strong reflectors. When several strong reflectors share a range with a weak one, the threshold can rise above the weak one and hide it (target masking). An ordered-statistic CFAR (OS-CFAR, after Rohling) avoids this by replacing the mean with an order statistic, here the median over the directions. A few strong cells hardly move the median. For the same false-alarm rate in pure noise (whose power is exponentially distributed, with a median of $\ln 2$ times the mean), its gain becomes $\alpha_{\mathrm{OS}} = \alpha / \ln 2 = 2.16$. Figure 9 puts two strong reflectors and one 12 dB weaker at the same range of 2.5 m. There, the mean lies at −8 dB and the median at −13 dB (relative to the strongest echo), so the RW-CFAR threshold (−11 dB) lies above the weak reflector (−13 dB) and removes it, while the OS-CFAR threshold (−14 dB) keeps it. The advantage is modest here, 2.4 dB, because the sidelobes of the strong reflectors raise the median as well, and the lower threshold also lets more of the sidelobes through (Fig. 9c). OS-CFAR costs a little more, a selection of the median among the directions per range cell instead of a sum, which is still cheap.

Two other variants are worth knowing. A two-dimensional cell-averaging CFAR uses as reference the energyscape smoothed with a Gaussian over both range and direction, so that it adapts locally. And for three-dimensional images, the lab's RTISSLAM 3D generalizes RW-CFAR to a local CFAR on the sphere of directions, which averages over neighboring directions with a heat kernel. In the limit of infinite smoothing, that average covers all directions, and it becomes RW-CFAR again.

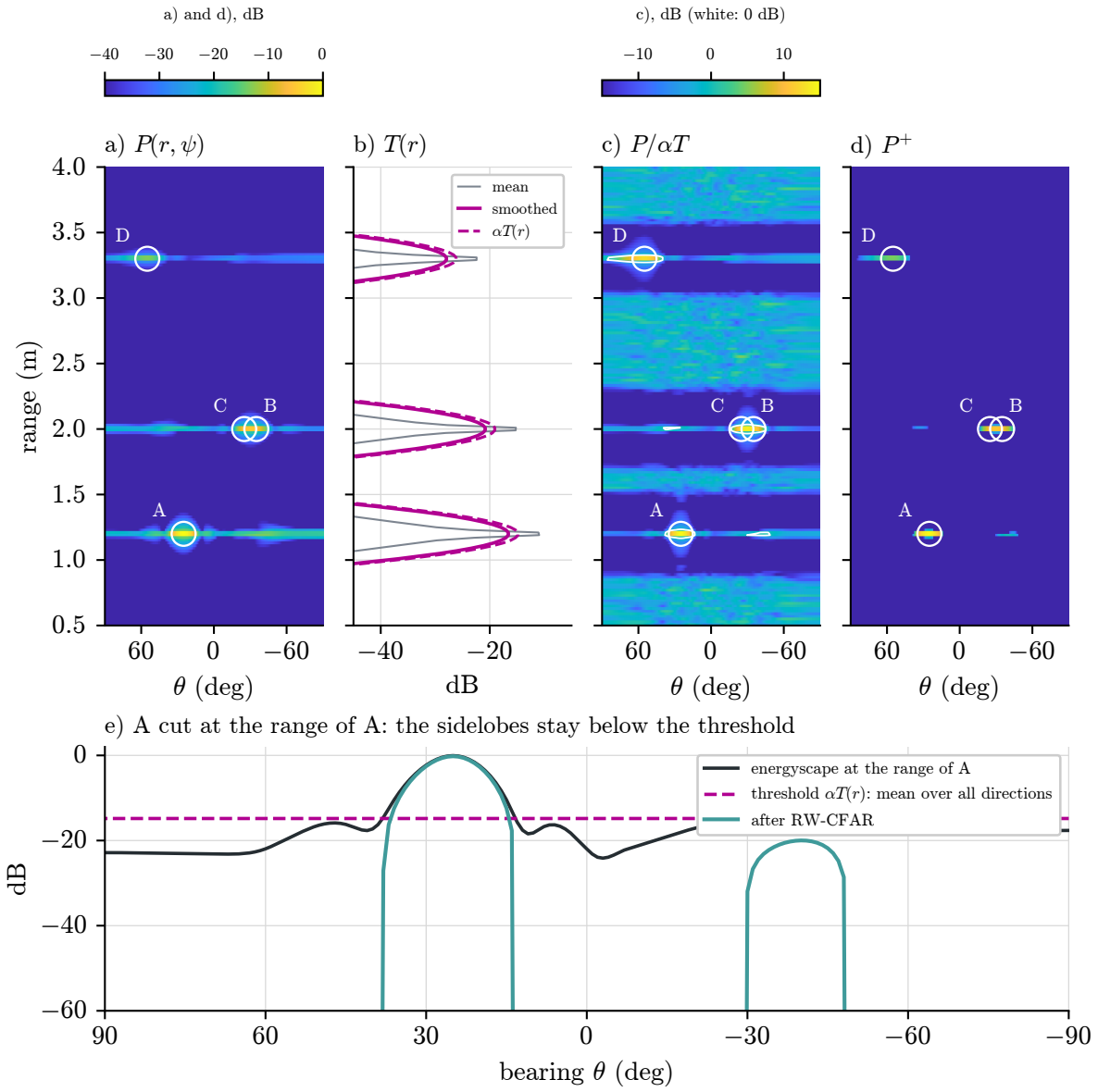


Figure 8: The energyscape and RW-CFAR step by step (range over bearing, 2 cm cells, bearing increasing to the left, as seen by the sensor; circles at the true positions of A to D). a) The energyscape $P(r, \psi)$. b) The reference $T(r)$, the mean over all directions (grey), smoothed along range (magenta), and the threshold $\alpha T(r)$ (dashed). c) The ratio $P/\alpha T$, with the 0 dB line in white where the energyscape is above −40 dB. d) What survives, P^+, on the color scale of a). e) A cut at the range of A: the energyscape, the threshold, and what survives.

step	what it does	what it buys
decimation	1-bit streams to 450 kHz samples	ordinary signals, 10 times fewer samples
matched filter	correlates with the call	range resolution ($c/2B$), about 21 dB of gain
beamforming	aligns and averages 32 microphones	direction, about 15 dB of gain
envelope	magnitude and zero-phase low-pass	energy instead of oscillation, fewer samples
CFAR	compares with a range-wise threshold	detections, without sidelobes and clutter

Table 1: The steps of the eRTIS pipeline and what each one buys.

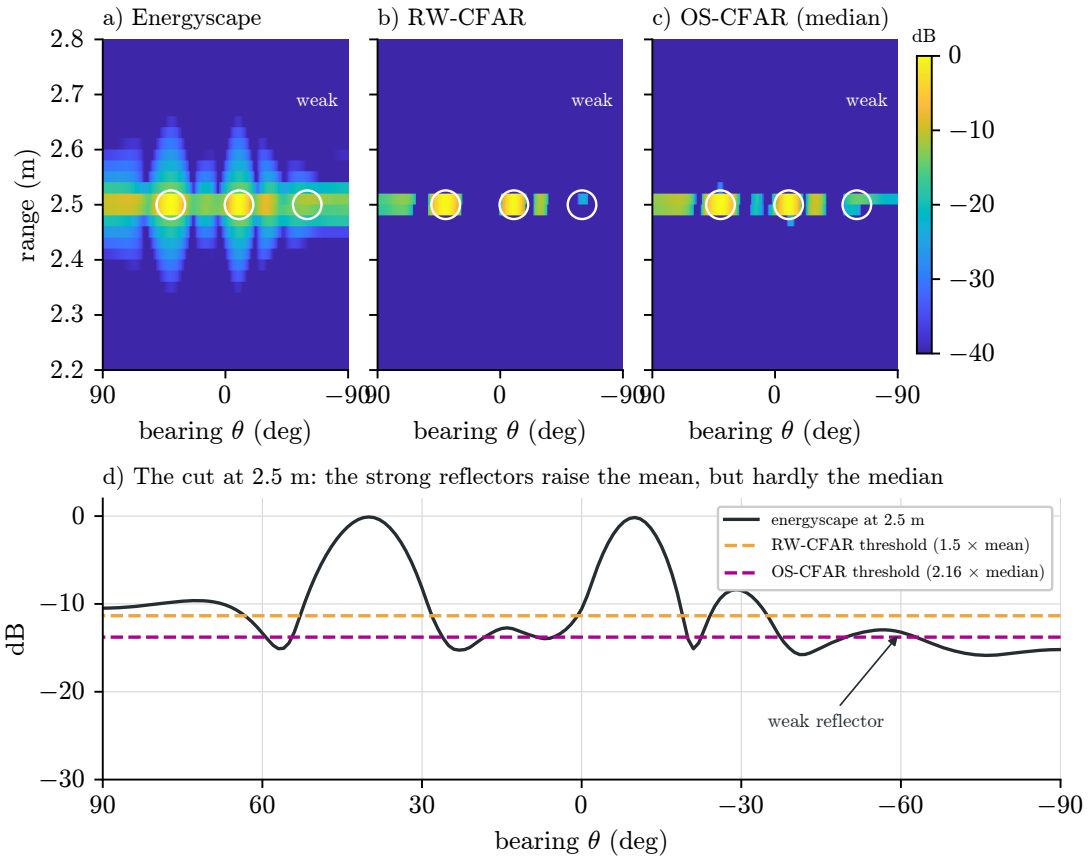


Figure 9: Target masking. Two strong reflectors and one 12 dB weaker share the range of 2.5 m (circles). a) The energyscape. b) RW-CFAR removes the weak reflector. c) OS-CFAR (median) keeps it, together with more of the sidelobes. d) The cut at 2.5 m with both thresholds.

Key idea

A single microphone hears long, overlapping echoes without any direction. The matched filter compresses every echo into a short pulse and so gives range, the beamformer combines the 32 microphones into beams and so gives direction, the envelope turns the result into an image of echo energy, and CFAR picks out the reflectors from their sidelobes and the clutter. Every step also adds sensitivity.

Exercise

1. The eRTIS software's default call sweeps from 25 to 50 kHz. What range resolution does it give, and what happens to the gain of the matched filter if the call stays 2.5 ms long?
2. Two reflectors lie at the same range, 5° apart. Would a larger array separate them? How large, if the beam width scales with the wavelength divided by the aperture?
3. With RW-CFAR, a single very strong reflector raises the threshold at its range. How strong must it be, relative to a weak reflector at the same range, to hide it, if its main lobe covers 10 of the 181 directions and $\alpha = 1.5$? How does the answer change with OS-CFAR?
4. The eRTIS software rounds the beamforming delays to whole samples at 450 kHz. How large is the resulting phase error at 75 kHz, and why might it matter less after the envelope?

Further Reading

- R. Kerstens, D. Laurijssen and J. Steckel, "eRTIS: A fully embedded real time 3D imaging sonar sensor for robotic applications," IEEE ICRA, 2019.
- G. L. Turin, "An introduction to matched filters," IRE Trans. Inform. Theory, 6(3), 311–329, 1960.
- H. L. Van Trees, *Optimum Array Processing*, Wiley, 2002.
- H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," IEEE Trans. Aerosp. Electron. Syst., AES-19(4), 608–621, 1983.