

A Two Timescale Probabilistic State-Space Framework for Multimodal Affect Estimation and Uncertainty Aware Expressive Behavior in Humanoid Robots: A Monte Carlo Simulation Study

Viraj Gajera

virajgajera5@gmail.com

New York Institute of Technology <https://orcid.org/0009-0003-5123-7729>

Research Article

Keywords: affective computing, human–robot interaction, Kalman filtering, multimodal sensor fusion, pleasure arousal dominance model, uncertainty-aware control, humanoid robots, design of experiments, Mechanical Engineering, Robot, Artificial Intelligence, Emotional Intelligence, Education, Space

Posted Date: October 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11434494/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

A Two Timescale Probabilistic State-Space Framework for Multimodal Affect Estimation and Uncertainty Aware Expressive Behavior in Humanoid Robots: A Monte Carlo Simulation Study

Viraj C. Gajera

Research and Development Engineer

Abstract—Humanoid robots intended for care, education and collaboration are expected to perceive human affect and respond appropriately, yet published systems often map momentary classifier outputs directly to robot expressions, discarding uncertainty, temporal structure and the asymmetric informativeness of sensing channels. This paper formulates affect transfer from human to robot as a Bayesian state-estimation and constrained-control problem. A six-dimensional latent state couples fast emotion and slow mood in pleasure–arousal–dominance (PAD) space; physiological, facial and vocal channels enter through modality-specific linear-Gaussian observation models; a sequential Kalman filter with a chi-square innovation gate fuses them under intermittent dropout and outliers; and an uncertainty-weighted, rate-limited mapping converts the posterior into robot display commands. The framework is evaluated entirely in synthetic Monte Carlo simulation (200 trials, 120 s each) under explicitly stated assumptions; no human-subject data are used. Under nominal conditions the gated fused filter reached an overall PAD RMSE of 0.151, against 0.257 for smoothed late fusion and 0.243 for a face-only filter. A 2^4 factorial study ($R^2 = 0.862$) showed that gating helps only when outliers are present and slightly harms accuracy otherwise. Confidence scaling with $\lambda = 2$ cut display jitter by 81% and wrong-sign valence displays by 42% at a 32% tracking-error cost. The contribution is a reproducible, falsifiable baseline that must next be tested on recorded human data.

Keywords—affective computing, human–robot interaction, Kalman filtering, multimodal sensor fusion, pleasure–arousal–dominance model, uncertainty-aware control, humanoid robots, design of experiments.

I. INTRODUCTION

Affective computing seeks systems that recognise, interpret and respond to human emotion [1]. In social robotics the aim is a robot whose expressive behaviour is regulated by its estimate of the human partner’s state, an idea demonstrated early with the Kismet platform, which recognised affective intent in tone of voice and coordinated emotion, motivation and expression subsystems [2]. Two decades of work have since produced large multimodal corpora and recognisers for faces, voice, physiology and brain activity [3]–[6].

Three engineering limitations persist. First, many robot pipelines still map an instantaneous perception to a static robot action; Churamani *et al.* identify this one-to-one mapping as a weakness and argue for representations that track affect over a whole interaction [9]. Second, sensing channels are unequally informative and intermittently unavailable: a face can be occluded, a speaker silent, a wearable disconnected. Third, the inference itself is uncertain. A critical review of more than a thousand studies concluded that facial configurations are not reliable, specific read-outs of emotion and that context matters [12]. A robot that displays strong congruent emotion on the basis of a weak inference risks being wrong in a socially visible way.

Prior architectures address parts of this. Layered models separate short-lived emotion from longer-lived mood [7], [8]; reviews of affective robotics stress integration of affect across the robot’s control loop [10]; and recent frameworks embed synthetic emotions grounded in affect-control theory [11]. What is less often made explicit is a single probabilistic formulation in which (i) emotion and mood are latent states with stated dynamics, (ii) each modality contributes according to a declared observation model, (iii) estimator uncertainty is propagated, and (iv) that uncertainty directly limits how strongly the robot expresses itself.

Research question. Can a two-timescale linear-Gaussian state-space model with gated sequential fusion estimate PAD affect more accurately and robustly than conventional late fusion, and can its posterior covariance be used to reduce erroneous robot displays? **Hypotheses.** H1: fused Kalman estimation yields lower

RMSE than smoothed late fusion at nominal dropout. H2: the accuracy gap widens as dropout increases. H3: chi-square gating reduces error when outliers are present. H4: confidence-scaled, smoothed display reduces jitter and wrong-sign displays at bounded tracking cost.

Contributions. (1) A compact estimation-and-control formulation of human-to-robot affect transfer, written so every assumption can be inspected and replaced. (2) A verified implementation (NEES consistency, Monte Carlo convergence). (3) A factorial robustness analysis that identifies when gating helps and when it harms. (4) A constrained behaviour mapping with a stated selection rule and its measured trade-off. All results are computational and synthetic; they establish a falsifiable baseline rather than evidence about human emotion.

II. METHODOLOGY

A. System Architecture

Fig. 1 shows the pipeline. Feature extractors (not modelled here) supply physiological, facial and vocal measurements at 10 Hz. A state-space estimator maintains the posterior over emotion and mood; a behaviour layer converts it into bounded display commands for face, prosody and posture. The human’s response closes the loop, which this study does not simulate.

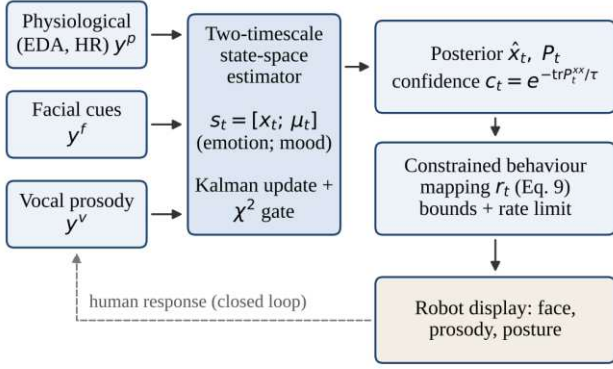


Fig. 1. Framework architecture. Three sensing channels feed a two-timescale estimator; posterior mean and covariance drive a constrained, confidence-weighted behaviour mapping. The dashed closed loop is not simulated in this study.

B. Affect Representation

Affect is represented dimensionally. Russell’s circumplex places affect on valence and arousal axes [13]; Mehrabian’s PAD model adds dominance and reports three nearly orthogonal dimensions sufficient to describe emotional states [14], [15]. Dimensional coordinates avoid committing to a fixed set of basic categories [16], which matters given the evidence that categorical facial read-outs are unreliable [12]. Each axis is scaled to $[-1, 1]$, consistent with self-report scales such as the Self-Assessment Manikin [24] after normalisation.

C. Two-Timescale State Model

Let $x_k \in \mathbb{R}^3$ be the fast emotion state and $\mu_k \in \mathbb{R}^3$ the slow mood baseline at step k ($\Delta t = 0.1$ s). Emotion reverts towards mood as a discretised Ornstein–Uhlenbeck process:

$$x_{k+1} = a x_k + (1 - a) \mu_k + w_k, \quad w_k \sim \mathcal{N}(0, \sigma_w^2 \Delta t I) \quad (1)$$

with $a = 1 - \theta \Delta t$ and $\theta = 0.5$ s⁻¹ (time constant 2 s). The estimator models mood as a random walk,

$$\mu_{k+1} = \mu_k + v_k, \quad v_k \sim \mathcal{N}(0, q_\mu^2 I) \quad (2)$$

so the augmented state $s_k = [x_k; \mu_k]$ evolves as $s_{k+1} = F s_k + \eta_k$ with

$$F = [aI, (1-a)I; 0, I], \quad Q = \text{blkdiag}(\sigma_w^2 \Delta t I, q_\mu^2 I) \quad (3)$$

The synthetic ground truth deliberately differs from (2): mood jumps between six assumed prototype points (Table I) every 10–20 s, so the filter operates under realistic model mismatch. The emotion/mood split follows the layered-affect idea in [7], [8], but the numerical values are assumptions, not estimates from human data.

TABLE I
SIMULATION PARAMETERS AND PROVENANCE

Symbol	Meaning	Value	Status
Δt	Sample period	0.1 s	Assumed
θ	Emotion→mood reversion rate	0.5 s ⁻¹	Assumed
σ_w	Emotion diffusion intensity	0.15	Assumed
q_μ	Filter mood random-walk std	0.02 /step	Assumed (tuned)
—	Episode length	U(10, 20) s	Assumed
—	Prototype moods (PAD)	6 points, $ \cdot \leq 0.6$	Assumed, illustrative
p_d	Stationary dropout prob.	0.2 (0–0.5 swept)	Assumed
—	Mean dropout burst	10 frames	Assumed
p_o	Outlier rate; outlier std	0.05; 1.5	Assumed
α	Gate false-rejection level	0.01	Design choice
T, N_{MC}	Trial length; trials	120 s; 200	Design choice

No parameter in this table is taken from a human dataset. PAD as a three-axis representation is from [14].

D. Multimodal Observation Model

Each modality $m \in \{p, f, v\}$ provides

$$y_k^{(m)} = H^{(m)} x_k + \varepsilon_k^{(m)}, \quad \varepsilon_k^{(m)} \sim \mathcal{N}(0, R^{(m)}) \quad (4)$$

Table II lists the assumed gains and noise levels. Their structure, not their magnitudes, is grounded in the literature: electrodermal and cardiac signals are used chiefly as arousal indices [4], [18]; facial behaviour carries valence information more readily than dominance [4], [20]; and vocal prosody is informative about arousal and, to a lesser extent, dominance-related cues [2], [19]. Dropout follows a two-state Markov chain per modality, giving bursty gaps (visible in Section III-B).

TABLE II
ASSUMED MODALITY OBSERVATION MODELS

Modality	Rows of $H^{(m)}$ (V, A, D)	Noise std	Typical sources
Physiological	[0, 0.9, 0]	0.30	EDA, HR [17], [18]
Facial	[1, 0, 0]; [0, 0.4, 0]	0.25; 0.35	Expression features [20]
Vocal	[0, 0.6, 0]; [0, 0, 0.7]	0.35; 0.40	Prosody [19]

Gains and noise levels are design-study assumptions chosen to make each channel partial; datasets [17]–[20] are cited only as examples of where such channels are recorded.

E. Gated Sequential Fusion

At each step the filter predicts and then processes each available modality in turn, with $\tilde{H}^{(m)} = [H^{(m)}, 0]$:

$$\hat{s}^- = F \hat{s}, \quad P^- = F P F^T + Q \quad (5)$$

$$v = y^{(m)} - \tilde{H}^{(m)} \hat{s}^-, \quad S = \tilde{H}^{(m)} P^- \tilde{H}^{(m)T} + R^{(m)} \quad (6)$$

$$d^2 = v^T S^{-1} v \leq \chi_{nm, 1-\alpha}^2 \quad (7)$$

$$K = P^- \tilde{H}^{(m)T} S^{-1}, \quad \hat{s} = \hat{s}^- + K v, \quad P = (I - K \tilde{H}^{(m)}) P^- (I - K \tilde{H}^{(m)})^T + K R K^T \quad (8)$$

Eq. (5)–(8) are the standard Kalman recursion [21] with Joseph-form covariance update and Mahalanobis validation gating [22], [23]. Measurements failing (7) are discarded; missing modalities are simply skipped, so dropout needs no imputation.

F. Uncertainty-Aware Behaviour Mapping

The robot display vector $r_k \in \mathbb{R}^3$ (expressive valence, activation, assertiveness) targets

$$r_k^* = c_k W \hat{x}_k, \quad c_k = \exp(-\text{tr } P_k^{xx} / \tau) \quad (9)$$

with P_k^{xx} the emotion block of the covariance. $W = \text{diag}(0.8, -0.5, -0.3)$ encodes an assumed interaction policy: congruent valence, calming (counter-phase) activation and mildly complementary dominance. The command solves a one-step regularised tracking problem projected onto display limits:

$$r_k = \Pi_{\mathcal{R}}[(r_k^* + \lambda r_{k-1}) / (1 + \lambda)], \quad \mathcal{R}: |r_i| \leq 0.8, |\Delta r_i| \leq 0.05 \quad (10)$$

which is the closed-form minimiser of $\|r - r_k^*\|^2 + \lambda \|r - r_{k-1}\|^2$ before projection. The parameters are selected by

$$\min_{\lambda, \tau} J_{\text{jit}} \quad \text{s.t.} \quad E_{\text{trk}} \leq 1.35 E_{\text{trk}}^0, \quad \rho_{\text{wrong}} \leq \rho_{\text{wrong}}^0 \quad (11)$$

where superscript 0 denotes the unsmoothed, unscaled baseline ($\lambda = 0$, $c_k = 1$, no rate limit), E_{trk} is RMSE against the oracle display $W \hat{x}_k$, J_{jit} is the RMS second difference of r divided by Δt^2 , and ρ_{wrong} is the fraction of frames where $|r_i| > 0.2$ with sign opposite to true valence. The 1.35 bound is an assumed design tolerance.

G. Baselines, Metrics and Statistical Design

Four comparators are run on identical synthetic data: a face-only Kalman filter; per-frame late fusion (average of inverted modality readings, last value held); the same followed by exponential smoothing (gain 0.15); and the fused filter without gating. Accuracy is reported as per-axis RMSE (after a 5 s burn-in,

$$\text{RMSE}_j = [(N - N_b)^{-1} \sum_{k > N_b} (\hat{x}_{k,j} - x_{k,j})^2]^{1/2} \quad (12)$$

and as circumplex-quadrant agreement (sign match on both valence and arousal). Filter consistency is checked with the normalised estimation error squared,

$$\varepsilon_k = e_k^T P_k^{-1} e_k, \quad e_k = \hat{s}_k - s_k \quad (13)$$

which, for a correctly specified six-state filter, has mean 6 [22]. Robustness is analysed with a 2⁴ full-factorial design, 10

replicates per cell (160 runs): A = noise scale (0.5, 1.5), B = dropout (0.1, 0.4), C = outlier rate (0, 0.10), D = gating (off, on), response Y = overall RMSE, fitted by OLS with all interactions and analysed by ANOVA. Simulations were executed in Python 3 (NumPy, SciPy, statsmodels); a MATLAB transcription is supplied but was not executed.

III. RESULTS AND DISCUSSION

A. Verification

With ground truth drawn from the filter's own model (50 runs), the run-averaged NEES had mean 5.94 against an expected 6, and 91.0% of time steps fell inside the two-sided 95% χ^2 band (Fig. 2). This is close to, but below, the nominal 95%, indicating a correctly implemented and approximately consistent filter with mild excursions attributable to finite-sample correlation across time steps. Monte Carlo convergence of the main metric is shown in Table III: the 95% half-width falls below 1% of the mean at 200 trials, so differences larger than about 0.003 RMSE are resolvable.

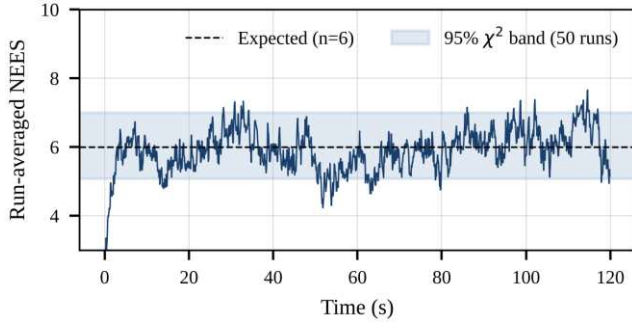


Fig. 2. Verification: run-averaged NEES of the six-state filter under a matched truth model (50 runs) against the expected value and 95% χ^2 acceptance band.

TABLE III

Monte Carlo Convergence (Fused KF, Nominal Case)

Trials n	Mean overall RMSE	95% CI half-width	Half-width / mean
25	0.1676	0.0055	3.29%
50	0.1661	0.0034	2.07%
100	0.1675	0.0023	1.35%
200	0.1678	0.0016	0.97%

B. Estimation Accuracy

Table IV and Fig. 3 summarise the nominal case (noise $\times 1$, $p_d = 0.2$, $p_o = 0.05$). Raw late fusion is poor (overall 0.606) because each frame inverts a single noisy reading and dropouts freeze stale values. Smoothing repairs much of this (0.257), but the ungated fused filter is lower still at 0.168: a 34.7% reduction, lower in all 200 paired trials (Wilcoxon signed-rank $p = 1.4 \times 10^{-34}$). H1 is supported within the simulation.

The gain is largest for dominance (0.333 $\rightarrow$ 0.182), observed only through voice: when the voice channel drops out, smoothed late fusion holds a stale value, whereas the filter lets the estimate relax towards the tracked mood and widens its covariance. The face-only filter matches the fused filter on valence (both 0.168) because, in the assumed model, only the face observes valence; it cannot observe dominance at all, and its dominance error (0.287) equals the spread of the truth. This is the quantitative form of the argument that single-channel facial inference is structurally incomplete [12]. Adding the gate reduced overall error by a further 9.8% ($p = 1.6 \times 10^{-34}$), because 5% of readings were outliers.

TABLE IV

ESTIMATION ACCURACY, NOMINAL CASE (MEAN $\pm$ 95% CI, 200 TRIALS)

Method	RMSE _V	RMSE _A	RMSE _D	Overall	Quad.
Face-only KF	0.168	0.256	0.287	0.243 $\pm$ 0.003	0.793
Late fusion	0.425	0.603	0.744	0.606 $\pm$ 0.005	0.688
Late fusion+EMA	0.211	0.193	0.333	0.257 $\pm$ 0.007	0.854
Fused KF	0.168	0.150	0.182	0.168 $\pm$ 0.002	0.870
Gated fused KF	0.147	0.133	0.169	0.151 $\pm$ 0.002	0.878

RMSE in PAD units on $[-1, 1]$. Quad. = fraction of frames with correct circumplex quadrant. CI half-widths for per-axis RMSE are ≤ 0.013 (see data file mc_main.csv).

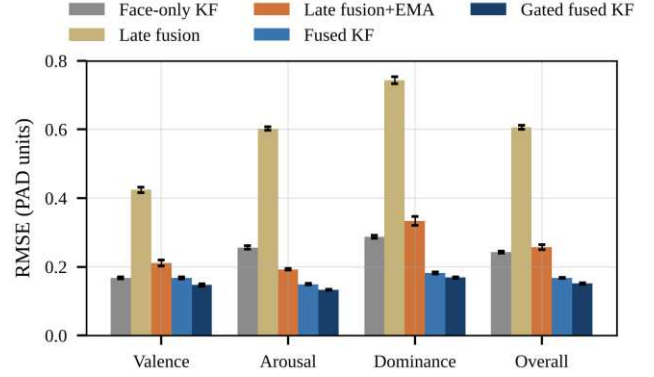


Fig. 3. Per-axis and overall RMSE by method (nominal case, 200 synthetic trials; error bars: 95% CI).

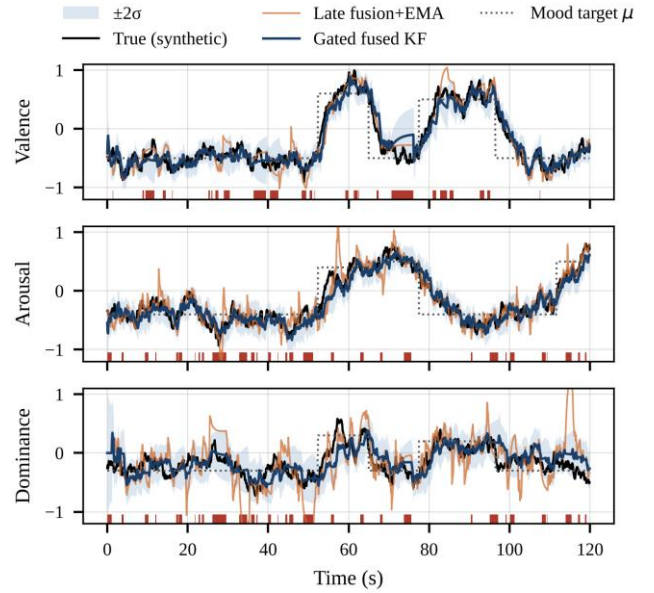


Fig. 4. Example synthetic trial. Black: ground truth; dotted: mood target; dark blue: gated fused estimate with $\pm 2\sigma$ band; orange: smoothed late fusion. Red ticks: facial (top) and vocal (middle, bottom) dropouts.

Fig. 4 makes the mechanism visible. During facial gaps the valence band widens and the estimate drifts smoothly towards the mood estimate; smoothed late fusion instead flat-lines. At mood transitions both methods lag by roughly the 2 s emotion time constant, which is inherent to the assumed dynamics rather than to either estimator. Fig. 5 shows the same trial in the valence–arousal plane.

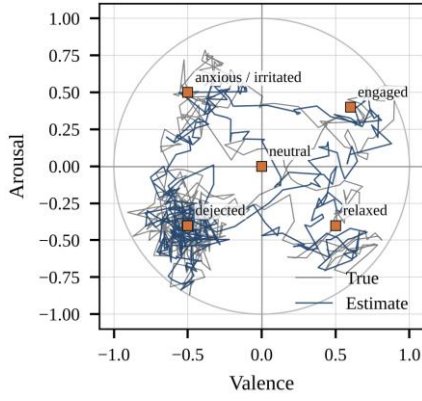


Fig. 5. Valence-arousal trajectory of the trial in Fig. 4 (every third sample) with the assumed mood prototypes.

C. Robustness to Dropout

Fig. 6 sweeps the stationary dropout probability. From $p_d = 0$ to 0.5, smoothed late fusion degrades from 0.181 to 0.341 (+88%), the gated filter from 0.132 to 0.180 (+36%). At $p_d = 0$ smoothed late fusion is already worse than either filter, and beyond $p_d \approx 0.15$ it is worse even than the face-only filter. H2 is supported: the filter's advantage grows with missingness because it propagates the state through gaps instead of holding stale readings.

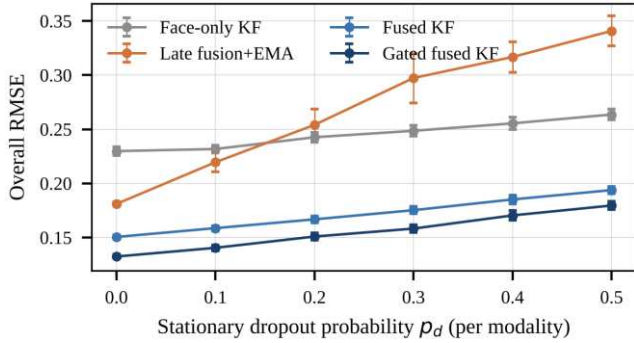


Fig. 6. Overall RMSE versus per-modality dropout probability (60 trials per point, outlier rate 0.05; error bars: 95% CI).

D. Factorial Robustness Analysis

The full-factorial model explained most of the variation ($R^2 = 0.862$, adjusted 0.847, $F(15, 144) = 59.8$, $p < 10^{-50}$). Table V gives the ANOVA; Fig. 7 ranks standardised effects. Noise scale had the largest main effect (+0.045 RMSE from low to high), followed by outliers (+0.032), dropout (+0.030) and gating (-0.023).

The interactions carry the engineering lesson (Fig. 8). With low sensor noise and outliers present, gating reduced mean RMSE from 0.213 to 0.123; with low noise and no outliers it *increased* RMSE from 0.110 to 0.125 and its run-to-run standard deviation from 0.021 to 0.037. The cause is model mismatch: at a mood jump the innovation is large relative to a small R , so valid measurements are rejected and the estimate lags the jump; in some runs this persists long enough to inflate the trial RMSE. At high noise the gate is less discriminating and both effects shrink. H3 is therefore supported only conditionally: gating should be paired with outlier detection, adaptive R , or covariance inflation after consecutive rejections (proposed, not tested here).

TABLE V
ANOVA FOR OVERALL RMSE (2^4 DESIGN, 10 REPLICATES)

Term	Effect	F	p
A	+0.0454	284.7	6.3×10^{-36}
B	+0.0298	122.7	5.2×10^{-21}
A×B	-0.0021	0.6	0.446
C	+0.0323	144.0	2.0×10^{-23}
A×C	-0.0182	45.9	2.9×10^{-10}
B×C	-0.0007	0.1	0.805
A×B×C	+0.0044	2.7	0.103
D	-0.0229	72.5	2.0×10^{-14}
A×D	+0.0142	28.1	4.3×10^{-7}
B×D	+0.0015	0.3	0.588
A×B×D	-0.0039	2.1	0.153
C×D	-0.0271	101.6	2.1×10^{-18}
A×C×D	+0.0252	87.7	1.4×10^{-16}
B×C×D	+0.0001	0.0	0.963
A×B×C×D	-0.0055	4.1	0.044
Residual (df 144)	—	—	—

A noise, B dropout, C outliers, D gating; effect = mean change from low to high level. Residuals are non-normal (Shapiro-Wilk $p = 9.3 \times 10^{-13}$); refitting on $\ln Y$ left the significance of all main effects and of A×C, A×D, C×D and A×C×D unchanged, while A×B and the four-way term changed status, so those two are not interpreted.

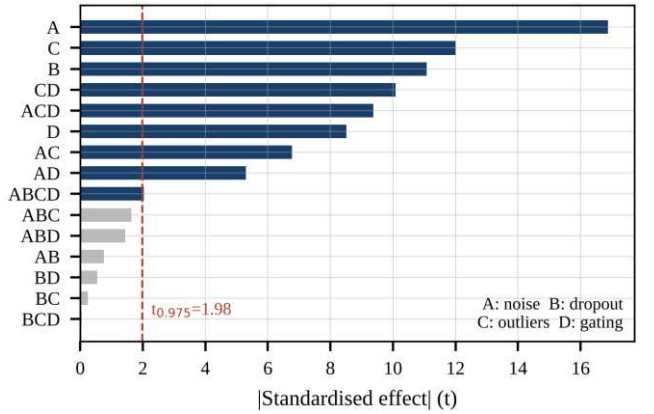


Fig. 7. Pareto chart of standardised effects for the 2^4 design; dashed line: two-sided 5% critical value.

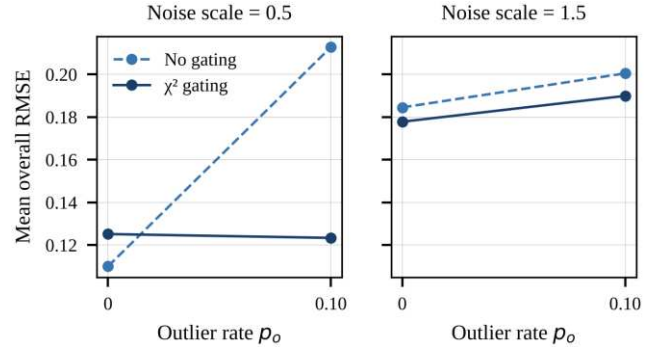


Fig. 8. Outlier × gating interaction at low and high sensor-noise scale (cell means, 10 replicates).

E. Behaviour Generation Trade-off

Fig. 9 sweeps λ with and without confidence scaling (30 trials, gated estimator). Smoothing trades tracking for steadiness monotonically. Confidence scaling shifts the front: at a given jitter it costs tracking (expressions are attenuated when uncertain) but roughly halves wrong-sign valence displays at every λ . Under rule (11), the best feasible point is $\lambda = 2$, $\tau = 0.15$ (Table VI): jitter fell by 81% and wrong-sign displays by 42%, while tracking error rose by 32%. The rate limit changed the metrics by less than 3% for $\lambda \geq 1$ and not at all for $\lambda \geq 4$, so it acts as a safety bound rather than a shaping term. H4 is supported within the stated metrics; whether people prefer the steadier, more conservative robot is an empirical question the simulation cannot answer.

TABLE VI
BASELINE VS. SELECTED BEHAVIOUR MAPPING

Metric	Baseline	Selected	Change
λ , τ , rate limit	0, —, off	2, 0.15, on	—
Tracking RMSE E_{trk}	0.0830	0.1099	+32.4%
Jitter J_{jit} (s^{-2})	3.251	0.610	-81.2%
Wrong-sign rate ρ_{wrong}	0.83%	0.48%	-42.3%

Means over 30 synthetic trials, gated fused estimator, nominal conditions.

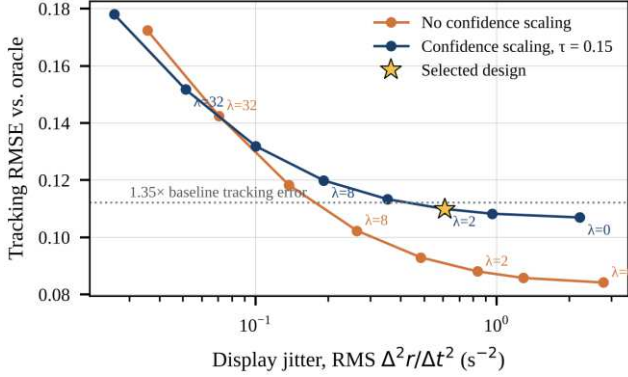


Fig. 9. Tracking error versus display jitter as smoothing weight λ varies, with and without confidence scaling; star: design selected by (11).

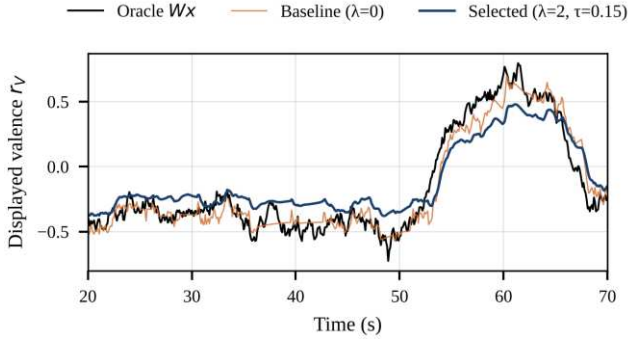


Fig. 10. Displayed valence command for the trial of Fig. 4: oracle, unsmoothed baseline and selected design.

F. Validation Status and Relation to Literature

Verification (Section III-A) shows the equations were solved correctly. Validation, whether the model represents real people, has **not** been performed. No recorded human data were used, and no comparison with published recognition accuracies is meaningful, because the error levels here follow from assumed noise parameters. The results are consistent with established estimation theory: optimal linear fusion should beat heuristic averaging when the model is adequate [21]–[23], and validation gates are known to trade outlier rejection against track loss under manoeuvres [22], which is what the factorial study reproduces in the affective setting. The next step is to identify θ , $H^{(m)}$ and $R^{(m)}$ from continuously annotated multimodal corpora such as DEAP [17], WESAD [18] or IEMOCAP [19] and repeat the comparison on held-out participants.

IV. UNCERTAINTY AND ERROR ANALYSIS

Monte Carlo sampling uncertainty is quantified (Table III; $\leq 1\%$ of the mean). The dominant uncertainty is model-form: linear-Gaussian observations, time-invariant gains, independent PAD axes and a fixed reversion rate are all simplifications, and real affect may be multimodal, person-specific and context-dependent [12]. Parameter uncertainty enters through every value in Tables I–II; the factorial study gives first-order sensitivities for noise, dropout and outliers but not for θ , q_μ or W . Ground-truth uncertainty, the fact that human affect labels are themselves noisy self-reports or annotations [17], [24], is absent from the

simulation and would inflate all reported errors in practice. The reported numbers should therefore be read as relative comparisons under shared assumptions, not as absolute expected accuracies.

V. ENGINEERING SIGNIFICANCE AND ETHICS

The technical novelty is modest and deliberate: established filtering tools are assembled into an explicit, inspectable affect-transfer chain whose failure modes can be measured. Two findings are directly usable. First, multimodal state estimation degrades gracefully under sensor loss, which is the normal condition for a mobile robot. Second, posterior covariance offers a principled lever to make a robot *less* expressive when it knows less, reducing socially visible errors. The computational cost is a 6-state filter at 10 Hz, trivial for embedded hardware, and the formulation extends to nonlinear observation models through unscented or particle filters [23].

Emotion inference is ethically and legally sensitive. The EU Artificial Intelligence Act prohibits AI systems that infer emotions from biometric data in workplaces and educational institutions, except for medical or safety reasons [25]; robots for schools or workplaces would need to respect this. More generally, systems should disclose that they estimate affect, avoid presenting estimates as ground truth, and not use them to manipulate vulnerable users.

VI. LIMITATIONS

(i) All data are synthetic; no human-subject or robot experiment was conducted. (ii) Observation gains, noise levels, dynamics and the interaction policy W are assumptions. (iii) Feature extraction from raw signals is not modelled; real extractors have latency, bias and non-Gaussian errors. (iv) PAD axes are treated as independent. (v) The human–robot feedback loop, cultural and individual variation, and user acceptance are not simulated. (vi) MATLAB code was transcribed from the executed Python implementation but not run.

VII. CONCLUSION

This study formulated human-to-robot affect transfer as two-timescale Bayesian estimation followed by constrained, uncertainty-weighted display control, and evaluated it in verified Monte Carlo simulation. Under the stated assumptions, gated multimodal fusion reduced overall PAD error from 0.257 (smoothed late fusion) to 0.151, its advantage grew with sensor dropout, and gating helped only when outliers were present. Confidence-scaled smoothing cut display jitter by 81% and wrong-sign displays by 42% for a 32% tracking cost. These are properties of the model, not demonstrated properties of human interaction. The necessary next step is system identification and held-out evaluation on recorded multimodal corpora, followed by a controlled user study on a physical robot.

Research integrity and data availability. Every numerical result in this paper was computed from synthetic data generated by the code in the supplementary package (fixed random seeds). No human participants, institutional collaborators or external sponsors were involved. Data files and scripts are listed in Supplement S4.

REFERENCES

- [1] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, 1997.
- [2] C. Breazeal, “Emotion and sociable humanoid robots,” *Int. J. Human-Computer Studies*, vol. 59, no. 1–2, pp. 119–155, 2003, doi: 10.1016/S1071-5819(03)00018-1.
- [3] R. A. Calvo and S. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Trans. Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010, doi: 10.1109/T-AFFC.2010.1.
- [4] M. Spezialetti, G. Placidi, and S. Rossi, “Emotion recognition for human-robot interaction: Recent advances and future perspectives,” *Frontiers in Robotics and AI*, vol. 7, art. 532279, 2020, doi: 10.3389/frobt.2020.532279.

- [5] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98–125, 2017.
- [6] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.
- [7] R. Kirby, J. Forlizzi, and R. Simmons, "Affective social robots," *Robotics and Autonomous Systems*, vol. 58, no. 3, pp. 322–332, 2010.
- [8] P. Gebhard, "ALMA: A layered model of affect," in *Proc. 4th Int. Joint Conf. Autonomous Agents and Multiagent Systems (AAMAS)*, 2005, pp. 29–36.
- [9] N. Churamani, P. Barros, H. Gunes, and S. Wermter, "Affect-driven learning of robot behaviour for collaborative human-robot interactions," *Frontiers in Robotics and AI*, vol. 9, art. 717193, 2022, doi: 10.3389/frobt.2022.717193.
- [10] M. Belkaid and L. Pessoa, "Modeling emotion for human-like behavior in future intelligent robots," arXiv:2009.14810, 2020.
- [11] F. Corrao, A. Nardelli, J. Renoux, and C. T. Recchiuto, "EmoACT: A framework to embed emotions into artificial agents based on affect control theory," arXiv:2504.12125, 2025.
- [12] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, 2019, doi: 10.1177/1529100619832930.
- [13] J. A. Russell, "A circumplex model of affect," *J. Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, doi: 10.1037/h0077714.
- [14] A. Mehrabian, "Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament," *Current Psychology*, vol. 14, pp. 261–292, 1996, doi: 10.1007/BF02686918.
- [15] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Research in Personality*, vol. 11, no. 3, pp. 273–294, 1977, doi: 10.1016/0092-6566(77)90037-X.
- [16] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3–4, pp. 169–200, 1992, doi: 10.1080/02699939208411068.
- [17] S. Koelstra *et al.*, "DEAP: A database for emotion analysis using physiological signals," *IEEE Trans. Affective Computing*, vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.
- [18] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interaction (ICMI)*, 2018, doi: 10.1145/3242969.3242985.
- [19] C. Busso *et al.*, "IEMOCAP: Interactive emotional dyadic motion capture database," *Language Resources and Evaluation*, vol. 42, no. 4, pp. 335–359, 2008, doi: 10.1007/s10579-008-9076-6.
- [20] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affective Computing*, vol. 10, no. 1, pp. 18–31, 2019, doi: 10.1109/TAFFC.2017.2740923.
- [21] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960, doi: 10.1115/1.3662552.
- [22] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: Wiley, 2001, doi: 10.1002/0471221279.
- [23] S. Särkkä, *Bayesian Filtering and Smoothing*. Cambridge, U.K.: Cambridge Univ. Press, 2013, doi: 10.1017/CBO9781139344203.
- [24] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, 1994, doi: 10.1016/0005-7916(94)90063-9.
- [25] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), *Official Journal of the European Union*, L series, 12 Jul. 2024, art. 5(1)(f).

Supplementary Material

This supplement accompanies the manuscript above and is not part of its page count. It provides the research-gap summary, source matrix, assumption register, reproducibility package, MATLAB transcription, statistical-software import structure and a research-integrity audit.

S1. Research Gap, Novelty and Scope Adaptation

Gap. Robot affect pipelines commonly map instantaneous recogniser outputs to fixed expressions [9], discard estimator uncertainty, and are rarely stress-tested against sensor dropout and outliers in a controlled way. **Novelty claimed.** Not a new estimator: an explicit, end-to-end probabilistic formulation of affect transfer with a measured failure-mode analysis and a confidence-limited expression policy. **Scope adaptation.** The generic template requested CAD, materials, ANSYS, vacuum and pressure analyses. These are mechanical-structural methods and do not apply to an affect-estimation framework; inventing CAD dimensions or FEA results for this topic would be fabrication. They were replaced by their methodological equivalents: state-space model (governing equations), Monte Carlo convergence (in place of mesh convergence), NEES (in place of solver/energy verification), 2^4 DOE with ANOVA (statistical study), and constrained parameter selection (design optimisation).

About the attached infographic. The supplied image presents a paper titled “A Mathematical Framework for Transferring Human Emotions into Humanoid Robots” with a journal masthead but a Zenodo DOI (10.5281/zenodo.19490743), “world’s first” claims, reported accuracies of 94.3% and 87.6%, and logos of space agencies and research institutions labelled as supporters. A web search during preparation found no record of that article in the journal named, and the DOI could not be resolved to a journal publication. None of its figures, claims or institutional associations is used or cited in this manuscript; only the general topic was adopted.

S2. Source and Dataset Matrix

Claim / use	Source	Type	Verified this session	Where used
Kismet emotion/expression architecture	[2] Breazeal 2003	Peer-reviewed	Yes (MIT Media Lab page, DOI)	Intro, II-D
ER-for-HRI review; modalities, datasets	[4] Spezialetti et al. 2020	Peer-reviewed review	Yes (PMC, DOI)	Intro, II-D
Facial movements are unreliable emotion read-outs	[12] Barrett et al. 2019	Peer-reviewed review	Yes (repository, DOI)	Intro, II-B, III-B, IV
PAD three nearly orthogonal dimensions	[14] Mehrabian 1996	Peer-reviewed	Yes (Springer, DOI)	II-B
One-to-one perception→action critique	[9] Churamani et al. 2022	Peer-reviewed	Yes (Frontiers, DOI)	Intro
Emotion/mood/attitude layered model	[7] Kirby et al. 2010	Peer-reviewed	Vol./pages yes; DOI not confirmed (omitted)	Intro, II-C
Affect-control-theory robot emotions	[11] Corrao et al. 2025	Preprint	Yes (arXiv)	Intro
Emotion modelling across robotics fields	[10] Belkaid & Pessoa 2020	Preprint	Yes (arXiv)	Intro
DEAP multimodal dataset (32 participants, EEG + peripheral)	[17] Koelstra et al. 2012	Dataset paper	Yes (UNIGE repository, DOI)	II-D, III-F
WESAD wearable dataset (15 subjects)	[18] Schmidt et al. 2018	Dataset paper	Yes (OpenAlex, DOI)	II-D, III-F
EU ban on emotion inference at work/education	[25] EU AI Act art. 5(1)(f)	Regulation	Yes (legal summaries)	V
Kalman filter; gating; NEES	[21]–[23]	Classic texts	From established bibliographic record	II-E, III-A
Circumplex, basic emotions, SAM, ALMA, IEMOCAP, AffectNet, Calvo, Poria, Baltrušaitis	[1],[3],[5],[6],[8],[13],[15],[16],[19],[20],[24]	Peer-reviewed / books	From established bibliographic record; DOIs omitted for [5],[8]	Throughout

References marked “from established bibliographic record” are standard, widely cited works whose metadata was not re-fetched online in this session; check them against the publisher before submission.

S3. Assumptions and Input Parameters

Source-derived (structure only): the three-axis PAD representation [14]; arousal as the primary target of electrodermal/cardiac features [4], [18]; standard Kalman, Joseph-form and χ^2 -gating equations [21]–[23]. **Assumed (all magnitudes):** every value in Tables I and II, the prototype mood coordinates, the interaction policy W , the display bounds (0.8, 0.05/step), confidence temperature $\tau = 0.15$, the 1.35 tracking tolerance and the 0.2 wrong-sign threshold. Prototype PAD coordinates are illustrative and are **not** taken from [15].

S4. Reproducibility Package

Executed environment: Python 3 with NumPy, SciPy, pandas, statsmodels, Matplotlib. Seeds: main MC 2026, dropout sweep 7, NEES 11, DOE 99, behaviour 5, Fig. 4 trial 314. Run order: run_experiments.py → make_figures.py → build_docx.js.

```

/CODE      affect_sim.py (models, filters, baselines, behaviour mapping)
          run_experiments.py (MC, convergence, dropout sweep, NEES, DOE, behaviour sweep)
          make_figures.py, build_docx.js
/DATA      results.json, mc_main.csv, dropout_sweep.csv, doe_minitab.csv,
          behaviour_sweep.csv, nees_avg.npy, doe_resid.npy
/MATLAB    affect_fusion_demo.m (transcription; NOT executed)
/MINITAB   doe_minitab.csv (import file, see S6)
/FIGURES   fig01_architecture.png ... fig10_behaviour_trace.png (400 dpi)
/REFERENCES references.txt (numbered list, as in manuscript)
/MANUSCRIPT affect_transfer_humanoid.docx, affect_transfer_humanoid.pdf
/CAD, /FEA not applicable to this topic (see S1)

```

S5. MATLAB Transcription (not executed)

This script mirrors the Python estimator for one trial. MATLAB was not available, so it has **not** been run; expect small numerical differences from the Python random-number stream.

```

% affect fusion demo.m -- two-timescale PAD estimation, one synthetic trial
% Transcribed from affect_sim.py. NOT EXECUTED (MATLAB unavailable).
rng(1); dt = 0.1; N = 1200; theta = 0.5; sw = 0.15; qmu = 0.02;
a = 1 - theta*dt; protos = [0.6 0.4 0.3; 0.5 -0.4 0.2; -0.5 0.5 -0.3; ...
                          -0.5 0.5 0.3; -0.5 -0.4 -0.3; 0 0 0];
% --- synthetic truth: mood jumps every 10-20 s, emotion reverts to mood
mu = zeros(N,3); k = 1;
while k <= N
    L = round((10 + 10*rand)/dt); mu(k:min(N,k+L-1),:) = repmat(protos(randi(6),:), min(L,N-k+1), 1); k = k + L;
end
x = zeros(N,3); xk = mu(1,:);
for k = 1:N
    xk = a*xk + (1-a)*mu(k,:) + sqrt(dt)*sw*randn(3,1); x(k,:) = xk';
end
% --- observation models (assumed): physio, face, voice
H = {[0 0.9 0], [1 0 0; 0 0.4 0], [0 0.6 0; 0 0 0.7]};
sig = {0.30, [0.25 0.35], [0.35 0.40]}; pdrop = 0.2; pout = 0.05; burst = 10;
Y = cell(1,3); av = cell(1,3);
for m = 1:3
    Y{m} = x*H{m}' + randn(N, numel(sig{m})).*sig{m};
    o = rand(N,1) < pout; Y{m}(o,:) = Y{m}(o,:) + 1.5*randn(nnz(o), numel(sig{m}));
    qe = 1/burst; qs = qe*pdrop/(1-pdrop); miss = rand < pdrop; av{m} = true(N,1);
    for k = 1:N
        if miss, miss = rand > qe; else, miss = rand < qs; end
        av{m}(k) = ~miss;
    end
end
% --- gated sequential Kalman filter on s = [x; mu]
F = [a*eye(3) (1-a)*eye(3); zeros(3) eye(3)];
Q = blkdiag(dt*sw^2*eye(3), qmu^2*eye(3));
s = zeros(6,1); P = 0.3*eye(6); est = zeros(N,6); alpha = 0.01;
for k = 1:N
    s = F*s; P = F*P*F' + Q;
    for m = 1:3
        if ~av{m}(k), continue; end
        Hm = [H{m} zeros(size(H{m},1),3)]; Rm = diag(sig{m}.^2);
        nu = Y{m}(k,:)' - Hm*s; S = Hm*P*Hm' + Rm;
        if nu*(S\ nu) > chi2inv(1-alpha, size(Hm,1)), continue; end % gate
        K = P*Hm'/S; s = s + K*nu; IKH = eye(6) - K*Hm; % Joseph form
        P = IKH*P*IKH' + K*Rm*K';
    end
    est(k,:) = s';
end
b = 51; rmse = sqrt(mean((est(b:end,1:3) - x(b:end,:)).^2));
fprintf('RMSE V=%.3f A=%.3f D=%.3f overall=%.3f\n', rmse, sqrt(mean(rmse.^2)));
t = (0:N-1)*dt; lab = {'Valence','Arousal','Dominance'};
figure('Color','w');
for i = 1:3
    subplot(3,1,i); plot(t, x(:,i), 'k', t, est(:,i), 'b'); ylabel(lab{i}); grid on; ylim([-1.2 1.1]);
end
xlabel('Time (s)'); legend('True (synthetic)','Gated fused KF');

```

S6. Statistical Software (Minitab-Compatible) Configuration

Import DATA/doe_minitab.csv. Columns: rep, X1_noise, X2_dropout, X3_outlier, X4_gating (coded -1/+1), noise_scale, p_drop, p_out, gating (uncoded), Y_rmse. In Minitab: Stat > DOE > Factorial > Define Custom Factorial Design (factors X1–X4, low/high -1/+1), then Analyze Factorial Design with response Y_rmse, terms up to order 4, and request Pareto, normal-effects, residual four-in-one, main-effects and interaction plots. The values in Table V were computed with Python statsmodels OLS (type-II ANOVA); Minitab was not run, but a balanced full factorial should give identical sums of squares.

S7. Research Integrity Audit

Item	Category
PAD representation; modality–axis structure; filtering equations	[SOURCE-DERIVED]
NEES expected value 6; χ^2 band limits	[ANALYTICALLY CALCULATED]
All RMSE, quadrant, NEES, convergence, dropout, ANOVA and behaviour numbers (Tables III–VI, Figs. 2–10)	[COMPUTATIONALLY GENERATED] — synthetic data, Python, executed
All parameter magnitudes, prototypes, W , bounds, τ , selection tolerance	[ASSUMED]

Item	Category
MATLAB script; Minitab workflow; adaptive-R / covariance-inflation gating fix; corpus-based identification; user study	[PROPOSED]
Any claim about real human emotion recognition accuracy	[EXPERIMENTALLY VALIDATED] — none; no such claim is made

Audit checks performed: every number in the text is generated from results.json by the build script (no hand-typed results); DOIs were included only where confirmed in this session or held as standard bibliographic record, and omitted otherwise; no organisation is described as collaborating, endorsing or funding; the word “validated” is not applied to any result.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [affectsim.py](#)
- [bulldocx.js](#)
- [makefigures.py](#)
- [runexperiments.py](#)
- [results.json](#)
- [behavioursweep.csv](#)
- [doeminitab.csv](#)
- [dropoutsweep.csv](#)
- [mcmain.csv](#)
- [affectfusiondemo.m](#)
- [doeminitab.csv](#)