Principles of Echolocation
in Bats and Machines

Note 6

Multiplying Instead of Adding

Delay-multiply-and-sum beamforming and the coherence factor

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

Version 2026-11-18 16:23

1 The Baseline: Delay and Sum

An eRTIS sensor records the echoes of its call with 32 microphones. To find out where an echo comes from, we steer the array: for every direction θ we delay the microphone signals $x_i(t)$ so that a sound from that direction arrives at all of them at the same time, and then we add them,

$$y_{\mathrm{DAS}}(t,\theta) = \sum_{i=1}^{N} x_i\big(t + \tau_i(\theta)\big), \tag{1}$$

where $\tau_i(\theta)$ is the steering delay of microphone i. This is delay-and-sum (DAS) beamforming. In the direction of a reflector, the N copies of its echo add up in phase. In every other direction they add up with partly random phases, and part of the echo remains as a sidelobe. After matched filtering and envelope detection, every direction and range cell gets a level, and together they form the sonar image.

The eRTIS array is small. Its microphones span only about 6.6 cm horizontally (Fig. 1a), so its beam is wide, and with 32 microphones its sidelobes are high. For the call band of 20 to 80 kHz, a single reflector at 2 m produces a mainlobe of 8.0° at −3 dB and sidelobes up to -16 dB everywhere else (Fig. 3a, gray). A strong echo therefore spills over the whole image, and weak echoes can hide in its sidelobes. Making the array larger is not always an option on a robot, or on a bat. Delay-multiply-and-sum (DMAS) beamforming attacks the problem in the processing instead, and our implementation follows Jansen et al. [1].

2 Delay, Multiply and Sum

The idea of DMAS is to reward agreement between the microphones. Instead of adding the aligned signals, we multiply them in pairs (or triplets) and add the products. A product of two aligned copies of the same echo is large and positive. A product of two signals that do not agree, such as the misaligned echo in a sidelobe direction, or independent noise, changes sign at random and averages away.

Multiplying would also square (or cube) the amplitudes, so that the image would no longer scale with the echo strength. DMAS therefore first takes a signed root of every aligned signal,

$$s_i = \operatorname{sgn}(x_i)\,|x_i|^{1/n}, \tag{2}$$

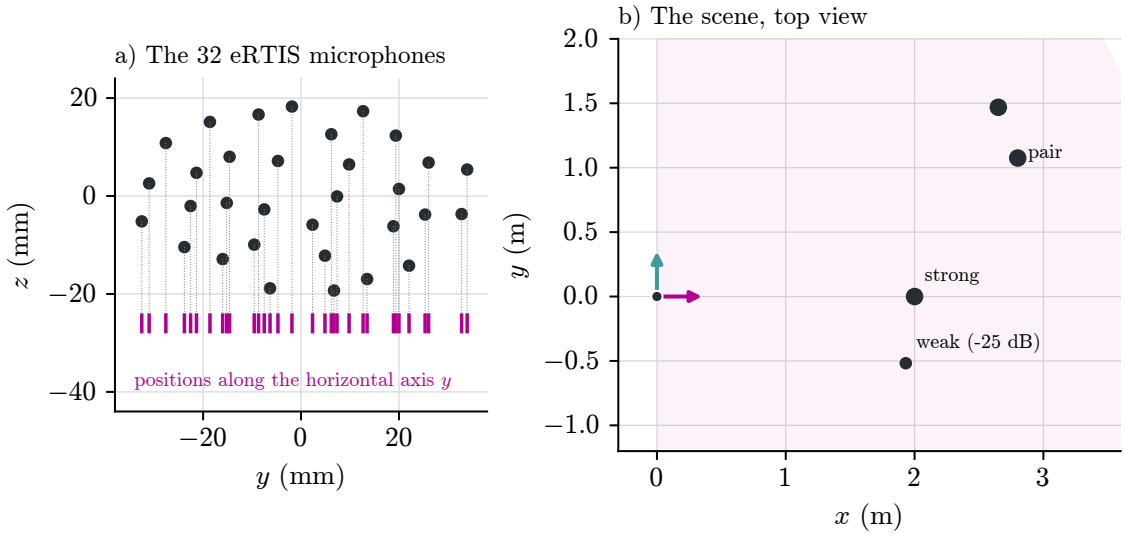


Figure 1: a) The 32 microphones of an eRTIS (dots), and their positions along the horizontal axis (magenta), which is all that counts for bearing in the horizontal plane. b) The simulated scene: a strong reflector at 2 m, a pair at 3 m, and a weak reflector (-25 dB) at the same range as the strong one. The sonar is drawn as a triad.

so that a product of n of them is again an amplitude. The DMAS output of order n is the sum of all products of n different microphones,

$$y_{\mathrm{DMAS}n} = \sum_{i_1<i_2<\cdots<i_n} s_{i_1} s_{i_2} \cdots s_{i_n}, \tag{3}$$

which is the elementary symmetric polynomial of order n of the s_i. For $N = 32$ microphones, that is 496 pairs for $n = 2$ and 4,960 triplets for $n = 3$, too many to compute one by one for every sample and direction. Fortunately, the same sums follow from the power sums $P_k = \sum_i s_i^k$ (Newton's identities):

$$y_{\mathrm{DMAS2}} = \tfrac{1}{2}\left(P_1^2 - P_2\right), \qquad y_{\mathrm{DMAS3}} = \tfrac{1}{6}\left(P_1^3 - 3P_1P_2 + 2P_3\right). \tag{4}$$

In general, the elementary symmetric polynomials e_k follow from the power sums with the recursion

$$e_0 = 1, \qquad e_k = \frac{1}{k}\sum_{i=1}^{k}(-1)^{i-1} e_{k-i} P_i, \tag{5}$$

which gives the two expressions above for $k = 2$ and 3, and any higher order in the same way. Our real-time eRTIS processing also uses order 5.

A worked example. Take four aligned and rooted signals at one instant, $s = (1, 2, -1, 3)$. The six pairs give $1\cdot 2 + 1\cdot(-1) + 1\cdot 3 + 2\cdot(-1) + 2\cdot 3 + (-1)\cdot 3 = 5$. With the power sums $P_1 = 5$, $P_2 = 15$ and $P_3 = 35$, the same result follows from $\frac{1}{2}(25 - 15) = 5$, without forming a single pair. For the four triplets, $\frac{1}{6}(125 - 3\cdot 5\cdot 15 + 2\cdot 35) = -5$, which again equals the sum of $1\cdot 2\cdot(-1)$, $1\cdot 2\cdot 3$, $1\cdot(-1)\cdot 3$ and $2\cdot(-1)\cdot 3$. Note the negative result: one microphone disagrees in sign, and the triplets that contain it pull the output down. That is exactly the mechanism that suppresses incoherent signals.

The cost. Forming all products directly would take $\binom{N}{n}$ products per output sample and direction: 496 for $n = 2$, 4,960 for $n = 3$ and 201,376 for $n = 5$ with our 32 microphones. Through the power sums, the cost is N signed roots, about $N(n-1)$ multiplications for the powers, Nn additions for the sums, and a handful of operations to combine them, so it grows only linearly with N and n. For 32 microphones and order 5, that is about 32 roots and 300

other operations per sample and direction, a few times the 31 additions of DAS. The steering delays, usually the most expensive part, are the same for all methods. The roots are the costly step, and a fast implementation approximates them, for example with a lookup table. The coherence factor adds N squares and one division.

The coherence factor. A second, related tool measures directly how well the microphones agree,

$$\mathrm{CF}(t,\theta) = \frac{(\sum_i x_i)^2}{N \sum_i x_i^2}, \tag{6}$$

the ratio of the coherent to the total energy. It is 1 when all aligned signals are equal and close to $1/N$ for noise. Multiplying the DMAS output by the CF suppresses incoherent contributions once more.

3 What It Does to the Image

Figure 2 shows the same simulated recording (Fig. 1b, with sensor noise), processed six times, and Table 1 lists the numbers. The images are computed exactly as in the lab's processing: matched filter, steering delays, beamformer, envelope, and the mean power per range cell.

Sidelobes and noise. The sidelobes drop from -16 dB with DAS to -31, -41 and -56 dB for DMAS of order 2, 3 and 5, and to -71 and -90 dB with the coherence factor (Fig. 3a). The noise floor of the image falls from -56 to between -72 and -123 dB. This is what makes DMAS images look so clean. Each step in the order buys roughly another 10 to 15 dB.

Resolution. The mainlobe narrows from 8.0° to 5.0° (order 2 and 3) and 4.0° (order 5, and with CF). The pair, two reflectors 8° apart, is barely separated by DAS (a dip of only -1.5 dB between them) and clearly separated by DMAS (-3.1 to -5.4 dB, Fig. 3b). The gain in resolution is real, but modest compared with the gain in sidelobes: the mainlobe is still set by the size of the array. One caveat holds for every beamformer: when we placed the two reflectors at exactly the same range, their echoes arrived at the array at the same time and merged into a single peak between them, whatever the method. The pair in the figure is 3 cm apart in range.

	DAS	DMAS order 2	DMAS order 3	DMAS 3 × CF	DMAS order 5	DMAS 5 × CF
mainlobe width at −3 dB (deg)	8.0	5.0	5.0	4.0	4.0	4.0
peak sidelobe level (dB)	-16	-31	-41	-71	-56	-90
noise floor of the image (dB)	-56	-72	-83	-106	-100	-123
dip between the pair (dB)	-1.5	-3.1	-3.9	-4.8	-4.8	-5.4
weak reflector alone in its range cell (dB)	-29.7	-29.7	-29.7	-29.7	-29.7	-29.7
weak reflector next to the strong one (dB)	-21	-33	-43	-83	-59	-100

Table 1: DAS and DMAS for the simulated eRTIS (32 microphones, 20–80 kHz). All levels in dB relative to the strongest reflector in the same image. The weak reflector has a true level of -25 dB at 1 m, which becomes −29.7 dB at its own range of 2.6 m.

Key idea 1

DMAS rewards agreement between the microphones. Coherent echoes survive the products, incoherent sidelobes and noise average away. The result is a narrower mainlobe and sidelobes and noise tens of dB lower, for little extra computation.

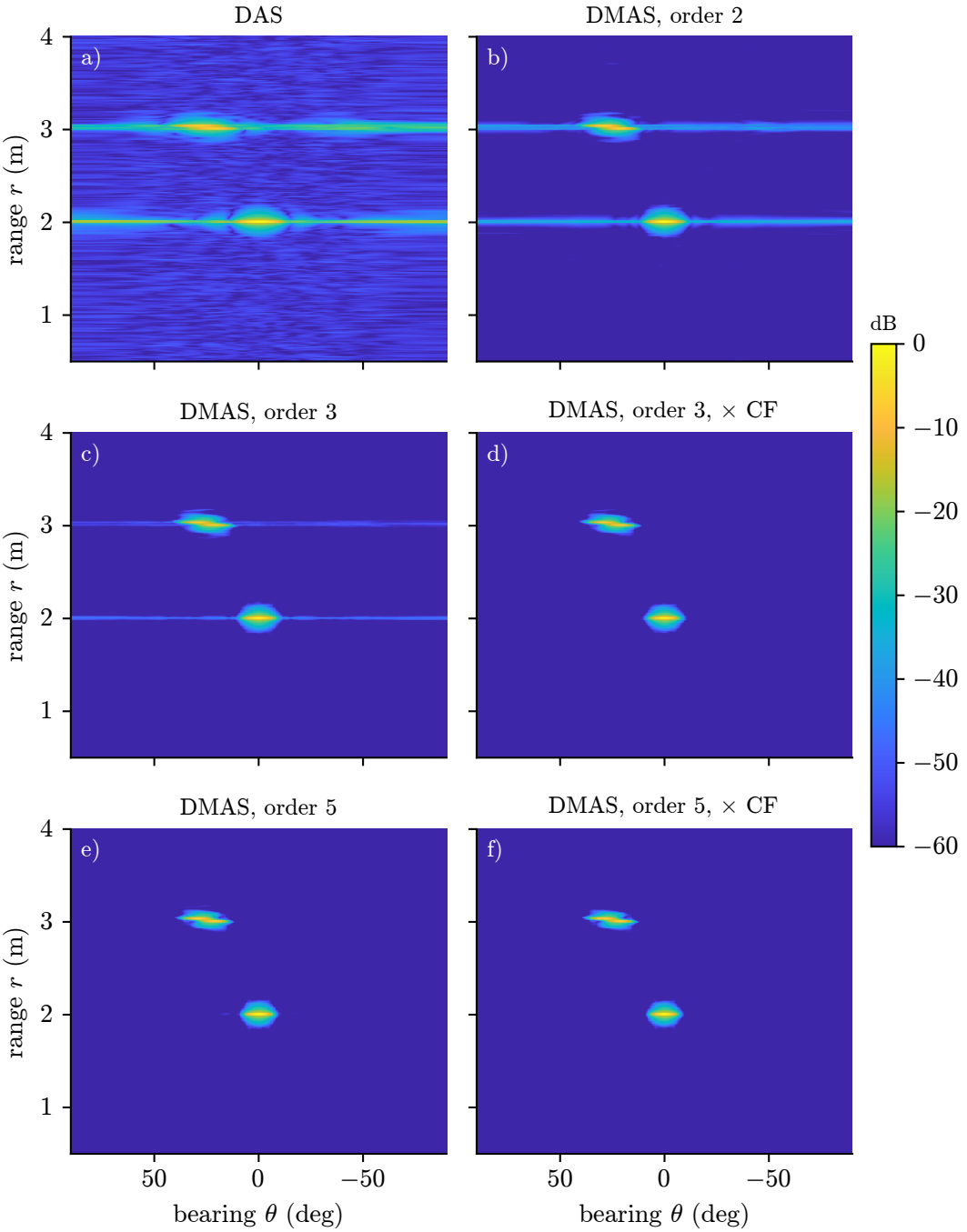


Figure 2: The same recording, beamformed six times (each image in dB relative to its own maximum). From DAS to DMAS of order 5 with the coherence factor, the sidelobes along the range lines of the reflectors and the noise floor drop by tens of dB, and the reflectors become compact. The weak reflector at $-15°$, at the same range as the strong one, is lost in the sidelobes with DAS and suppressed with DMAS.

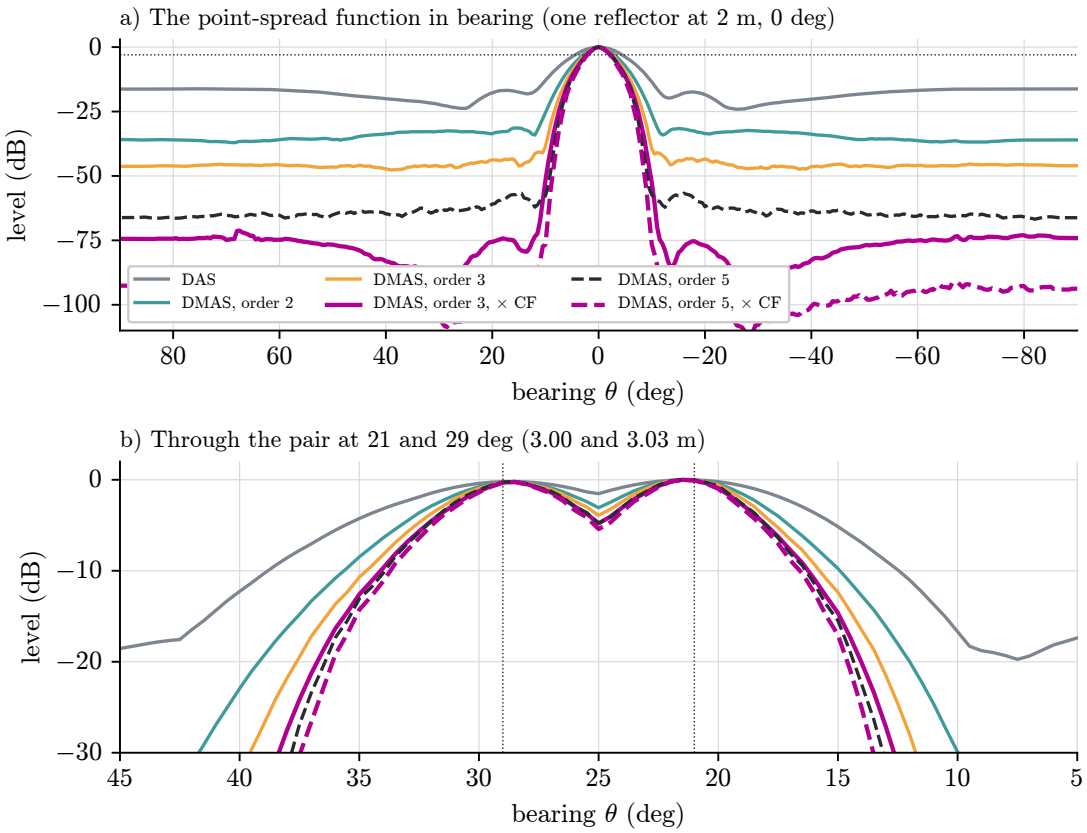


Figure 3: a) The point-spread function in bearing of every method, for one reflector at 2 m straight ahead (no noise). b) A cut through the pair, from the recording of Fig. 2. Dotted lines: the true bearings.

4 The Price

DMAS is not free of side effects, and the main one is subtle. For a reflector that is alone in its range cell, DMAS keeps the level exactly: the weak reflector, placed at its own range, appears at -29.7 dB with every method (Table 1). The signed roots make sure that the output is again an amplitude. But when the weak reflector sits at the same range as a strong one, the products mix the two. In the direction of the weak reflector, the strong echo arrives misaligned on all microphones, its contribution changes sign from microphone to microphone, and through the products it drags the weak echo down with it. The weak reflector then appears at -33, -43 and -59 dB (order 2, 3 and 5) instead of -21 dB, and the coherence factor, which sees the incoherent strong echo, suppresses it to -83 and -100 dB, effectively removing it from the image.

A two-source analysis. We can isolate the mechanism with a minimal model. At one frequency, a weak echo of amplitude w arrives aligned on all 32 microphones, and a strong echo of amplitude S arrives at the same time but misaligned, with a random phase on every microphone, as in a sidelobe direction. We measure how much of the weak echo is left in the output (its correlation with its own waveform), relative to its output when it is alone, averaged over 2,000 random phase patterns (Fig. 4). For DAS, the answer is always 0 dB: a linear beamformer adds the strong echo, but it does not touch the weak one. For DMAS, the weak echo is attenuated as soon as the strong one becomes comparable. With the strong echo 10 dB above the weak one, the weak echo loses -18 dB with DMAS of order 2, -29 dB with order 3 and -61 dB with order 5. The reason lies in the roots and the products. Every rooted signal s_i is dominated by the strong echo, and the weak echo only shifts it slightly, by an amount that shrinks as the root compresses the large values. In a product of n signals, that small coherent shift is multiplied by $n-1$ factors of the strong echo with random signs, which average away. The

higher the order, the more factors, and the harder the weak echo is crushed. In the simulated scene, the sidelobe of the strong reflector lies about 6 dB above the weak one at its position, and the measured losses follow the same ordering.

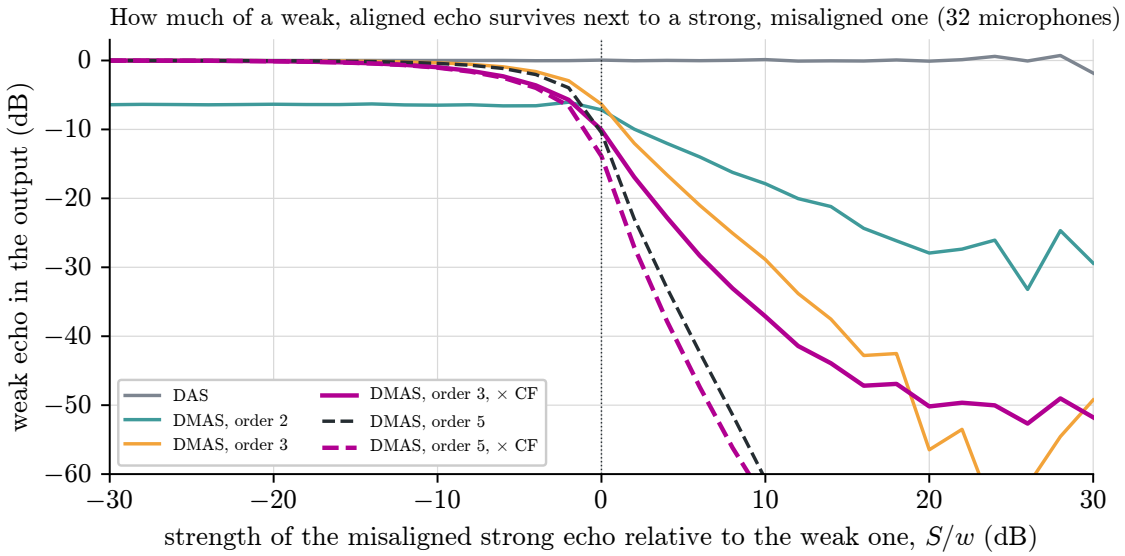


Figure 4: The two-source model: a weak echo, aligned on all 32 microphones, and a strong echo with a random phase on every microphone. The weak echo's share of the output, relative to its output alone, as a function of the strength of the strong echo (2,000 random phase patterns per point).

DMAS images thus have a large dynamic range for isolated reflectors, but a weak echo next to a strong one at the same range is attenuated, and the levels in a DMAS image are no longer a linear sum of what the reflectors contribute.

Two further costs are worth knowing. The DMAS output is a nonlinear function of the signals, so the classic tools of linear beamforming, such as computing the image of a scene as the sum of the images of its reflectors, or deconvolving the point-spread function, no longer apply directly. And although Newton's identities keep the computation linear in N, DMAS needs the aligned signals of all microphones at every sample and direction, so it cannot be reduced to a single weighted sum as DAS can.

Key idea 2

DMAS keeps the level of an isolated reflector, but it is nonlinear. A weak echo at the same range as a strong one is attenuated, and the coherence factor can remove it altogether. Clean images come at the price of quantitative amplitudes.

5 Another Route to Resolution

DMAS improves the image by making the beamformer nonlinear. There is a second route, which stays linear in the data but chooses the weights of the microphones from the data themselves. Adaptive beamformers such as the minimum-variance distortionless-response (MVDR, or Capon) beamformer place nulls on the strong interfering echoes, and subspace methods such as MUSIC separate the directions of the reflectors from the noise. They can resolve reflectors far closer than the width of the beam, but they need a good estimate of the covariance between the microphones, and they are sensitive to errors in the array model. These methods are the subject of Note 5, *Beyond Delay-and-Sum*.

Exercise

1. Show that $\frac{1}{2}(P_1^2 - P_2)$ equals the sum of all products $s_i s_j$ with $i < j$. Derive the expression for order 3 in the same way.
2. Why does the signed root keep the output of DMAS proportional to the amplitude of a single reflector? What happens to the levels without the root?
3. Two reflectors at exactly the same range merge into one peak, whatever the beamformer. Explain why, and what a robot could do to separate them (think of motion, see Note 2).
4. For which tasks would you prefer DAS over DMAS, despite its sidelobes?
5. Count the operations per sample and direction of DMAS of order 4 for 32 microphones, once directly and once through the power sums.

References

[1] W. Jansen, D. Daems and J. Steckel, "Delay-multiply-and-sum beamforming for real-time in-air acoustic imaging," arXiv:2511.09165, 2025.

[2] H. B. Lim, N. T. T. Nhung, E.-P. Li and N. D. Thang, "Confocal microwave imaging for breast cancer detection: Delay-multiply-and-sum image reconstruction algorithm," IEEE Trans. Biomed. Eng., 55(6), 1697–1704, 2008.

[3] G. Matrone, A. S. Savoia, G. Caliano and G. Magenes, "The delay multiply and sum beamforming algorithm in ultrasound B-mode medical imaging," IEEE Trans. Med. Imaging, 34(4), 940–949, 2015.