

Supplementary Material for Behavior-Supervised Alignment for Cross-Subject Neural Decoding Across Non-Corresponding Electrode Grids

Ward Ceyssens^a, Axel Faes^{b,*}

^a*Data Science Institute (DSI), Hasselt University, Hasselt, Belgium*

^b*Biomedical Data Sciences, Data Science Institute (DSI) and Biomedical Research Institute (BIOMED), Hasselt University, Hasselt, Belgium*

S1. Implementation and reproducibility

The alignment for finger f on subject i is built from the training split alone. Writing $H_i \in \mathbb{R}^{T \times C_i}$ for the z-scored 70–150 Hz high-gamma envelope and $y_{i,f}$ for the z-scored flexion of finger f , the rank-1 filter is the normalized channel-behavior correlation

$$w_{i,f}^{(1)} = \frac{H_i^\top y_{i,f}}{\|H_i^\top y_{i,f}\|_2}, \quad (\text{S1})$$

and the remaining $R - 1$ filters are the leading left singular vectors of the correlation-reweighted high-gamma covariance $\tilde{H}_i^\top \tilde{H}_i$, where $\tilde{H}_i = H_i \text{diag}(|w_{i,f}^{(1)}|)$ restricts the covariance to the finger-driven electrodes, each sign-aligned to $w_{i,f}^{(1)}$. Stacking the $R = 4$ filters for all five fingers gives $W_i \in \mathbb{R}^{D \times C_i}$ with $D = 20$. Table S1 lists the remaining settings. The alignment, source pool, and decoder retrain end to end on a single GPU.

S2. Method detail

Figure S1 unpacks three parts of the pipeline that the main-text flowchart states only in outline: what a behavior-aligned filter looks like on the cortex,

*Corresponding author.

Email address: `axel.faes@uhasselt.be` (Axel Faes)

how one shared axis behaves in two subjects whose grids do not correspond, and how the decoder is wired.

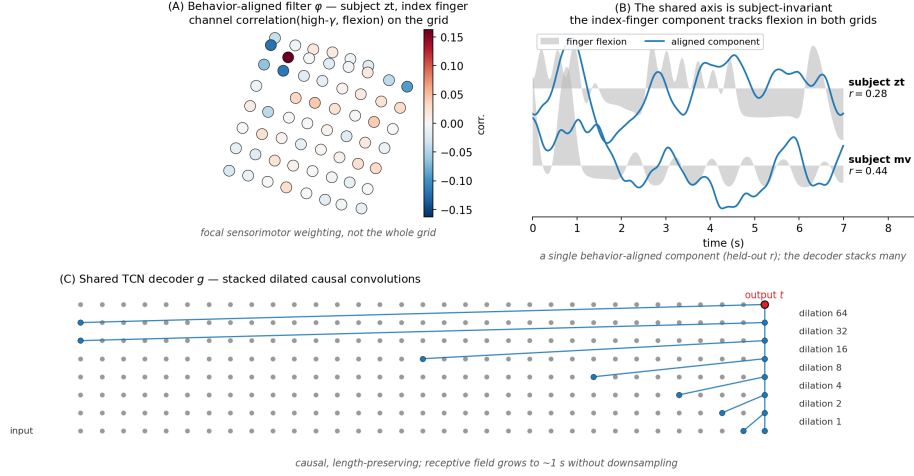


Figure S1: (A) A behavior-aligned filter as a cortical spatial map: the per-channel correlation between high-gamma activity and index-finger flexion for subject *zt*. (B) The same index-finger axis tracks flexion in two subjects whose grids do not correspond, using a single behavior-aligned component (held-out r shown). (C) The shared TCN decoder: stacked dilated causal convolutions with dilations growing $1 \rightarrow 64$, giving a receptive field that spans the full 512-frame (~ 5 s) training segment.

S3. Per-subject decoding accuracy

Table S2 breaks the three main-text accuracy figures out per target subject. The means and across-subject standard deviations on the last row are the numbers quoted in the main text; the per-subject rows are given here so that every reported mean can be audited against its parts.

S4. Hyperparameter sensitivity

Thirteen configurations were run under the full leave-one-subject-out protocol on the seven target subjects, sweeping one axis at a time from the reference

configuration ($R = 4$, $D = 20$, decoder width 128, maximum dilation 64). Table S3 gives every configuration with its accuracy in both transfer regimes and its calibration cost in wall-clock seconds, and Figure S2 plots the three swept axes.

R and D are not independent hyperparameters in this construction: $D = 5R$ by definition, one bundle of R filters per finger, so a sweep over R is simultaneously a sweep over D and there is no separate D axis to sweep. $R = 4$ is a genuine optimum on the minimal-calibration metric (0.530, 0.539, 0.557, 0.569, 0.561, 0.554 for $R = 1, 2, 3, 4, 6, 8$), and calibration cost rises monotonically with R , so the chosen value is also the cheapest of the configurations that reach the top of the curve. The dilation axis is flat where it matters: 32 and 64 give 0.5708 and 0.5687, a gap of 0.002 against a seed-to-seed standard deviation of 0.013 over twelve seeds of the reference configuration, so we report the range as flat and claim no optimum on it. The width axis carries a deployment observation rather than an accuracy one: width 64 reaches 0.5667 with 435,205 parameters against width 128’s 0.5687 with 1,673,221, so nearly all of the accuracy is available at a quarter of the parameter count.

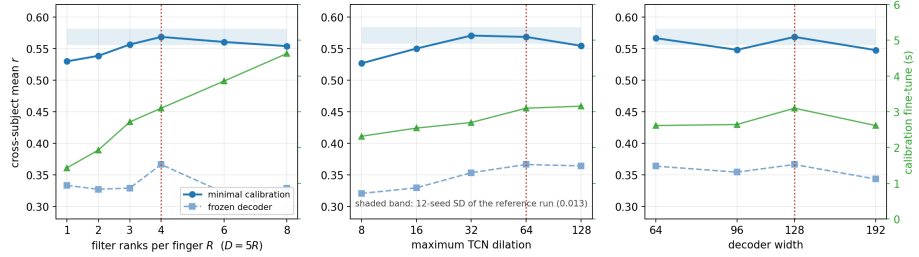


Figure S2: Accuracy and calibration cost across the three swept axes. Solid blue is minimal-calibration cross-subject r , dashed light blue is frozen-decoder r , green triangles are calibration fine-tuning wall clock on the right-hand axis. The dotted red line marks the configuration used in the paper. The shaded band is the twelve-seed standard deviation of the reference run (0.013) around the best value on each axis; differences inside it are not resolvable.

S5. Computational cost

Table S4 reports parameter count, decoder arithmetic, single-window latency on both GPU and CPU, and peak GPU memory, for the proposed method and for both channel-truncation baselines. Latency is per one-second window at the 100 Hz feature rate, measured over repeated forward passes with the median reported and the 95th percentile alongside it. Calibration fine-tuning costs between 1.5 and 4.4 s of wall clock across the seven target subjects.

The row that matters for a deployment argument is the last one. The decoder is cheap, at 0.334 GFLOP and 1.71 ms per second of signal, and the alignment projection is cheaper still at 0.41 ms per second, but the filterbank front end costs 339.98 ms per second of signal, two orders of magnitude more than the decoder and shared by every arm including the baselines. Any comparison of deployment cost against a large pretrained model should therefore be made on total pipeline cost, not on decoder size, and our advantage on that comparison is smaller than the parameter counts alone would suggest.

S6. Component clustering detail

Every quantity in this section is computed on the held-out test slice, which no alignment filter and no decoder has seen, so none of it is circular. Each component is described in two ways: functionally, by its correlation with each of the five flexion traces in each frequency band, giving a $5 \times \text{bands}$ signature; and anatomically, by the filter-weighted centroid of its electrodes in MNI coordinates. All comparisons are made across subjects and never within one.

The null throughout is the label-permuted control that the main-text ablation already uses: each subject’s filters are rebuilt against independently shuffled finger labels. That control is the right null here because a supervised projection with no shared axis still denoises and still keeps within-subject structure, so any metric that survives it is not measuring supervision alone.

Table S5 gives the four metrics. Cross-subject finger retrieval, which asks whether the nearest component in another subject carries the same finger,

reaches 0.3036 over 840 queries against a permuted null of 0.1976 and chance of 0.200. The adjusted Rand index of k -means with $k = 5$ against finger identity is 0.072 against a null of -0.0083 . Both are modest in absolute terms and the silhouette is slightly negative at -0.0265 , so the components do not form clean per-finger clusters; what they do carry is cross-subject information about which finger they encode, above what a subject-specific supervised projection alone provides. The adjusted Rand index against subject identity is higher than against finger identity (0.1466), which says the same thing from the other side: subject is still the stronger organizing factor in the pooled signatures, and the finger signal sits on top of it.

The anatomical columns are the negative half of the result and are reported as such. The mean MNI centroid distance between same-finger components in different subjects is 35.54 mm and between different-finger components 35.51 mm, a gap of -0.031 mm with a permutation p of 0.53105 over 10^4 permutations. Same-finger components are anatomically no closer than different-finger ones. The correspondence the method constructs is functional and not anatomical, which is consistent with the main-text finding that a principled anatomical alignment collapses, and it is the reason we make no claim of anatomical homology.

S7. Coverage audit of a second cohort

The method states a precondition: a behavior-defined axis can be estimated only if some electrode observes the cortical territory that the calibration behavior drives. This section measures that precondition on a cohort where it is not guaranteed, so that the boundary of the method is quantified rather than assumed.

The cohort is twelve patients recorded during long-term clinical epilepsy monitoring, with 63 to 117 good electrodes each and annotated arm reaches with simultaneous pose tracking [2]. It is a deliberate stress case: the implants were placed for seizure localization, not for motor research, so coverage of hand

and arm motor cortex is incidental rather than designed. For each subject, epochs were cut from 3 s before to 3 s after each annotated reach onset, up to 120 reaches; the 70–200 Hz envelope was computed, z-scored against the pre-onset baseline, and averaged over epochs. Two numbers come out per subject: the peak post-onset z on the most responsive electrode, and the number of electrodes exceeding $z = 2$. The wrist speed profile over the same epochs is the control: if speed does not rise after the annotated onsets, the annotations or the time base are wrong and the high-gamma result would mean nothing.

Table S6 gives the audit. The control passes: wrist speed rises after onset in 11 of 12 subjects, so the annotations and the time base are sound. The precondition does not: only 6 of 12 subjects have any electrode above threshold. Three of those are unambiguous, sub-03 with peak z 10.8 on 7 electrodes, sub-05 with 20.7 on 5, and sub-11 with 4.6 on 14; the other three are marginal, at one to three electrodes each. In a cohort implanted for clinical rather than motor-research reasons, the precondition therefore holds for roughly half the subjects, and this measurement is what we recommend running before applying the method to a new cohort.

S8. Full comparison of methods that build a shared space

Table 1 of the main text carries the two closest precedents, hyperalignment and the shared response model, against ours. Table S7 here gives the full itemisation, adding the three further families that a reader may want checked against the same properties: Euclidean Alignment as the representative generic statistical alignment, anatomical interpolation onto a common cortical grid, and foundation-model pretraining. The properties are the ones that decide whether a method can be applied at all when the grids do not correspond. *Paired samples* means temporally synchronized or otherwise matched observations of the same event in two subjects; *sensor correspondence* means a bijection between the two subjects’ recording channels. Only the last row requires neither.

References

- [1] W. Yang, Y. Sun, B. Van Dyck, B. Calvo Merino, M. M. Van Hulle, CORTEG: Foundation models enable cross-modality representation transfer from scalp to intracranial brain recordings (2026). arXiv:2605.10337.
- [2] S. M. Peterson, S. H. Singh, B. Dichter, M. Scheid, R. P. N. Rao, B. W. Brunton, AJILE12: Long-term naturalistic human intracranial neural recordings and pose, *Scientific Data* 9 (2022) 184. doi:10.1038/s41597-022-01280-y.
- [3] J. V. Haxby, J. S. Guntupalli, A. C. Connolly, Y. O. Halchenko, B. R. Conroy, M. I. Gobbini, M. Hanke, P. J. Ramadge, A common, high-dimensional model of the representational space in human ventral temporal cortex, *Neuron* 72 (2) (2011) 404–416. doi:10.1016/j.neuron.2011.08.026.
- [4] P.-H. Chen, J. Chen, Y. Yeshurun, U. Hasson, J. V. Haxby, P. J. Ramadge, A reduced-dimension fMRI shared response model, in: *Advances in Neural Information Processing Systems* 28 (NeurIPS 2015), 2015, pp. 460–468.
- [5] H. He, D. Wu, Transfer learning for brain-computer interfaces: A Euclidean space data alignment approach, *IEEE Transactions on Biomedical Engineering* 67 (2) (2020) 399–410. doi:10.1109/TBME.2019.2913914.

Table S1: Reproducibility settings for the alignment and the shared TCN decoder. The feature rate is 100 Hz throughout.

Component	Setting
Sampling / features	1000 Hz raw; 24 log-spaced bands over 2–300 Hz; fourth-order Butterworth bandpass then Hilbert envelope; decimated to 100 Hz
Referencing	per-channel z-score, then across-channel median subtracted
Alignment φ	rank-1 correlation filter (Eq. S1) + $R-1=3$ SVD directions of the $ \text{corr} $ -reweighted high-gamma covariance; $R=4$, $D=5R=20$
High-gamma band	70–150 Hz
Target smoothing	Gaussian, $\sigma = 6$ frames (60 ms), applied within each contiguous split
Decoder	dilated causal TCN; 7 residual blocks, dilations 1, 2, 4, $\dots$, 64; width 128; kernel size 7; Batch-Norm; GELU; dropout 0.2; 1×1 input and output convolutions
Training segments	length 512 frames, stride 256
Loss	$\text{MSE} + 0.5(1 - \text{mean Pearson } r)$
Optimizer	AdamW, lr 2×10^{-3} , weight decay 10^{-4} , cosine schedule; batch 32; ≤ 120 epochs, early stop (patience 14) on validation r
Fine-tune (calibration)	40 epochs, lr 5×10^{-4} , warm-started from the source model
Ensemble	3 seeds, test predictions averaged
Data split	train 0–65% / validation 65–80% / test 80–100% (contiguous, temporal order)
Compute	single GPU

Table S2: Leave-one-subject-out decoding accuracy (mean Pearson r over five fingers). Within-subject, frozen-decoder cross-subject, and minimal-calibration (fine-tuned) cross-subject. The alignment is fitted from labeled target movement in both cross-subject columns; the columns differ in whether target labels also update the shared decoder.

Target subject	Within	Frozen-decoder	Fine-tuned (min.-calib.)
bp	0.452	0.336	0.556
zt	0.679	0.528	0.686
jp	0.494	0.321	0.571
ht	0.433	0.221	0.412
mv	0.523	0.426	0.613
wc	0.445	0.321	0.530
jc	0.577	0.333	0.595
Mean $\pm$ SD	0.515 $\pm$ 0.088	0.355 $\pm$ 0.097	0.566 $\pm$ 0.084

Reference: CORTEG (foundation-model SOTA, fine-tuned/calibrated): 0.554 [1]

The canonical cross-subject figure is 0.566 (seven-subject, three-seed ensemble). The other nearby values are matched-seed sub-runs reported for specific contrasts: 0.550 (full nine-subject library), 0.573 (seven-subject subset within the nine-subject run), 0.557 (ablation-table seed set), 0.570 (permute-control seed pair), and 0.532 (single-seed mean for the seed-spread estimate).

Table S3: Hyperparameter sweep, thirteen configurations. Accuracy is the mean over the seven target subjects; pretrain and fine-tune columns are mean wall-clock seconds per target. The first row is the reference configuration used throughout the paper.

Swept axis	R	D	Width	Max dil.	Params	Within	Frozen	Fine-tuned	Pretrain (s)	Fine-tune (s)
reference	4	20	128	64	1,673,221	0.515	0.366	0.569	21.7	3.1
R	1	5	128	64	1,627,141	0.487	0.333	0.530	10.2	1.4
R	2	10	128	64	1,642,501	0.509	0.327	0.539	15.1	1.9
R	3	15	128	64	1,657,861	0.510	0.329	0.557	18.4	2.7
R	6	30	128	64	1,703,941	0.499	0.322	0.561	38.9	3.9
R	8	40	128	64	1,734,661	0.490	0.329	0.554	35.5	4.6
Width	4	20	64	64	435,205	0.514	0.364	0.567	17.3	2.6
Width	4	20	96	64	953,861	0.523	0.354	0.548	17.9	2.6
Width	4	20	192	64	3,714,053	0.485	0.343	0.548	17.9	2.6
Dilation	4	20	128	8	982,789	0.468	0.320	0.527	18.1	2.3
Dilation	4	20	128	16	1,212,933	0.490	0.329	0.550	20.1	2.5
Dilation	4	20	128	32	1,443,077	0.518	0.353	0.571	20.1	2.7
Dilation	4	20	128	128	1,903,365	0.507	0.364	0.555	21.7	3.2

Table S4: Computational cost of the proposed method and the two channel-truncation baselines. FLOP and latency figures are per second of signal at the 100Hz feature rate. The front-end and alignment columns are shared across arms and are reported once in the text.

Model	Input dim.	Params	GFLOP/s	GPU ms (p95)	CPU ms (p95)	Peak GPU (MiB)
Behavior-aligned + TCN (ours)	480	1,673,221	0.334	1.713 (1.778)	8.028 (8.220)	7.3
First-43 channels + TCN	1032	1,743,877	0.348	1.715 (1.769)	8.114 (8.360)	8.1
First-43 + Euclidean Alignment + TCN	1032	1,743,877	0.348	1.703 (1.782)	8.216 (8.434)	8.1

Shared across all three arms: the filterbank front end costs 339.98 ms per second of signal, and the alignment projection costs 0.41 ms per second. Calibration fine-tuning costs 1.5 to 4.4 s per target subject.

Table S5: Cross-subject structure of the shared components against the label-permuted null. Retrieval is over 840 cross-subject queries with a chance level of 0.200; ARI is against k -means with $k = 5$ on the pooled component signatures; MNI columns are mean centroid distances between components in different subjects.

Components	Retrieval	n	Chance	Silhouette	ARI (finger)	ARI (subject)	MNI same	MNI diff.
Shared behavior-defined axes	0.3036	840	0.200	-0.0265	0.072	0.1466	35.54 mm	35.51 mm
Permuted-label null	0.1976	840	0.200	-0.0511	-0.0083	0.2219	35.49 mm	35.52 mm

Same-finger versus different-finger MNI gap: -0.031 mm, permutation $p = 0.53105$ over 10^4 permutations, for the shared axes.

Table S6: Functional coverage audit on a twelve-subject clinical ECoG cohort [2]. Peak z is the largest post-onset high-gamma response on the most responsive electrode; the count is of electrodes exceeding $z = 2$. Speed rise is the ratio of post-onset to pre-onset wrist speed and is the control on the annotations and time base.

Subject	Reaches	Electrodes	Wrist	Peak z	Peak t (s)	Above $z=2$	Speed rise
sub-01	58	91	R	0.71	1.20	0	3.55
sub-02	80	81	L	0.79	0.42	0	1.42
sub-03	114	79	R	10.81	1.34	7	2.17
sub-04	13	67	L	2.20	0.02	1	0.71
sub-05	120	100	L	20.74	1.88	5	1.81
sub-06	82	70	R	0.90	1.22	0	3.95
sub-07	93	63	L	2.36	0.18	3	2.10
sub-08	117	82	L	2.19	0.66	1	5.68
sub-09	31	106	R	1.43	1.36	0	3.94
sub-10	111	117	R	1.04	1.70	0	2.33
sub-11	120	103	R	4.57	1.26	14	2.93
sub-12	118	111	L	0.96	0.68	0	1.48
Coverage present in 6 of 12 subjects (any electrode above threshold); wrist speed rises after onset in 11 of 12.							

Table S7: Families of methods that build a space shared across subjects, itemised on the properties that decide applicability across non-corresponding grids. The first two rows and the last are the compact comparison of main-text Table 1; the three middle rows extend it.

Family	Shared axes defined by	Mapping estimated from	estimated from	Paired samples?	Sensor cor-resp.?	What transfers
Hyperalignment [3]	responses to a stimulus stream common to all subjects	time-synchronized responses of several subjects to that stream		yes	no	the common representational space
Shared response model [4]	latent factors common to subjects viewing one stimulus	synchronized multi-subject responses		yes	no	shared latent time courses
Euclidean Alignment [5]	the subject's own channel space, whitened	each subject's trial covariance, unlabeled		no	yes	a classifier on the shared channel set
Anatomical interpolation	a common cortical coordinate frame	registered electrode coordinates		no	no, but needs registration	a decoder on the common grid
Foundation-model pretraining [1]	a latent learned by self-supervision at scale	a large multi-subject corpus, then target calibration		no	no	the encoder plus a light head
Behavior-supervised alignment (ours)	the decode target, one axis bundle per finger	each subject's own recording and behavior alone		no	no	one decoder on the behavior-defined axes