ENP162 Assignment 2: Signal Detection

Kat Allen

October 10, 2021

1 Experiment Design

This is a signal detection experiment using grids of stars * and dots .
The user is informed in advance that there will be either the same amount of “A” and”B” prompts, 3 times as many As as Bs, or 1/3 as many As as Bs, and is given examples of both A and B prompts. The user is told that most, but not all, B prompts have more stars than A prompts.

The experiment uses **PEBL: The Psychology Experiment Building Language**[1]

2 Data Analysis

Table 2 shows the results of the 50:50 experiment. For analysis, the “hit rate” is defined with B being the detection goal, so if the user chose B and the stimulus was actually B, that’s a hit, if they chose A and it was actually a A stimulus, that’s a “Correct Reject”. A “Miss”, therefore, is when they chose A but the stimulus was B, and a “False Alarm” is if they chose B but the displayed stimulus was A. The sensitivity (d’) is middling (but better than both the 3:1 and 1:3 scenarios!)

Hit Rate	73.13%	38.75%	False Alarm Rate
Miss Rate	26.88%	61.25%	Correct Reject Rate

Sensitivity (d’)	.91
Criterion	3.31

Table 1: Confusion Matrix for the SDT demo experiment with a 50:50 ratio

This is a difficult task for me, with a moderately high false alarm rate. Digging deeper into the data (presented as figure 2), you can see that more clearly: while there were 8 times where an “A” signal was presented but the % fill was over 54%, I identified all but one of them as “B”. Similarly, in the cases where a “B” prompt was offered but its % fill was less than 44%, only once (of the 4 prompts) did I identify it correctly as a B.

The data for 45-55% in both cases is much messier. 36/110 (32%) of the time the prompt was a B and in that range , I misidentified it as an A. 52/93 times the prompt was an A and in that range, I misidentified it as a B (56%). That is a lot of errors - overall 43%, which is barely better than a coin flip.

Table 2 and figure 3 show the confusion matrix and ROC curve for the 3:1 scenario, where there were more A prompts than B. In this case, the hit rate is very low (33%!) but the correct reject rate is high - I was more likely to misidentify a B prompt as A, even after being told that there were more A prompts.

[1] `http://pebl.sourceforge.net`

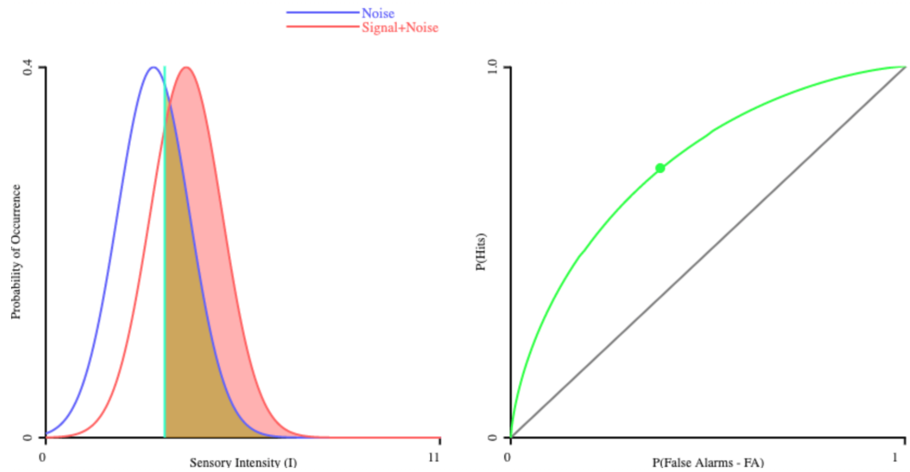

Figure 1: 50-50 ROC curve

Hit Rate	33%	18.33%	False Alarm Rate
Miss Rate	67%	82%	Correct Reject Rate

Sensitivity (d')	.46
Criterion	3.91

Table 2: Confusion Matrix for the SDT demo experiment with a 3:1 ratio

Table 2 and figure 4 show the results of the scenario where there were 3 times as many B prompts as A (so the signal occurred more freqquently.) In this case, the hit rate was very good (85%!) but the false alarm rate was also very high - I didn't identify the less frequent A prompts very well, and set a low criterion for what must be B.

Hit Rate	85%	69%	False Alarm Rate
Miss Rate	15%	31%	Correct Reject Rate

Sensitivity (d')	.52
Criterion	2.5

Table 3: Confusion Matrix for the SDT demo experiment with a 1:3 ratio

The last version of the experiement involved improving my ability to discern A from B stimuli by reducing the mean number of stars in an A prompt (from 46 to 40), and increasing the mean number of stars in a B prompt (from 54 to 60). This makes the task easier by making the two prompts more different.

Table 2 and figure 6 show the dramatic difference - a hit rate of 85% and a false alarm rate of only 20%! As predicted, the sensitivity is increased to 1.88.

Hit Rate	85%	20%	False Alarm Rate
Miss Rate	15%	80%	Correct Reject Rate

Sensitivity (d')	1.88
Criterion	3.83

Table 4: Confusion Matrix for the SDT demo experiment with a 50:50 ratio, but with $A_{mean} stars = 40$ and $B_{mean} = 60$

Count of choice		choice		
Signal	% fill	A	B	Times Presented
A	33	1		1
	34	1		1
	35	1		1
	36	5		5
	37	3	1	4
	38	4		4
	39	5		5
	40	2		2
	41	4		4
	42	9		9
	43	13	1	14
	44	8	4	12
	45	9	8	17
	46	6	4	10
	47	7	4	11
	48	7	8	15
	49	4	9	13
	50	4	3	7
	51	1	8	9
	52	1	1	2
	53	2	4	6
	54		2	2
	55		1	1
	56		1	1
	57	1	1	2
	58		2	2
A Total		98	62	160
B	39	1		1
	42	2		2
	44		1	1
	45		4	4
	46	1	3	4
	47	4	4	8
	48	5	8	13
	49	2	4	6
	50	9	6	15
	51		4	4
	52	6	7	13
	53	4	12	16
	54		11	11
	55	5	11	16
	56	3	7	10
	57		8	8
	58		5	5
	59		5	5
	60	1	7	8
	61		4	4
	62		1	1
	63		3	3
	64		1	1
	68		1	1
B Total		43	117	160

Figure 2: 50-50 data pivot table

3 Discussion

This was a difficult task for this subject/experimentor! It was difficult even with the pre-existing knowledge of the frequency of the signals. A rare signal was harder to identify than a frequent one, but knowing a signal would happen frequently caused many false alarms.

It became much easier in the scenario where the prompts were more different, and differentiating them further should only help more in improving sensitivity.

4 Applications to System Design

To apply this to system design, let's take an example system. In space, it's very important not to run into anything. This sounds obvious, but it's worse than you think.

Because it can cost thousands of dollars per pound of payload/spacecraft weight to launch something into orbit, spacecraft are often made of very very thin materials.

In addition, because there is no drag from air resistance, a very small object that goes flying off (like a dropped screw, or a piece of a decomissioned satellite that has broken up) will keep whatever its initial velocity was. Orbital velocities are necessarily very large, in order to keep the object in orbit, with higher velocities the closer you are to the earth. The ISS orbits the earth at 28,165 kilometers per hour, so a 1 gram screw that fell off the ISS might hit something with a force of $F = mv^2$ (so with some unit conversions) $= .001kg * (7823.61m/s)^2 = 78.23Newtons$. If you aren't familiar with what a Newton feels like, that's like having 7 kilograms fall on you. Not a nice feeling!

And there are a **lot** of pieces of space junk out there. The DoD's Space Surveillance Network tracks over 27,000 pieces of orbital debris, most of which are softball sized or larger.[2]. They estimate 100 million pieces in the millimeter size range, just in the most

[2] https://www.nasa.gov/mission_pages/station/news/orbital_debris.html

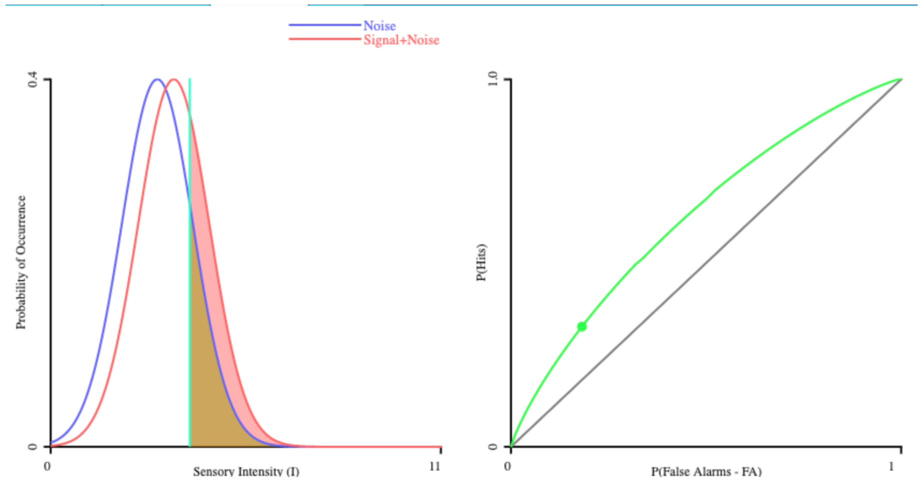

Figure 3: More A: the 3:1 ROC curve

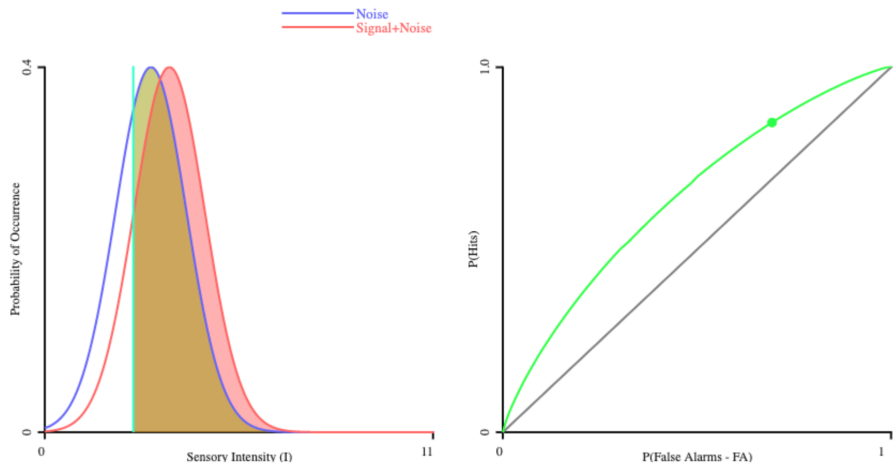

Figure 4: More B: the 1:3 ROC curve

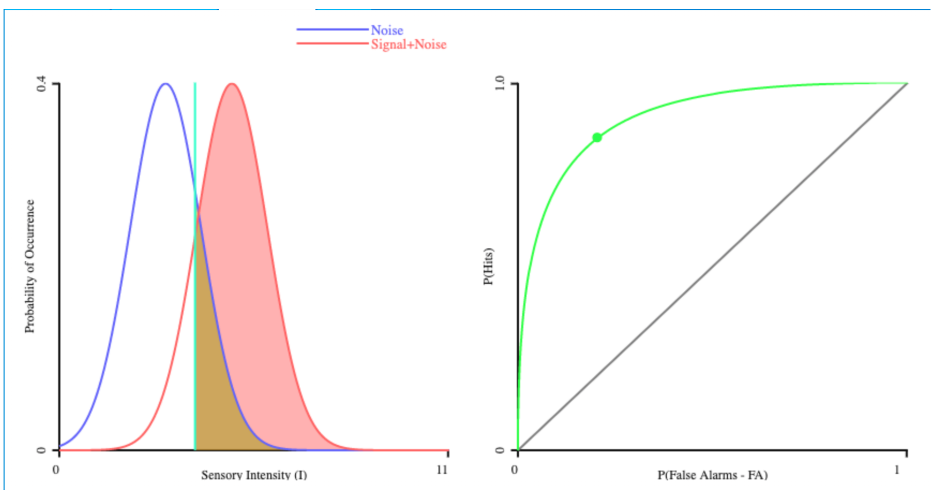

Figure 5: 50-50 ROC curve with higher sensitivity

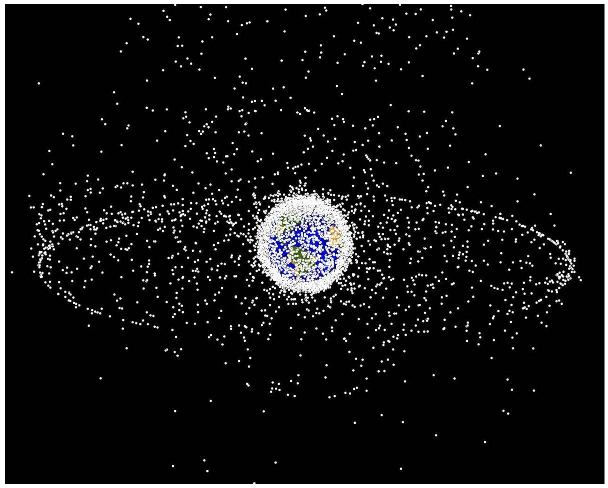

Figure 6: An artist's conception of space debris around the Earth

useful part of Earth's orbit for human space systems.

So with all that "junk" out there, a meteorite/space junk identification system whose job is identifying pieces of "stuff" that might be on a collision course with your spacecraft is pretty important. The operator needs to know with a pretty high certainty if an object is going to encounter the spacecraft, and how soon.

If an automated system is used to flag potential threats, signal detection can help by setting the automatic system's criterion at an appropriate place to balance the threat from potential impacts against the time demands on the user to investigate false alarms.

Even once a potential threat has been identified, the system needs to alert the user about the threat and they need to analyze it. A display for identifying micrometeorite threats needs to have a clear and uncrowded display to avoid misidentification - perhaps in the list of all tracked items, the potential threats could be identified in bold red while the non-flagged items are in faint grey. This will improve the user's sensitivity to detecting the output of the automation, which will allow them to more quickly evaluate the threat and prepare for it if needed.

5 Acknowledgements

Visualization Tool from "Signal Detection Theory and the Receiver Operating Curve": `https://isle.hanover.edu/Ch02Methods/Ch02SDT_ROC.html`

PEBL: The Psychology Experiment Building Language `http://pebl.sourceforge.net`

NASA's Space Debris and Human Spacecraft page: `https://www.nasa.gov/mission_pages/station/news/orbital_debris.html`