

Counting the cost: validating discrete driver–IVIS interaction count as a practical, equipment-free index of IVIS distraction

Angus McKerral^{1,2}, Jon-Paul Cavallaro¹, Rachael A. Wynne¹, Alexander Thorpe¹, Ami Eidels¹, Scott D. Brown¹, Guy E. Hawkins¹, Kristen Pammer^{1,2}

¹The University of Newcastle ²The University of Melbourne

Funded by the Australian Automobile Association Road Safety Research Program (DD2021-02). Partners: ABMARC (AARC proving ground) and Red Scientific (D. Strayer, J. Cooper).

Validated measures of IVIS demand do not scale



Criterion measures

- DRT (ISO 17488): residual attentional capacity – **cognitive demand**
- VOT: total shutter open time – **visual demand**
- Both require specialised equipment within a controlled setting and trained assessors

Research gap

Task structure predicts demand (Getty et al., 2018), but no cross-validated model links a **multimodal step count** to DRT and VOT

Three questions

1. Is step count stable across test contexts?
2. Does it predict the criterion measures out of sample?
3. What does each step cost?

20 drivers across 12 infotainment systems



Participants

- 20 drivers (15 male), aged 26–50 (M = 32)
- ~5h driving a week; full Australian licence
- $n=19$ (one excluded for protocol non-compliance)
- Three vehicles each over two consecutive days
- Testing at the AARC proving ground, Victoria, winter 2023

System	Vehicle	Category
Android Auto	Toyota Corolla 2021	Smartphone mirroring
Apple CarPlay	Toyota Corolla 2021	Smartphone mirroring
Audi MMI	Audi Q7 2020	Large SUV (European)
BMW iDrive 7.0	BMW 330i 2023	European sedan
Cupra MIB3	Cupra Ateca 2022	Medium SUV
Ford SYNC4A	Ford Ranger 2023	4WD ute
Mazda Connect	Mazda CX-9 2021	Large SUV
MG iSMART*	MG ZS 2022	Small SUV
NissanConnect	Nissan X-Trail 2023	Medium SUV
Android Automotive OS	Polestar 2 2022	Electric
Tesla touchscreen	Tesla Model Y 2022	Electric
Toyota Link	Toyota Corolla 2021	Small car
Toyota Link	Toyota Hilux 2019	4WD ute

IVIS interaction tasks



Task	Example instruction	Complete when
Initiate call	“Please call Alex”	Call connects to the correct contact
Initiate text	“Text Alex and ask if he could buy you a coffee”	Message sent; driver says “Done”
Media	Play a named song from a 15-song playlist	Playback starts
Navigation	Set a destination within 40 km (e.g. Bells Beach)	Route found
Radio	Tune to a named local AM/FM station	Station tuned
Receive call	Answer the researcher’s call	Driver says “Hello”
Receive text	Have the system read an incoming text	Reading complete

On-track vs stationary testing method



On the track

4.2 km closed highway circuit, 80 km/h, safety driver

Tactile DRT: vibration every 3–5 s, finger switch on the wheel

Outcome: **DRT miss rate** (cognitive)

In the parked car

VOT: occlusion goggles, 1,500 ms self-initiated views

Outcome: **total shutter open time** (visual)

Tracking task on a windscreen-mounted laptop; outcome: **tracking error**

A step is any discrete driver input



What counts as a step

- Touchscreen taps
- Physical buttons, dials and rotary controllers
- Steering-wheel controls
- Voice commands, including wake words
- Verbal responses to system prompts

Failed or repeated inputs count: a misrecognised command still costs attention

Coded from video by trained observers

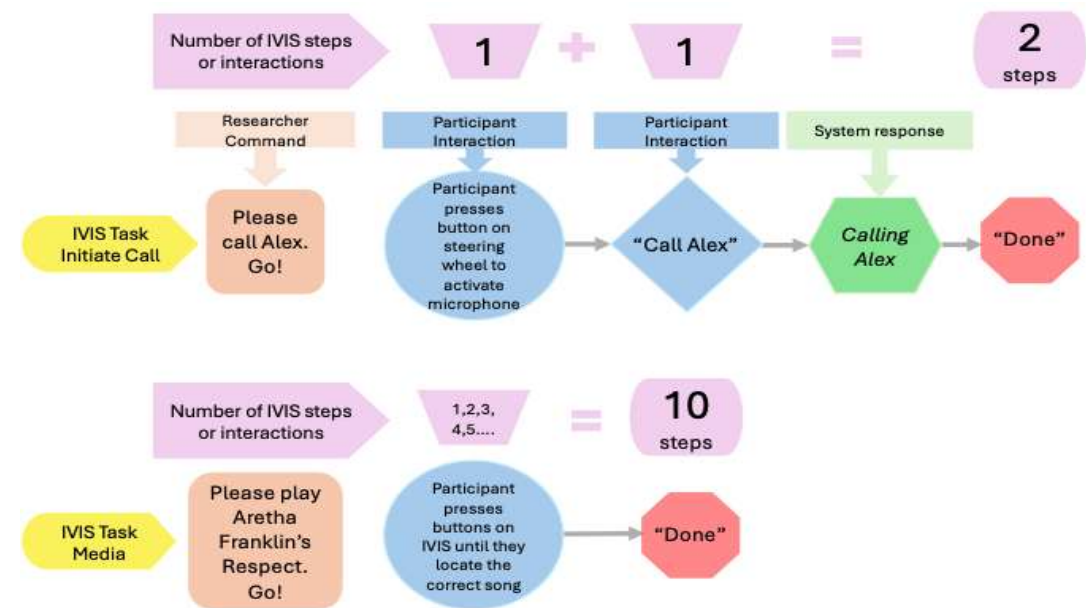
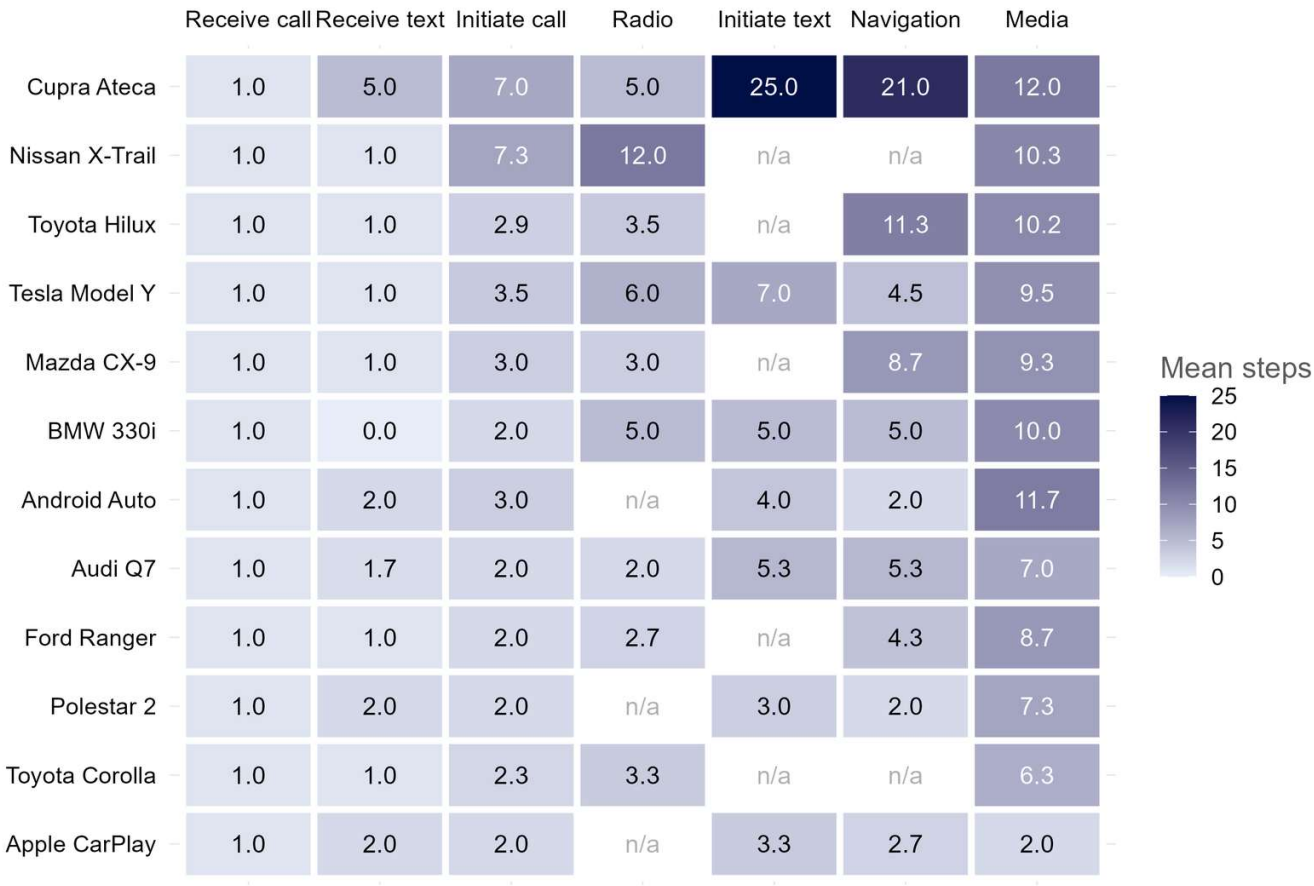


Fig 1. Voice call = 2 steps; browsing for a song = 10 steps

Variation in step count across task and system



Steps per participant × system × task

Context	Mean (SD)	Median	Range
On-track	4.37 (4.27)	2	0–25
Parked VOT	4.69 (5.10)	3	0–43
Tracking	4.51 (4.98)	2	0–39

Fig 2. Mean steps per system and task, on-track drive; n/a = task not supported

Cross-context validation of step counts between stationary and track testing

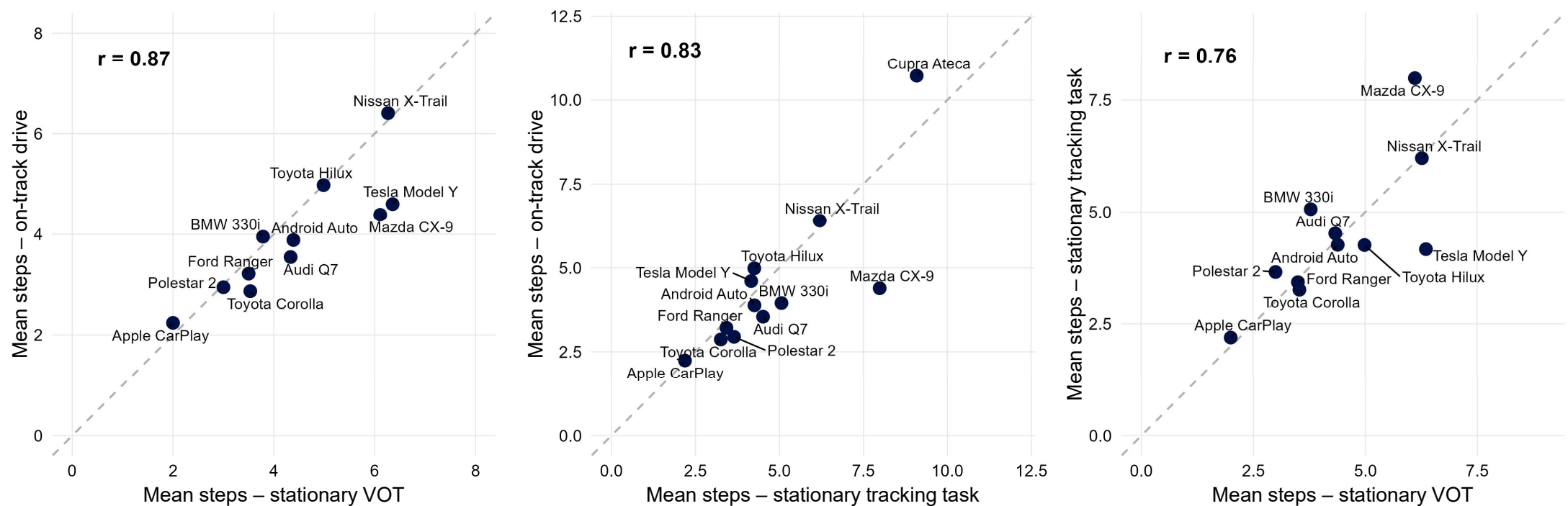


Fig 3. Model-estimated mean steps per system in each context; dashed line = equality. Cupra Ateca excluded from VOT comparisons (system failures).

Model selection



Selected on leave-one-out cross-validated r (over in-sample fit)

Specification	lme4 formula	DRT r	VOT r
1 Random intercept	Outcome $\sim$ nSteps + (1 participant)	.71	.65
2 + system-specific slopes	Outcome $\sim$ nSteps + nSteps:IVIS + (1 participant)	.73	.65
3 Random intercept and slope	Outcome $\sim$ nSteps + (nSteps participant)	.75	.63

Model 1 selection

Predictive stability: highest out-of-sample r for the VOT; extra parameters overfit

Standardisation: one specification for DRT and VOT; Model 3's DRT gain did not justify divergent methods

Robustness: Model 1 converged in every fold; Model 3 often did not

Parsimony: one system-agnostic coefficient per step

One input predicts cognitive, visual and visuo-manual demand out of sample

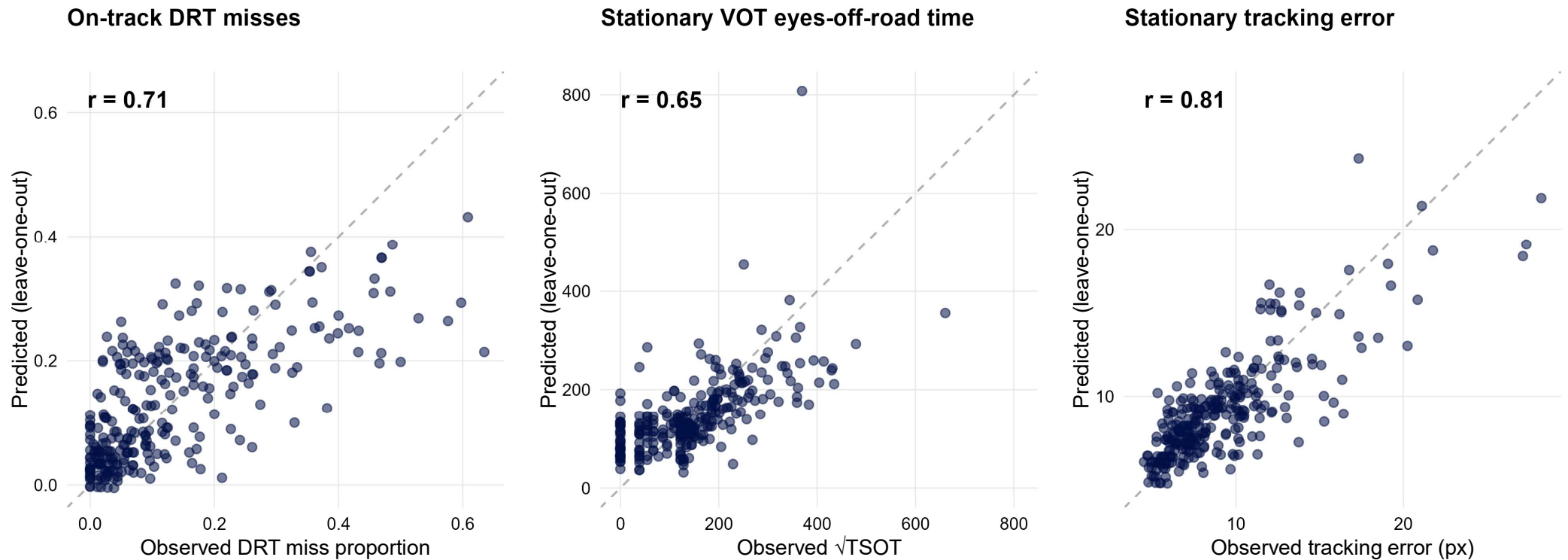


Fig 4. Leave-one-out predictions from the step-count model. DRT and VOT: N = 276 cells, Cupra Ateca excluded; tracking: N = 290, Cupra included; 19 participants.

Each step adds ~1 pp of DRT misses and 0.3–0.8 s of eyes-off-road time



Cognitive: predicted DRT miss rate

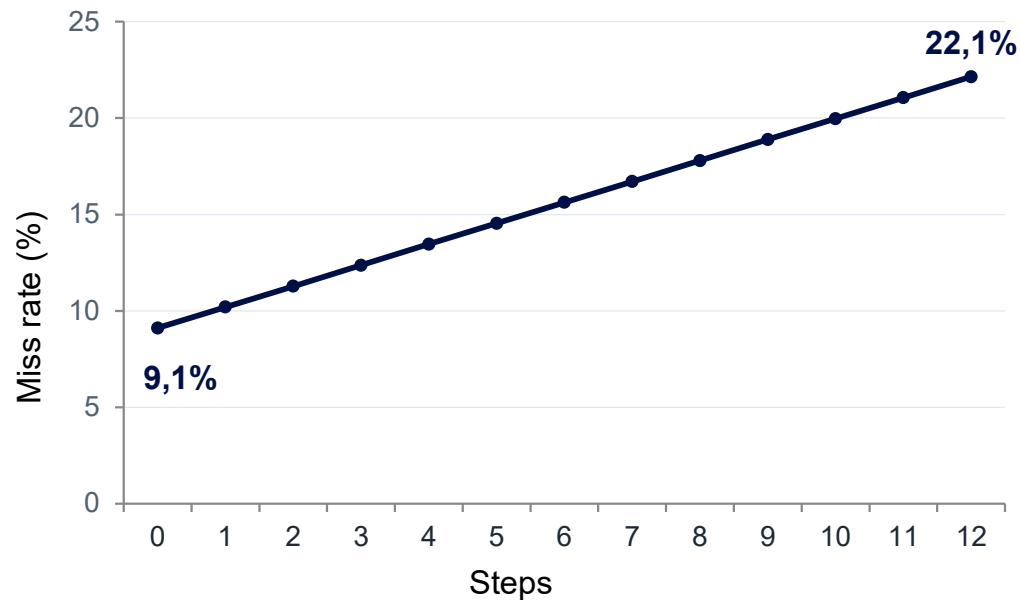


Fig 5. Predicted DRT miss rate (%) by number of steps

Visual: eyes-off-road time per task

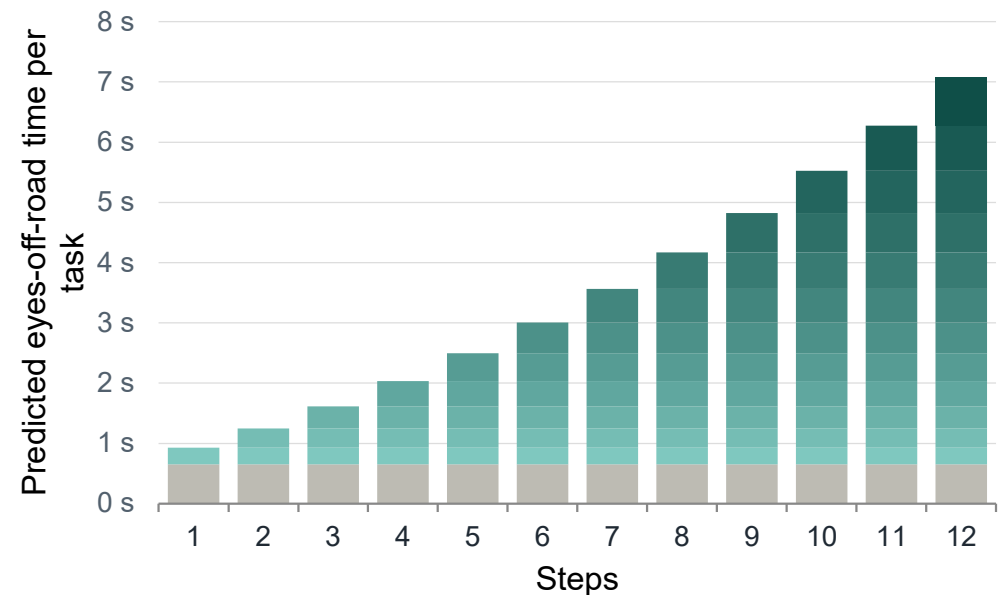


Fig 6. Predicted eyes-off-road time per task (s)

Generalisability constraints



Simplicity over best fit

Model 1 trades a minimal DRT accuracy (.71 vs .75) for stability and parsimony

Task type

Short, continuous tasks done as fast as possible; self-paced tasks untested

Fleet

12 systems tested in 2023; new modalities need re-validation on track

Tracking error

Supporting evidence only; its embedded DRT missed on-road demand

Counting steps predicts attentional cost



Validity

Step count is stable across contexts and predicts DRT, VOT and tracking criteria out of sample ($r = .65-.81$)

Interpretable

Each step costs about 1pp of DRT misses and 0.3-0.8 s of eyes-off-road time

Practical

A non-expert observer can score a system in a parked car in about 45-50 minutes, with a video camera or second observer and no other instrumentation

Application

The empirical basis for a consumer-facing IVIS rating

Design

Later steps cost more eyes-off-road time, so removing steps gives a compounding benefit

Thank you



Acknowledgements

Funded by the Australian Automobile Association Road Safety Research Program (Project DD2021-02). Findings do not necessarily reflect the views of the AAA or its member clubs.

ABMARC: AARC engineering, telemetry and safety drivers. Red Scientific: D. Strayer and J. Cooper. Protocol groundwork: M. Regan and P. Prabhakaran (UNSW).

Our thanks to the 20 participants.



**Australian
Automobile
Association**



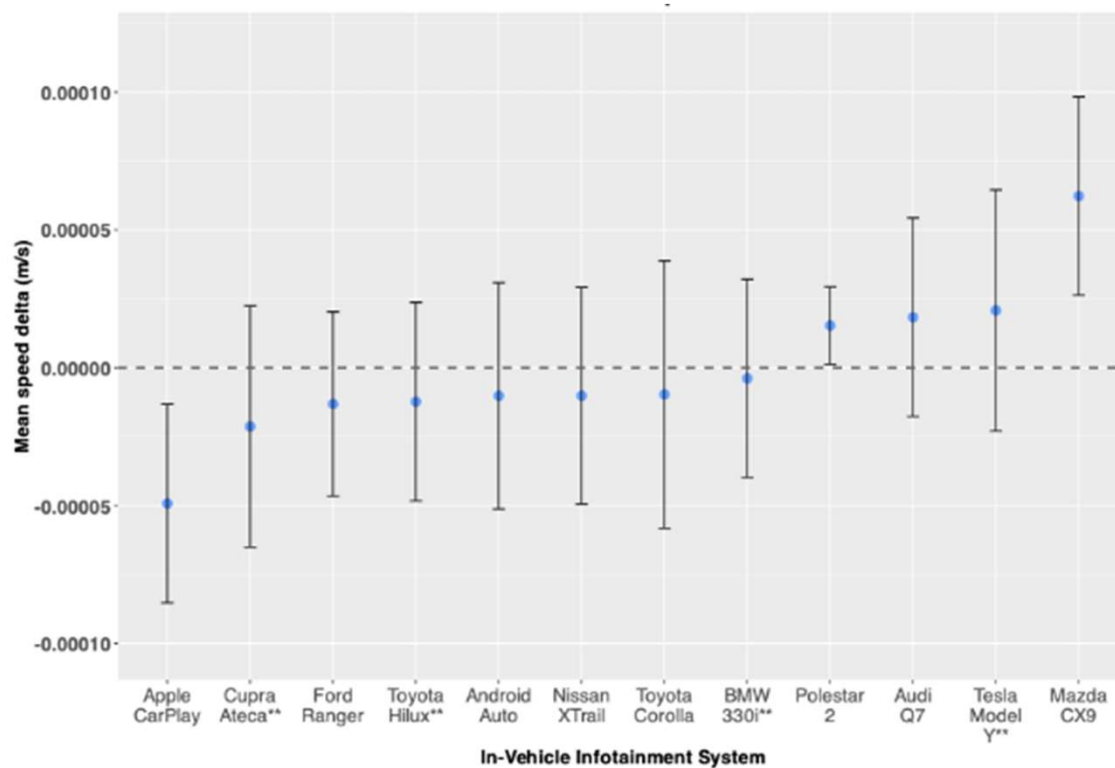
THE UNIVERSITY OF
NEWCASTLE
AUSTRALIA

Step-band rating of the 2023 fleet



Rating	Mean steps	System	VOT steps	DRT steps	Combined	Rating
★★★★★	1–2	Apple CarPlay	2.00	2.24	2.12	★★★★★
★★★★★	3–4	Polestar 2	3.00	2.95	2.97	★★★★★
★★★★★	5–6	Ford Ranger	3.50	3.22	3.36	★★★★★
★★★★★	7–8	Toyota Corolla	3.53	2.87	3.20	★★★★★
★★★★★	9–10	BMW 330i	3.79	3.95	3.87	★★★★★
★★★★★	≥ 11	Audi Q7	4.33	3.55	3.94	★★★★★
★★★★★		Android Auto	4.39	3.88	4.14	★★★★★
★★★★★		Toyota Hilux	4.99	4.98	4.98	★★★★★
★★★★★		Mazda CX-9	6.11	4.39	5.25	★★★★★
★★★★★		Tesla Model Y	6.36	4.60	5.48	★★★★★
★★★★★		Nissan X-Trail	6.27	6.41	6.34	★★★★★
★★★★★		Cupra Ateca*	–	10.72	10.72	★★★★★

Vehicle telemetry did not distinguish between systems



No meaningful system effects

Lateral acceleration and speed change varied little between systems

Interpretation

Consistent with the cognitive control hypothesis: drivers protect vehicle control while detection suffers

References



- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Getty, D., Biondi, F., Morgan, S. D., Cooper, J. M., & Strayer, D. L. (2018). The effects of voice system design components on driver workload. *Transportation Research Record*, 2672(35), 10–19. <https://doi.org/10.1177/0361198118777382>
- International Organization for Standardization. (2016). *Road vehicles – Transport information and control systems – Detection-response task (DRT) for assessing attentional effects of cognitive load in driving* (ISO 17488:2016). <https://www.iso.org/standard/59887.html>
- McDonnell, A. S., Imberger, K., Poulter, C., & Cooper, J. M. (2021). The power and sensitivity of four core driver workload measures for benchmarking the distraction potential of new driver vehicle interfaces. *Transportation Research Part F*, 83, 99–117. <https://doi.org/10.1016/j.trf.2021.09.019>
- Metz, B., Landau, A., & Just, M. (2014). Frequency of secondary tasks in driving – Results from naturalistic driving data. *Safety Science*, 68, 195–203. <https://doi.org/10.1016/j.ssci.2014.04.002>
- Strayer, D. L., Cooper, J. M., Goethe, R. M., McCarty, M. M., Getty, D., & Biondi, F. (2017). *Visual and cognitive demands of using in-vehicle infotainment systems*. AAA Foundation for Traffic Safety.
- Strayer, D. L., Cooper, J. M., McCarty, M. M., Getty, D. J., Wheatley, C. L., Motzkus, C. J., Goethe, R. M., Biondi, F., & Horrey, W. J. (2019). Visual and cognitive demands of CarPlay, Android Auto, and five native infotainment systems. *Human Factors*, 61(8), 1371–1386. <https://doi.org/10.1177/0018720819836575>
- Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In *Psychology of Learning and Motivation* (Vol. 54, pp. 29–58). Academic Press. <https://doi.org/10.1016/B978-0-12-385527-5.00002-4>
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449–455. <https://doi.org/10.1518/001872008X288394>