

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*

Abstract

Cross-subject neural decoding asks a learning system to reuse what it has learned from one set of brains on a new brain whose recording sensors do not line up with the old ones. For electrocorticography (ECoG), every implant places a different electrode grid over a different patch of cortex, so the input spaces of two subjects share no feature index and no geometric correspondence. This breaks the assumption, standard in domain adaptation and transfer learning, that source and target live in a common feature space. We recast cross-subject decoding as transfer across non-corresponding input spaces and solve it with a behavior-supervised alignment. For each subject, a per-subject spatial filter projects that subject’s grid onto a shared, behavior-defined component space: the axes are the cortical projections of each finger’s flexion, estimated from training behavior alone. A single temporal convolutional decoder, trained on pooled source subjects, then reads these shared components and predicts continuous finger trajectories. The alignment is estimated from labeled target movement in every regime reported here, so no regime is label-free or calibration-free; the regimes differ only in whether target labels also update the shared decoder. On the Stanford fingerflex dataset, leave-one-subject-out, our minimal-calibration cross-subject decoder reaches a mean Pearson r of 0.566 ± 0.084 over five fingers

*Corresponding author.

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

on seven subjects, and 0.550 across the full nine-subject library. Run under the protocol of the CORTEG foundation model, on all nine subjects with their data division and their evaluation rule, the same decoder reaches 0.530 (SD 0.157) against their published 0.554, a difference well inside the across-subject spread, so we read the two as occupying one accuracy band rather than as a win either way. The decoder uses about four to five minutes of target calibration instead of ten to thirty, and keeps an interpretable, auditable alignment front end and shared space. A same-protocol ablation shows that generic statistical alignment cannot bridge non-corresponding grids (Euclidean Alignment: 0.037 frozen-decoder, 0.357 fine-tuned), whereas behavior-supervised alignment can (0.371 frozen-decoder, 0.557 fine-tuned; the contrast is significant in both regimes, Wilcoxon signed-rank $p = 0.016$). A randomly initialized alignment head of the same capacity, trained identically but without behavioral supervision, falls to 0.319 with calibration and 0.260 with the decoder frozen, which places the gain on the behavioral correspondence rather than on trainable capacity or on the fine-tuning protocol, and the pattern replicates on an independent four-patient cursor-tracking cohort with a different behavior and different grids. The alignment, not the decoder, is what crosses the grid gap.

Keywords: transfer learning, domain adaptation, cross-subject decoding, electrocorticography, brain-computer interface, representation learning

1. Introduction

Continuous finger movement can be decoded from ECoG accurately enough to drive useful control in patients who can still produce the movements, as in the epilepsy-monitoring cohorts on which these decoders are built. The bottleneck is not the decoder’s ceiling on any one patient but the cost of reaching that ceiling for each new patient. Building a within-subject decoder demands a calibration session in which the patient produces labeled movements while the implant records, and the resulting model serves that patient only. Every new patient pays the full calibration cost again, and for patients who fatigue quickly or whose

movement repertoire is limited, a long calibration session is itself a barrier. A decoder that could be trained once on a population and then transferred to a new patient with a few minutes of calibration would change the deployment economics of clinical brain-computer interfaces [1, 2]. The accuracy ceilings on individual subjects are already adequate for useful control; the open problem is amortizing the cost of reaching them across patients, which is a transfer problem.

The obstacle to such transfer is structural, and it is a learning problem rather than a signal-processing detail. Two ECoG subjects rarely share electrode counts, electrode positions, or the cortical territory under the grid. Implant geometry is dictated by clinical need, the surgeon places electrodes to map a patient’s seizure focus or eloquent cortex, not to match some other patient’s array. One subject may carry a dense 8×8 grid over hand cortex while another carries a sparse strip that grazes it. There is no channel k that means the same thing across subjects, so the input spaces do not correspond. Classical domain adaptation and transfer learning assume the opposite: source and target samples are drawn over a common feature space, and the task is to correct a distribution shift on that shared space [3, 4, 5]. When the feature index itself differs from subject to subject, the premise that makes alignment meaningful is gone: one cannot match marginals, match second moments, or learn a domain-invariant representation over coordinates that name different things in each domain. There is no shared support to begin with.

It helps to separate two notions that the cross-session literature usually conflates. Cross-session transfer within one subject keeps the electrode array fixed and only the recording conditions drift, so the channel index is stable and a distribution-matching correction is well posed. Cross-subject transfer across implants changes the channel index itself. The first is a covariate-shift problem, the second a correspondence problem, and most of the BCI transfer toolkit was designed for the first and is silently reused for the second, which is where it breaks.

Generic statistical alignment, the workhorse of cross-session and cross-subject BCI, inherits exactly this limitation. Correlation Alignment whitens and recol-

ors a source covariance to match a target covariance [6]; Euclidean Alignment re-references each domain’s trials by the inverse square root of its mean covariance [7]. Both are maps from a fixed channel space to itself, so across non-corresponding grids they either require an arbitrary truncation to a common channel count, which destroys spatial meaning, or do not apply at all. Our ablation makes this concrete: truncating to a common channel set and then applying Euclidean Alignment leaves cross-subject decoding near chance (frozen-decoder $r = 0.037$).

Foundation models take the opposite route and win by scale. Large neural recordings are pretrained with self-supervised objectives, after which a light head and a calibration session adapt the model to a new subject and task [8, 9, 10, 11], which sidesteps the correspondence problem by absorbing it into a high-capacity model. CORTEG, trained across many ECoG subjects, reaches a cross-subject finger-decoding mean r of 0.554 [8]. These models are effective, and they pay for it on three fronts: their internal axes have no neurophysiological reading, so a decode failure cannot be traced to a cortical site or a movement; they depend on a large and often inaccessible pretraining corpus; and they typically need ten to thirty minutes of target calibration to specialize, which for a fatigued patient is a long session. For a clinical decoder those two costs determine whether the system can be trusted and whether it can be set up at the bedside.

We propose a different bridge across the grid gap, one that is supervised by behavior. For each subject we fit a per-subject spatial filter that projects that subject’s electrodes onto a small set of components whose meaning is fixed across subjects by the decode target itself. Concretely, for each of the five fingers we extract spatial patterns from the training-set correlation between each channel’s high-gamma envelope and that finger’s flexion, giving $5R$ behavior-defined components. These components are *the same axes* for every subject, not because the grids match, but because each axis is defined as “the cortical projection of finger f ,” a behavioral concept that exists independent of geometry. On this shared component space a single temporal convolutional network (TCN), trained on the pooled source subjects, decodes continuous finger trajectories.

The subject-specific part of the model is only the spatial filter φ_i ; the decoder g is shared. Following standard domain-adaptation usage we call the subjects the shared decoder is trained on the *source* subjects, or source domain, and the held-out subject the *target*.

The intuition is that movement is the invariant the recordings lack. Two subjects flexing the same finger engage homologous cortical machinery, even though the electrodes that observe it are placed differently. Defining the shared axes by the movement rather than by the sensors makes them comparable across subjects for the same reason the movement is, and leaves the spatial filter one job: translating each subject’s idiosyncratic grid into the language of those axes.

One thing this construction does not do is remove the need for target labels, and we are explicit about it throughout. The alignment φ_t is estimated from labeled target movement in both transfer regimes reported here. What the regimes separate is the second use of target labels, updating the shared decoder, which the minimal-calibration regime performs and the frozen-decoder regime does not. Neither regime is label-free, and neither is calibration-free.

The payoff is a cross-subject decoder that is at once accurate, cheap, and legible. With minimal calibration it attains a leave-one-subject-out mean r of 0.566 ± 0.084 (SEM 0.032, $n = 7$; 0.550 across the full nine-subject library), reaching the accuracy band of the CORTEG foundation model at 0.554 [8]. The 0.012 difference is smaller than the decoder’s own seed-to-seed spread (single-seed SD 0.013), so we treat the two as comparable rather than as a win. Run under CORTEG’s own protocol, on all nine subjects with their chronological data division and their evaluation rule, the same decoder reaches 0.530 (SD 0.157) against their published 0.554, which puts the comparison on one protocol and leaves the reading unchanged. The decoder needs about four to five minutes of target calibration rather than ten to thirty, and it stays inspectable: the leading alignment filter for each finger is a forward pattern one can plot on the cortex and read against known motor somatotopy [12, 13].

Our contributions are learning-systems results about transfer across non-corresponding input spaces. The fixed, hand-computed alignment φ is a de-

liberate choice, not a limitation we settled for: it is a strong and transparent baseline for the transfer problem, and pairing it with a stock temporal convolutional decoder keeps the credit on the alignment rather than on a bespoke architecture.

- We formalize cross-subject neural decoding as transfer across non-corresponding input spaces, where source and target share neither a feature index nor a geometry, and cast it as a label/behavior-supervised common-subspace heterogeneous domain adaptation without paired or pivot instances [3, 4, 14].
- We show that source pretraining on this shared space acts as a transferable prior: the fine-tuned cross-subject decoder meets or exceeds from-scratch within-subject decoding for six of the seven subjects, with the seventh (ht) marginally below (0.412 vs 0.433); the mean gain is +0.05 (Wilcoxon signed-rank $p = 0.078$, a trend), not a uniform per-subject win.
- We introduce behavior-supervised alignment: a per-subject spatial filter that maps each heterogeneous grid onto a shared latent space whose axes are defined by the decode target, making the shared space common by construction and independent of grid geometry.
- We pair this alignment with a single shared TCN decoder and a transfer protocol (frozen-decoder and minimal-calibration), and through a same-protocol ablation isolate the alignment as the load-bearing component: behavior alignment carries cross-subject transfer where generic statistical alignment, including Euclidean Alignment on a forced common channel set, cannot [7]. A randomly initialized alignment head of matched capacity, trained identically without behavioral supervision, separates the contribution of the correspondence from that of trainable capacity and of the fine-tuning protocol.
- We report the method on a second, independent ECoG cohort, four cursor-tracking patients who appear in no part of the fingerflex library and per-

form a different behavior, and we run our decoder under the published protocol of the foundation-model reference so that the external comparison is on one data division, one subject count and one evaluation rule.

2. Related Work

Domain adaptation and transfer learning give us the vocabulary for cross-subject decoding, but their guarantees presuppose a shared input space. The canonical taxonomy organizes transfer by what is shared between source and target, and the feature-representation branch assumes a common feature space on which distributions can be matched [3, 4, 15]. Whether that match is made by kernel embeddings of layer activations [16], by adversarial domain confusion as in ADDA and DANN [17, 18, 19], by a meta-learned initialization [20], or by contrastive prediction of future latents [21], the objective requires that coordinate k of a source sample and coordinate k of a target sample be comparable. Across non-corresponding grids there is no such coordinate. Heterogeneous domain adaptation relaxes that assumption, aligning domains that share neither features nor instances, one line of it through labels rather than instance correspondences [14]. Our setting is an instance of it, with no paired or pivot instances and the shared axes defined by the decode target rather than by matching distributions.

Within BCI, transfer usually takes the form of generic statistical alignment, and there the common-channel assumption is sharp. Correlation Alignment recolors features to match second-order statistics [6, 22], Euclidean Alignment whitens each subject’s trials by the inverse square root of the average trial covariance [7], Riemannian methods transport or recenter distributions of covariances on their manifold [23, 24, 25], and joint decorrelation builds its operators from the same channel-by-channel covariance [26]. Every one of these operators is a map on a single fixed channel space: the whitener in Euclidean Alignment is $C_i \times C_i$ and acts on the same C_i channels in source and target. Across non-corresponding grids that correspondence does not exist, and forcing it by

truncating every grid to a common channel count pairs up channels observing unrelated cortex. Section 4.3 measures the degeneracy: Euclidean Alignment on a forced common channel set falls back to the no-alignment baseline. The operator is sound for the cross-session problem it was built for and inert for the cross-subject one.

A third line builds explicit shared and private subspaces: domain separation networks split representations under orthogonality and reconstruction constraints [27], and recent neural-data work mixes per-subject tokens into a shared latent [28], aligns latent spaces across recordings [29], and aligns brain recordings to a shared space for speech decoding [30]. The closest precedent is functional alignment for neuroimaging: hyperalignment maps non-corresponding voxel spaces into a common space defined by shared task and stimulus responses [31], and the shared response model learns such a space from a shared external variable [32]. These are the sister idea to ours, bringing correspondence in from shared task structure rather than from the recordings.

Table 1 itemises those two precedents against ours on the properties that decide whether a method applies at all when the grids do not correspond; Supplementary Section S8 extends the itemisation to Euclidean Alignment, anatomical interpolation onto a common grid, and foundation-model pretraining.

The distinguishing claim is the last two columns: our shared axes are estimated from each subject’s own recording and behavior alone, so the construction needs no synchronized stimulus and no correspondence between the sensor sets, where hyperalignment and the shared response model reach a comparable structural goal only from temporally synchronized instances. Euclidean Alignment needs a shared channel set, and anatomical interpolation needs registered coordinates and collapses even with them (Section 4.3). Elsewhere in the shared-latent literature the shared space is discovered by unsupervised reconstruction, by adversarial invariance, or by contrastive alignment to a reference, never by the behavior being decoded. Making the decode target itself the supervisor grounds each axis in the task rather than in distributional similarity, and leaves it legible before any decoder is trained on it, because each axis names a movement rather

Table 1: The distinguishing requirement, set against the two closest precedents. *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. Supplementary Section S8, Table S7 itemises three further families on the same properties.

Family	Shared axes defined by	Mapping estimated from	Paired samples?	Sensor cor- resp.?
Hyperalignment [31]	responses to a stimulus stream common to all subjects	time-synchronized responses of several subjects to that stream	yes	no
Shared response model [32]	latent factors common to subjects viewing one stimulus	synchronized subject responses	yes	no
Behavior-supervised alignment (ours)	the decode target, one axis bundle per finger	each subject’s own recording and behavior alone	no	no

than an arbitrary direction in a learned basis.

Foundation models for neural data are the fourth camp and the current accuracy benchmark for cross-subject decoding. Large self-supervised models pretrain on pooled recordings and then transfer with a small head: POYO and population-transformer style models for spiking and field data [33, 9], BrainBERT-style masked models for intracranial signals [34, 10], LaBraM for EEG [11], and broad decoding benchmarks built on top of them [35]. CORTEG is the most directly comparable, reporting cross-subject finger-decoding mean $r = 0.554$ after calibration [8]. This camp answers the correspondence problem by not answering it: a sufficiently expressive pretrained encoder accommodates whatever grid a new subject brings, at the price of the scale, the opacity and the calibration that buy that flexibility. Our aim is to reach the same accuracy band by answering the correspondence problem directly and cheaply, with a small behavior-defined map.

The decoder itself sits in a well-developed lineage: linear and ridge models on spectral features [36, 37], optimal-control and Kalman approaches for closed-loop stability [2], high-performance intracortical speech and handwriting decoders [1], and the U-Net-style FingerFlex [38], the within-subject benchmark on the Stanford finger task, which rests on spatial-filter and pattern interpretations of ECoG motor decoding [39]. We borrow that pattern view for a different purpose, defining a shared cross-subject axis rather than a within-subject filter. Our own decoder is a stock dilated causal temporal convolutional network [40, 41, 42] with a U-Net-like encoder-decoder shape [43], chosen for stable training on long continuous trajectories and a causal receptive field that suits an online clinical setting. Compact EEG and ECoG networks [44, 45, 46] and tensor-structured decoders [47] build spatial structure in through learned spatial convolutions, where we factor that step out into the explicit, behavior-supervised alignment. The lineage supplies the decoder; the contribution is the alignment that lets one such decoder serve many subjects.

3. Methods

Figure 1 is an end-to-end flowchart of the pipeline; the parts follow in detail.

3.1. Problem formulation

We are given N source subjects and one target subject, using the source and target domain terminology standard in transfer learning [3, 4]. Source subject i records ECoG on a grid G_i with C_i electrodes at subject-specific cortical positions; the target records on G_t with C_t electrodes. The grids do not correspond: there is no bijection between the electrodes of G_i and those of G_j , the electrode counts differ, and the cortical territory covered by each grid differs. For each subject, the data are a multichannel time series $X_i \in \mathbb{R}^{C_i \times T}$ and a five-dimensional flexion trajectory $Y_i \in \mathbb{R}^{5 \times T}$, one continuous channel per finger.

A within-subject decoder learns a map $f_i : \mathbb{R}^{C_i \times T} \rightarrow \mathbb{R}^{5 \times T}$ from that subject’s own data. Cross-subject transfer would reuse such a map on a new subject,

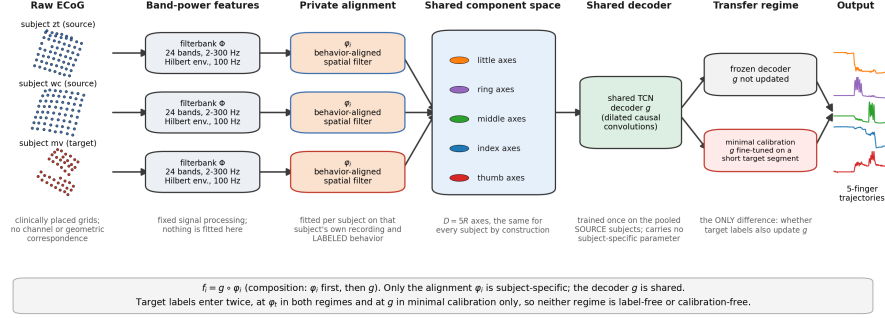


Figure 1: End-to-end flowchart of the pipeline, from raw recordings to predicted trajectories. Non-corresponding electrode grids enter on the left and are turned into 24-band envelope features (Section 3.2); a private, behavior-aligned spatial filter ϕ_i projects each subject’s grid onto a shared finger-motor component space (Section 3.3); a single shared TCN decoder g , trained on the pooled source subjects (Section 3.4), reads that space and outputs continuous finger trajectories under one of the two transfer regimes (Section 3.5). Each stage is annotated with the data it is fitted on. Target labels enter twice and the two entries are distinct: they estimate the target alignment ϕ_t in both regimes, and they additionally update the shared decoder in the minimal-calibration regime only. Only the spatial filter is subject-specific; the decoder is the same for every subject.

but f_i takes inputs on G_i , and $G_t \neq G_i$ with no correspondence, so f_i cannot be applied to target data at all. The standard transfer-learning fix, correcting a distribution shift over a shared feature space [3, 4], is unavailable: there is no shared feature space to correct over.

We therefore factor the decoder as

$$f_i = g \circ \phi_i, \quad (1)$$

where $\circ$ denotes function composition, read right to left: the per-subject alignment ϕ_i is applied first and the shared decoder g is applied to its output. Here $\phi_i : \mathbb{R}^{C_i \times T} \rightarrow \mathbb{R}^{D \times T}$ is a per-subject alignment that maps subject i ’s grid onto a common D -dimensional component space, and $g : \mathbb{R}^{D \times T} \rightarrow \mathbb{R}^{5 \times T}$ is a single decoder shared by all subjects. The component space is fixed across subjects, so g has a well-defined input regardless of which grid produced it. Only ϕ_i is subject-specific. Transfer to a new target reduces to fitting ϕ_t for the target and

reusing the shared g learned from the source subjects. The grid gap is crossed once, inside φ_i , and never seen by the decoder.

This factorization places the entire subject heterogeneity in one component and the entire shared knowledge in the other, which is what makes the two parts trainable from different data. The alignment φ_i needs only subject i 's own recording and behavior; it never compares subject i to subject j , so it imposes no requirement that grids match. It does need that subject's behavior labels: the correlation that defines the filter is computed against the measured flexion, so fitting φ_t on the target consumes labeled target movement. The decoder g needs only data already mapped into the shared space; it never sees a raw grid, so it carries no knowledge of any particular geometry. The dimension D of the shared space is the same for every subject by construction, set by the number of behavioral axes rather than by any electrode count, which is precisely the property that a fixed channel truncation cannot provide.

3.2. Features

This subsection has been moved forward, so that the features are defined before the alignment and the decoder that consume them. Features are spectral power in 24 log-spaced bands spanning 2–300 Hz, capturing both the high-gamma band that carries the sharpest finger-specific motor information and the low-frequency motor potential (LMP) that tracks movement at slower timescales [13, 48, 49]. Each channel is z-scored and then re-referenced by subtracting the across-channel median, a common reference; each band power is the Hilbert envelope of a fourth-order Butterworth bandpass, computed per channel and decimated to a 100 Hz feature rate. The flexion targets are Gaussian-smoothed (standard deviation six feature frames, 60 ms) to suppress sensor noise in the continuous trajectory, and the smoothing is applied within each contiguous split so its kernel does not straddle the calibration/test boundary. The high-gamma envelope used to define the behavior-alignment filters (Section 3.3) is a 70–150 Hz band from this same pipeline. The alignment filter for each finger is estimated from that high-gamma band but applied to all 24 bands, on the ra-

tionale that a channel’s proximity to a finger’s cortical representation governs its coupling in the lower bands too; the shared spatial axis reflects where a finger projects on the cortex, not one frequency. After the alignment projects each grid onto the $D = 20$ behavior-defined components, the decoder input at each time step is $20 \text{ components} \times 24 \text{ bands}$.

3.3. Behavior-aligned spatial filtering

The alignment φ_i is the core of the method, and its design is what makes the component space common across non-corresponding grids. We define each component not by a geometric or statistical property of the grid, but by the behavior it explains.

Consider one finger $f \in \{1, \dots, 5\}$ and one subject i . From the training portion of that subject’s recording we compute, for each electrode c , the Pearson correlation between the electrode’s high-gamma envelope $h_{i,c}(t)$ and that finger’s flexion trajectory $Y_{i,f}(t)$. Collecting these correlations over all electrodes gives a spatial weighting that scores how strongly each electrode’s high-gamma activity tracks finger f . The first spatial filter for finger f is this correlation pattern, ℓ_2 -normalized. This rank-1 filter is a SPoC-type supervised spatial filter: SPoC finds spatial filters whose band power co-varies with a continuous behavioral target [50], which is exactly what the high-gamma-versus-flexion correlation does here. Our contribution is not the filter but its repurposing as a cross-subject shared axis, one per finger, that manufactures correspondence across non-corresponding grids; SPoC itself is a within-subject construction. The remaining $R - 1$ filters are the leading singular vectors of the high-gamma covariance re-weighted by the magnitude of that correlation pattern, which restricts the covariance to the finger-driven electrodes; each is sign-aligned to the correlation pattern so that the whole bundle points the same way along the finger’s cortical projection. This yields R filters per finger and $5R = D$ filters in total ($R = 4$, $D = 20$ in our experiments), spanning the finger-driven subspace. Stacking these filters into $W_i \in \mathbb{R}^{D \times C_i}$, the alignment projects the subject’s

per-band features through W_i :

$$\varphi_i(X_i) = W_i \Phi(X_i), \quad (2)$$

where Φ produces the per-channel spectral features described in Section 3.2.

Three properties follow from this construction. First, the component axes are common across subjects *by construction*. Axis f, r always means “the r -th cortical projection of finger f ,” a behaviorally defined quantity that exists for any subject regardless of how many electrodes the grid has or where they sit. Subject i and subject j may use entirely different electrode weightings to realize axis f, r , but the axis denotes the same behavioral source in both. The electrode count, spacing and cortical location all vary freely; what does not vary is the set of five fingers and the R projections kept for each, so a subject with sixty-four electrodes and a subject with forty-eight produce vectors in the same twenty-dimensional space, comparable because their axes are defined by the same five movements.

Second, the construction is leakage-safe. W_i is estimated from the training split only, so no target label information from the evaluation portion enters the alignment. In the cross-subject regimes, the target’s training split is used to fit φ_t , but the decoder either never sees target labels (the alignment still uses target labels; frozen-decoder) or sees only the short calibration segment (minimal-calibration), and in both cases the evaluation portion is held out entirely. The correlation that defines the filter is a training-set statistic, not a fitted parameter of the decoder, so it cannot memorize evaluation data.

Third, the leading spatial map for each finger is interpretable as it stands and the remaining filters are not. The two are produced by the same routine, and the distinction follows in three steps.

(i) A forward model, or activation pattern, states how strongly a hypothesized source is expressed at each sensor; a backward model, or extraction filter, states how the sensors are weighted to recover that source. Haufe et al. [12] establish that only the forward one may be read as a spatial distribution of cortical activity, because a backward filter can place a large weight on a channel

carrying no signal at all, purely to cancel noise picked up by another channel.

(ii) The rank-1 filter is, entry by entry, the correlation between one channel’s high-gamma envelope and that finger’s flexion (Supplementary Eq. S1). Each entry is a univariate statistic of that channel alone and depends on no other channel’s weight, which is the definition of an activation pattern: the map records where the behavior is expressed, not how channels are combined to recover it. It is therefore forward by construction, and no backward-to-forward transform is needed or applied.

(iii) The remaining $R - 1$ filters are the leading singular vectors of the correlation-reweighted high-gamma covariance. A singular vector is a multivariate direction selected for the variance structure of the channel set; its entries are defined jointly, and none is a per-channel relationship to the behavior. They are backward-model directions, the Haufe caveat applies in full, and we make no cortical-source claim for them; they are retained because they add decodable subspace and are read as capacity rather than as anatomy. The interpretability claim is therefore bounded to the rank-1 axis of each finger, the one Figure 3 plots, and nothing else in the model.

Because that axis is a physical cortical pattern, it can be rendered on the cortex and compared with known motor representation [39, 51], and the shared front-end is auditable. One can check, before trusting a decode, that the index-finger axis localizes over the expected sensorimotor territory rather than over an artifact channel, a form of validation a learned latent basis does not offer. The TCN decoder that reads the shared space is not itself interpretable; the interpretability claim is about the alignment front-end and the shared space, not the decoder.

The contrast with generic statistical alignment is the heart of the argument. Correlation Alignment and Euclidean Alignment take their axes from the eigenvectors of a covariance matrix, which are properties of the recording rather than of the task, and across non-corresponding grids those eigenvectors describe different cortical geometries [6, 7]. Behavior-supervised alignment instead defines its axes by the decode target. The axes match across subjects because finger

movement matches across subjects, even when nothing about the grids does.

Supplementary Figure S1 shows the alignment as a cortical map, the shared axis at work across two mismatched grids, and the decoder that reads it.

3.4. Shared TCN decoder

The decoder g is a dilated causal temporal convolutional network that maps the D -component, multi-band feature stream to the five-finger trajectory while preserving sequence length. Stacked convolution blocks use exponentially increasing dilations $1, 2, 4, \dots, 64$, so the receptive field spans the entire training segment: with two kernel-7 convolutions per residual block the theoretical field is about fifteen seconds at the 100 Hz feature rate, which exceeds the 512-frame (5.12s) training segment, so every output frame sees the full segment of context without downsampling, in the WaveNet and TCN tradition [42, 40, 41]. Each block is causal and length-preserving, and a U-Net-style skip structure links matching depths so that fine temporal detail survives to the output [43, 41]. The input at each time step is the D behavior-aligned components across the spectral bands of Section 3.2; the output at each time step is the five continuous flexion values. We use a hidden width of 128 channels.

The decoder is trained on the pooled source subjects. Because every source subject’s data has already been mapped through its own φ_i into the same D -component space, the pooled training set is consistent: a single g sees component f, r from every source subject as the same input dimension. The pooling is only possible because of the alignment; without it, the source recordings would be incommensurable and could not be concatenated into one training set. We train three random seeds and average their outputs, an ensemble that reduces seed-to-seed variance in the continuous predictions and stabilizes the per-subject correlations we report. The decoder carries no subject-specific parameters, which is what lets the same g serve a new subject once that subject’s φ_t is fitted. The decoder is deliberately ordinary: a causal dilated TCN is a standard sequence model, so the transferability is attributable to the alignment rather than to a bespoke decoder.

3.5. Transfer regimes

We evaluate two transfer regimes that differ in how much target information the decoder sees. Target labels are used for two distinct purposes in this work, and the regimes separate them. The first use is *alignment estimation*: the target filter φ_t is computed from the correlation between the target’s own high-gamma envelope and the target’s own measured flexion, so it consumes labeled target movement. The second use is *decoder fine-tuning*: gradient updates to the shared decoder g on a short labeled target segment. Both regimes below perform the first. Only the minimal-calibration regime performs the second. Consequently neither regime is label-free and neither is calibration-free, and we do not describe them as such anywhere in this paper.

In the *frozen-decoder* regime, the shared decoder g is taken as trained on the source subjects and applied to the target through the target-fitted alignment φ_t , with the decoder weights held frozen. The name refers to the decoder and to nothing else: the alignment still uses labeled target behavior to define the projection. It tests whether the source-trained decoder reads the shared component space well enough that a new subject’s aligned signal is decodable without any decoder adaptation.

In the *minimal-calibration* regime, the source-pretrained decoder is lightly fine-tuned on a short segment of target calibration data. Source pretraining here acts as a transferable prior in the sense of fine-tuning a learned initialization [20, 16]: the decoder starts from a population solution and is nudged toward the target with a small number of gradient steps on a few minutes of target movement. The two regimes bracket the deployment options. The frozen-decoder regime is the no-adaptation extreme and isolates how much the shared space alone carries; minimal-calibration is the cheap-calibration extreme, and it is the regime in which we make the headline accuracy claim. The contrast between them measures the value of the source prior, since both use the same target alignment and differ only in whether the decoder adapts. We quantify how little calibration suffices in Section 4 with a calibration-amount curve.

3.6. Evaluation

We evaluate by leave-one-subject-out (LOSO). Within each subject the recording is partitioned contiguously in time, preserving temporal order: the first 65% is training, 65–80% is validation, and the final 20% is the held-out test segment. Target calibration in the transfer regimes draws on a fraction of the 0–80% region only, and the test segment is never seen during fitting or model selection. For each held-out target the remaining subjects are the source set: the shared decoder is trained on them, the target is decoded under each regime, and all architecture and rank choices are fixed in advance or selected on the validation split, never on the test segment, so the reported numbers are not test-set maxima. The metric is the mean Pearson correlation r over the five fingers on the continuous predicted trajectory against the continuous flexion ground truth. We average and compare r without a Fisher z -transform; at these trajectory lengths the resulting bias is negligible. Electrode locations for the cortical maps use the MNI/Talairach coordinates provided with the Stanford fingerflex library [52, 53].

The library contains nine subjects. The seven-subject set used throughout excludes cc (63 electrodes) and wm (38 electrodes), which a “first 43 channels” truncation in the pipeline removed; the exclusion tracks no property of the data, and cc carries more electrodes than most of the retained subjects. Because the behavior-supervised alignment accepts any electrode count by construction, we report both that set and, for a matched-subject-count comparison against CORTEG’s nine subjects, the full nine-subject library.

Independent cohort. To test the method outside the fingerflex library we use the cursor-tracking recordings from the same public ECoG library [52]: four patients (fp, gf, rh, rr), none of whom appear among the nine fingerflex subjects, tracking a target with a joystick while a continuous two-dimensional cursor trajectory is recorded. The behavior and the grids differ from the main experiment while the acquisition does not, so a difference in the result is not a difference in acquisition. The method carries over unchanged except for the number of behavioral axes, which follows the target dimension: $D = 2R$ rather

than $5R$. Both transfer regimes are run leave-one-subject-out over the four patients against the same first- k -channel no-correspondence control used on fingerflex, and a within-subject decoder is fitted on the same cohort as its accuracy ceiling.

Protocol-matched external comparison. Comparing our mean r against CORTEG’s published mean r places the two numbers on different protocols. We therefore additionally run our own decoder under CORTEG’s protocol as published [8]: all nine subjects rather than our seven, their chronological data division, taking the first 400s of any recording of at least 600s and otherwise the first two thirds with the remainder as test and a validation slice from the tail of train, and their evaluation rule, the mean over the five fingers of the Pearson correlation on their test window. We also rebuild their input pipeline and fit a plain ridge regression on it, which anchors the comparison: if our reconstruction reproduces their own published linear baseline, the data is theirs in every respect that matters.

Alignment heads. To measure what the fixed correlation filter costs, and where the transfer gain comes from, we compare four alignment heads with decoder, features, regimes and seeds held identical and only the head varying. *Fixed* is the published correlation filter, frozen. *Learned* uses it as an initialization and trains it jointly with the decoder. *Nonlinear* adds a residual per-component band-mixing multilayer perceptron of grouped 1×1 convolutions whose second layer is zero-initialized, so training begins exactly at the linear solution and any gain is attributable to the added capacity. *Random* is the learned head from a random orthonormal initialization with no behavioral supervision: it has the capacity of the learned head and the same training, and differs only in that its axes are not defined by behavior.

Component correspondence and coverage. Two further measurements support the claims about what the shared components are, both computed on held-out data and both detailed in the supplementary material. The first characterizes each component functionally, by its per-band correlation with each flexion trace, and anatomically, by the filter-weighted MNI centroid of its elec-

trodes, then compares components across subjects against the label-permuted null (Supplementary Section S6). The second audits the coverage precondition on a second, twelve-subject ECoG cohort recorded during clinical epilepsy monitoring rather than for motor research [54], counting per subject how many electrodes carry a reach-locked high-gamma response (Supplementary Section S7).

3.7. Implementation and reproducibility

The closed form of the rank-1 filter and of the $R - 1$ singular directions, together with the full list of feature, referencing, decoder, optimizer, ensembling and data-split settings and the compute used, is given in Supplementary Section S1 and Supplementary Table S1.

4. Results

There are two transfer regimes, frozen-decoder and minimal-calibration; within-subject decoding is a sanity reference, not a transfer setting, and we report it to show the alignment does not throw away single-subject accuracy. The same alignment and the same decoder architecture run in every setting; only the amount of target supervision changes. Both regimes fit the target alignment from labeled target movement, and only minimal-calibration additionally updates the decoder. Every mean and spread quoted below is stated in the text; Supplementary Table S2 gives the same numbers broken out per subject.

4.1. Decoding accuracy and calibration cost

Trained on a single subject from scratch, our model reaches a within-subject mean r of 0.515 ± 0.088 , with the best-covered subject (zt) at 0.679 and a representative mid-range subject (mv) at 0.523 (Supplementary Table S2). This is below the within-subject convolutional state of the art on this task, FingerFlex [38], which reports roughly 0.6–0.65; our within-subject figure is a sanity reference computed under our own protocol, not a re-run of that method. It rules out one alternative reading of the transfer results below: a method could buy

transferability with an input representation so generic that it decodes everyone equally but poorly, and this figure shows the behavior-aligned representation is not that, since it keeps enough subject-specific signal to decode a single subject well.

Applying the source-trained decoder directly through the target alignment, with no decoder fine-tuning, gives a cross-subject mean r of 0.355 ± 0.097 (Supplementary Table S2). The spread across subjects is informative: *zt* reaches 0.528 and *mv* 0.426 frozen-decoder, while subjects with poorer cortical coverage of the relevant motor territory (*ht* at 0.221) transfer less well. The pattern tracks coverage: subjects whose grids sit well over hand sensorimotor cortex hand the shared decoder a clean signal, while a grid that grazes the territory gives the alignment less to work with. That a decoder whose weights have never been updated on this subject already decodes well above the permutation null is direct evidence that the behavior-aligned component space is genuinely shared. A label-permutation null, decoding against shuffled finger trajectories, gives $r \approx 0$; the 0.042 no-alignment entry in Table 3 is not that null but a broken-alignment decoder, a real decoder fed an incoherent correspondence.

The headline result is the minimal-calibration regime. With the source-pretrained decoder lightly fine-tuned on a short target segment, the cross-subject mean r rises to 0.566 ± 0.084 (Supplementary Table S2). This is competitive with the CORTEG foundation model’s reported cross-subject finger-decoding mean of 0.554 [8], with no large-scale pretraining corpus and no opaque backbone. To quantify how much of the 0.012 difference is signal, we ran the seven-subject LOSO across twelve single seeds: the single-seed mean is 0.532 ± 0.013 (SD), so the seed-to-seed spread alone (0.013) is larger than the 0.012 gap to CORTEG. We therefore read the two as reaching the same accuracy band, not as a win either way. Five of the seven subjects exceed 0.55 (*zt* 0.686, *mv* 0.613, *jc* 0.595, *jp* 0.571, *bp* 0.556). The jump from frozen-decoder 0.355 to fine-tuned 0.566 shows that the short calibration is not cosmetic: it lets the decoder adapt to the residual per-subject structure the shared space leaves on the table, recovering most of the gap to within-subject accuracy. Section 4.2 puts this number on

the reference method’s own protocol.

Reporting the full nine-subject library, on a matched subject count to CORTEG, tells the same story. The fine-tuned mean over all nine is 0.550 (the seven-subset within the same run is 0.573, consistent with the 0.566 headline), with frozen-decoder 0.336 and within-subject 0.499. The two subjects excluded by the code artifact are the best and the worst in the library: cc, the best-covered subject at 0.750 fine-tuned, and wm, a low-channel subject at 0.190; the remaining seven fall in between (zt 0.699, jc 0.617, mv 0.598, bp 0.589, wc 0.543, jp 0.530, ht 0.436). The $0.566 \rightarrow 0.550$ change from seven to nine subjects is within seed noise and is driven almost entirely by the low-channel wm; that cc, the best subject, was one of the two dropped confirms the exclusion had no bearing on the headline.

Source pretraining acts as a transferable prior. The fine-tuned cross-subject decoder meets or exceeds the from-scratch within-subject decoder for six of the seven subjects, with the seventh (ht) marginally below (Supplementary Table S2: 0.556 vs 0.452 for bp, 0.686 vs 0.679 for zt, 0.571 vs 0.494 for jp, 0.412 vs 0.433 for ht, 0.613 vs 0.523 for mv, 0.530 vs 0.445 for wc, 0.595 vs 0.577 for jc). The mean gain is +0.05, a trend at this sample size (Wilcoxon signed-rank $p = 0.078$); we do not claim a uniform per-subject win. The effect is largest for the mid-range subjects (bp gains 0.104, mv gains 0.090, wc gains 0.085, jp gains 0.077), the cases where a single subject’s data leave the most room for a population prior to fill, consistent with the view of pretraining as a transferable initialization [16, 20]. Varying the number of source subjects used to train the shared decoder, over resampled subsets, the fine-tuned cross-subject mean r moves from 0.476 ± 0.08 with a single source subject to 0.494 ± 0.09 with two, 0.529 ± 0.08 with four, and 0.525 ± 0.08 with six (single-seed, so below the ensembled headline; the shape is the point). The prior helps and then saturates around four, where the error bars for four and six overlap, and even a single source subject already transfers well.

Figure 2 plots cross-subject r against the amount of target calibration. Accuracy reaches $r \approx 0.54$ at about four minutes and the CORTEG accuracy band

(0.554) by about 4.8 minutes, so we report the operating point as four to five minutes; this is roughly 80–90% of the full 5.6-minute calibration set, not a small fraction of it. CORTEG reports comparable accuracy after ten to thirty minutes of calibration [8], though that figure is from a different protocol, so the calibration-time comparison is across protocols rather than head-to-head. Accuracy climbs steeply in the first few minutes and then flattens, so most of the achievable transfer is captured very early. The source prior is what makes this possible: the decoder starts from a population solution, so target calibration has only to adapt rather than to learn the task.

Two settings questions are worth one number each here, with the full sweeps in the supplementary material. On hyperparameters, a thirteen-configuration sweep (Supplementary Section S4, Table S3 and Figure S2) puts $R = 4$ at a genuine optimum: minimal-calibration r runs 0.530, 0.539, 0.557, 0.569, 0.561 and 0.554 across $R = 1, 2, 3, 4, 6, 8$. Note that R and D are not independent knobs in this construction, since $D = 5R$ by definition, one bundle of R filters per finger, so sweeping R sweeps D ; and the dilation range is flat between 32 and 64 (0.5708 and 0.5687, a gap of 0.002 against a seed-to-seed SD of 0.013 over twelve seeds), so we report it as flat and claim no optimum there. On computational cost (Supplementary Section S5, Table S4), the deployed model has 1,673,221 parameters, costs 0.334 GFLOP per second of signal in the decoder, runs at 1.71 ms on GPU and 8.03 ms on CPU per second-long window with a 7.3 MiB peak GPU footprint, and calibration fine-tuning costs 1.5 to 4.4 s of wall clock across the seven target subjects; the honest caveat is that the decoder is not what dominates deployment cost, since the filterbank front end costs 339.98 ms per second of signal, two orders of magnitude more, while the alignment projection itself costs 0.41 ms per second.

4.2. Comparison with existing methods under a matched protocol

The comparison above is across protocols: our leave-one-subject-out split on seven subjects against CORTEG’s published figure on nine under their own division. Table 2 removes that difference for the rows that can be matched. Run

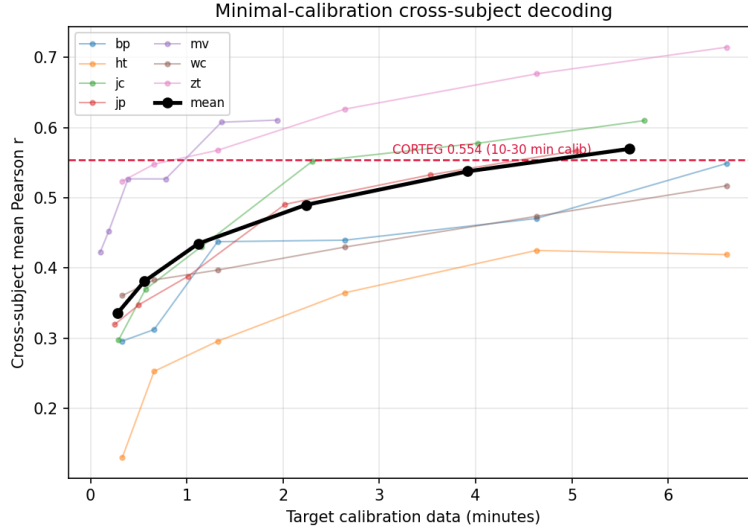


Figure 2: Cross-subject decoding accuracy as a function of target calibration time. The mean reaches the CORTEG-level accuracy band (around 0.55) at about four to five minutes of calibration ($r \approx 0.54$ at four minutes, reaching about 0.554 near 4.8 minutes), whereas the CORTEG foundation model reaches comparable accuracy after ten to thirty minutes under a different protocol [8]. The plotted means are 0.336, 0.381, 0.435, 0.490, 0.538 and 0.570 at 0.3, 0.6, 1.1, 2.2, 3.9 and 5.6 minutes; the full calibration set is 5.6 minutes.

under CORTEG’s protocol, on all nine subjects with their chronological data division and their evaluation rule, our decoder reaches a mean r of 0.530 (SD 0.157) against their published 0.554. CORTEG is ahead by 0.024, well inside the 0.157 across-subject spread of our own arm, so the protocol-matched result reproduces the reading this paper already gave: the two occupy one accuracy band. We make no claim to beat CORTEG.

The ridge rows are what make this a comparison rather than an assertion. A plain ridge regression fitted on the windows we rebuilt from their input pipeline reaches 0.380 on their split, against the 0.336 their paper reports for the same baseline, so our reconstruction of their inputs reproduces their own linear baseline and the data on which the two methods are compared is theirs in every respect that matters. The final two rows of Table 2 are marked as cross-paper

Table 2: Comparison with existing methods. The protocol column states how each number was obtained: the first four rows are all under CORTEG’s published protocol on the same nine subjects with the same chronological data division and the same evaluation rule, so they are directly comparable; the last two are cited across papers and are given for orientation only.

Method	Protocol of the reported number	n	Mean r
CORTEG [8], as published	their protocol, cross-subject with calibration	9	0.554
Ours, run under their protocol	their protocol, cross-subject with minimal calibration	9	0.530 (SD 0.157)
Ridge on their inputs, our rebuild	their protocol, their input pipeline rebuilt by us	9	0.380
Ridge [8], as published	their protocol, their own linear baseline	9	0.336
Ours, our protocol	our leave-one-subject-out split, minimal calibration; <i>cross-paper</i>	7	0.566 (SD 0.084)
FingerFlex [38], as published	within-subject, not cross-subject; <i>cross-paper</i>	9	≈ 0.60 – 0.65

rather than protocol-matched: our seven-subject headline and the published within-subject FingerFlex figure are given for orientation, and neither is a controlled comparison.

4.3. The alignment is what crosses the grid gap

To isolate the alignment’s contribution we run a same-protocol ablation in which only the cross-subject correspondence scheme changes; the decoder, features, regimes, and seeds are held fixed (Table 3). Behavior-aligned components give frozen-decoder $r = 0.371$ and fine-tuned $r = 0.557$. (This “ours” row is the

same method as Supplementary Table S2 at matched seeds; the 0.557 versus 0.566 difference is seed variance across the three-seed ensembles.) Replacing the alignment with a naive common channel set, the first 43 channels of every grid with no alignment, collapses frozen-decoder transfer to $r = 0.042$ and caps fine-tuned transfer at 0.339. Adding Euclidean Alignment on top of that forced common channel set does not help: frozen-decoder $r = 0.037$, fine-tuned $r = 0.357$ [7]. One might object that no-alignment is a strawman and that a sensible anatomical alignment would do better. It does not. A principled coordinate-based common grid, built by placing k -means anchors on the real electrode positions and interpolating each grid onto them with a Gaussian kernel, with a 20-anchor grid matched to our component count, also collapses (frozen-decoder 0.071 ± 0.055 , fine-tuned 0.273 ± 0.132). A coherent anatomical correspondence is therefore not enough; it is behavioral supervision specifically, not alignment in general, that crosses the grid gap. Generic statistical alignment, the standard cross-subject tool, cannot bridge non-corresponding grids, because the channel set it operates on is not shared in any meaningful sense. The behavior-versus-Euclidean-Alignment contrast is significant in both regimes (Wilcoxon signed-rank over the seven held-out subjects, $p = 0.016$ frozen-decoder and fine-tuned alike). The behavior-supervised alignment, which defines its axes by the task rather than by the recording, is what makes cross-subject transfer possible. The decoder is the same in all three rows; the alignment is the difference between near-chance and foundation-model-level transfer.

One objection remains: perhaps any supervised front-end, not the *shared* behavioral axis specifically, is what helps. A further control separates the two. We keep the same behavior-supervised filter but permute the target’s twenty component axes, which preserves the supervised denoising exactly while destroying the cross-subject correspondence. In a seed-matched three-seed pair, permuting the axes drops frozen-decoder transfer from 0.358 to 0.162 and fine-tuned transfer from 0.570 to 0.525 (Table 3; Wilcoxon signed-rank $p = 0.016$ in both regimes, all seven subjects favoring the shared axes). The reading is nuanced. The permuted front-end still beats the unsupervised schemes (0.162

frozen-decoder versus about 0.04–0.07 for the first-43, Euclidean-Alignment, and anatomical rows), so supervised denoising does contribute. But shared behavioral correspondence adds a further significant increment on top, and it is largest in the frozen-decoder regime, where the decoder cannot relabel the permuted axes; fine-tuning recovers part of the permutation, as expected. Shared behavioral definition, not supervised denoising alone, is what drives frozen-decoder transfer. With seven subjects the Wilcoxon signed-rank test bottoms out at $p = 0.016$ (all seven favoring the reference); these tests are uncorrected and underpowered, so we read them as consistent directional evidence rather than as precise p -values.

The permutation control removes one of three candidate explanations for the transfer. The alignment heads in the lower block of Table 3 remove the other two. All four heads share the decoder, the features, the regimes and the seeds, and differ only in how the per-subject projection is obtained. A randomly initialized head, with the capacity of the learned head and the same training but no behavioral supervision at all, drops minimal-calibration transfer from 0.5746 to 0.3194 ($p = 0.0156$, all seven subjects favoring supervision) and frozen-decoder transfer to 0.2603, below even the fixed filter. Capacity is held constant, so the gain is not the trainable capacity of the alignment; the decoder and the fine-tuning protocol are identical across the four heads, so it is not those. With the permutation control removing the remaining candidate, supervised denoising, what is left is the cross-subject behavioral correspondence.

The same experiment answers a separate question about the fixed filter, and the answer is a qualified concession. Once the decoder is free to fine-tune, the three supervised heads are indistinguishable: 0.5746 fixed, 0.5775 learned, 0.5691 nonlinear, none of them separable (Wilcoxon $p = 1.00$ and $p = 0.58$). In the frozen-decoder regime, where only the per-subject head adapts, the fixed filter does cap capacity: a learned linear head lifts 0.3529 to 0.4648, better on 6 of 7 subjects ($p = 0.031$), and a nonlinear head to 0.5173, better on 7 of 7 ($p = 0.0156$). The fixed filter is a sound choice when calibration is available and a real constraint when it is not, and Section 6 takes up what follows from that.

A last question is whether components carrying the same finger label in different subjects are the same kind of object (Supplementary Table S5). Given a component in one subject, the nearest component in another subject carries the same finger in 0.3036 of 840 cross-subject queries, against 0.1976 for the label-permuted null and 0.200 for chance; the adjusted Rand index of finger identity against k -means on the pooled signatures is 0.072 against a null of -0.0083 . The absolute values are modest 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 purely subject-specific supervised projection would give, which is what the shared axes require. Anatomy tells the complementary half of the story: the mean MNI centroid distance between same-finger components in different subjects is 35.54 mm against 35.51 mm for different-finger components, a gap of -0.031 mm with permutation $p = 0.53$. Same-finger components are anatomically no closer than different-finger ones, so the correspondence this method builds is functional rather than anatomical; Section 5.1 takes that up. Full clustering detail is in Supplementary Section S6.

4.4. Replication on an independent cohort

The results above are all from one dataset. To test whether they are a property of the method or of that library, we repeated the experiment on an independent cohort: four patients performing continuous two-dimensional cursor tracking, none of whom appear among the nine fingerflex subjects (Table 4). A within-subject decoder reaches $r = 0.316$ on this cohort, which is the ceiling the transfer arms are read against.

The central pattern replicates. In the frozen-decoder regime, behavior-aligned transfer reaches $r = 0.2427$ against $r = 0.0250$ for the no-correspondence control, a difference of $+0.2177$, with the aligned arm ahead on 3 of 4 subjects. With minimal calibration the figures are 0.4191 against 0.2936, a difference of $+0.1255$, again ahead on 3 of 4. The correspondence is clearest when each cohort is read against its own within-subject ceiling: frozen-decoder transfer recovers

Table 3: Same-protocol ablation isolating the behavioral alignment. Mean cross-subject r ; only the cross-subject correspondence scheme changes in the upper block, and only the alignment head in the lower block.

Correspondence scheme	Frozen-decoder	Fine-tuned (min.-calib.)
Behavior-aligned components (ours)	0.371	0.557
Behavior-aligned, axes permuted [†]	0.162	0.525
First-43 channels, no alignment	0.042	0.339
First-43 + Euclidean Alignment [7]	0.037	0.357
Anatomical-coordinate alignment	0.071	0.273
<i>Alignment head[‡]</i>		
Fixed correlation filter (ours, as published)	0.3529	0.5746
Learned linear head	0.4648	0.5775
Nonlinear head	0.5173	0.5691
Random initialization, no behavioral supervision	0.2603	0.3194

[†] Supervised-but-not-shared control: same behavior-supervised filter with the target’s twenty axes permuted. Paired against a seed-matched shared run (0.358 frozen-decoder / 0.570 fine-tuned); Wilcoxon signed-rank $p = 0.016$ in both regimes.

[‡] Four alignment heads with decoder, features, regimes and seeds held identical; this block is a separate seed set from the upper block, so compare within a block. The random-initialization row has the capacity of the learned head and the same training, and differs only in that its axes are not defined by behavior.

76.7% of that ceiling with behavior-supervised alignment and 7.9% without it, where the same two figures on the fingerflex cohort are 72.1% and 8.2%. An independent cohort, a different behavior and a different set of grids, and the alignment carries the same fraction of the achievable signal while the control collapses to the same near-zero.

Two limits apply to this replication. With four subjects a paired Wilcoxon signed-rank test cannot fall below $p = 0.125$ whatever the data, so this is a replication of the size and shape of the effect and we make no significance claim on it. And subject rh is a partial exception, slightly behind the control in both regimes (0.1379 against 0.1495 frozen-decoder, 0.1533 against 0.3342 fine-tuned);

Table 4: Replication on an independent cohort: four cursor-tracking patients (fp, gf, rh, rr), none of them fingerflex subjects, with a two-dimensional continuous target rather than five-finger flexion. Behavior-aligned transfer against the first- k -channel no-correspondence control, mean r over the two cursor dimensions.

Regime	Behavior-aligned	No corresp.	Aligned ahead
Frozen decoder	0.2427	0.0250	3 of 4
Minimal calibration	0.4191	0.2936	3 of 4
Within-subject (gate)	0.316		
Read against each cohort’s own within-subject ceiling, frozen-decoder transfer recovers 76.7% of the ceiling with alignment and 7.9% without, against 72.1% and 8.2% on fingerflex. With $n = 4$ a paired Wilcoxon test cannot reach below $p = 0.125$, so no significance claim is made here. Subject rh runs slightly behind the control in both regimes.			

the effect is a majority pattern on this cohort, not a uniform one.

5. Discussion

5.1. Electrode, anatomical, and functional correspondence

Three notions of correspondence are easy to conflate, and this method engages only one of them. *Electrode-level correspondence* is a bijection between the channels of two subjects, so that channel k names the same thing in both. Our grids do not have it and the method does not construct it. *Anatomical correspondence* maps both subjects into a common cortical coordinate frame and treats nearby coordinates as equivalent. The method does not build that either, and the ablation shows that building it is not sufficient: a principled interpolation onto a common anchor grid collapses to frozen-decoder $r = 0.071$. *Functional correspondence* is what remains and what the method constructs. Each subject’s grid is mapped onto axes defined by the behavior, and two components count as the same axis when they were built for the same finger, whatever electrodes realize them.

The distinction is measured rather than asserted. Same-finger components in different subjects are anatomically no closer than different-finger ones, 35.54

against 35.51 mm with permutation $p = 0.53$ (Supplementary Table S5). Had the method been performing a covert anatomical alignment, that gap would have been large and negative. It is not, which is also why anatomical interpolation fails while the behavior-defined axes work: what makes the axes comparable is not where the electrodes sit.

The assumption the method does require is that components built for the same finger in different subjects are comparable, and the evidence for it is the cross-subject retrieval figure of 0.3036 against a permuted-label null of 0.1976 and chance of 0.200, with an adjusted Rand index of 0.072 against -0.0083 . That is evidence of shared functional structure. It is not evidence of anatomical homology, and we stop short of that claim, which our own MNI result does not support. What the components share is a relationship to the same movement, which is what the construction supervises for and what the shared decoder consumes. Section 6 states the condition under which even that much fails.

5.2. What the shared maps show

Each behavior-aligned map is a univariate channel-behavior correlation pattern, an activation pattern in the sense of Haufe et al. [12], so it can be read straight off the cortex (Figure 3). A clinician can look at the map for a given finger and see which cortical sites drive its decode, a reading a foundation model’s internal representations do not offer [8, 11]. The somatotopic ordering was not built in. It emerges because the behavior correlation that defines each map independently lands on the cortical territory that physiology assigns to that finger [13], so the shared axes are the physiologically expected cortical correlates of the movements rather than arbitrary directions that happen to decode.

5.3. Deployment cost and relation to shared-latent methods

The practical consequence is a lower deployment bar. Foundation models reach their accuracy by pretraining on large neural corpora that are expensive to assemble and often unavailable outside a few labs [8]. For continuous clinical decoders, where each new patient is a fresh calibration session and where decoders

historically required substantial within-subject training [1, 2], the calibration budget of Section 4.1 is the lever that matters. The smaller resource footprint also has a methodological upside: a six-subject pipeline can be retrained and audited end to end on a single machine, so the alignment, the source pool, and the decoder are all inspectable, which keeps the system accountable in a way a large pretrained backbone is not. One caveat belongs here. The filterbank front end dominates the pipeline at 339.98 ms per second of signal, against 1.71 ms for a decoder forward pass and 0.41 ms for the alignment projection (Supplementary Section S5), so a deployment comparison against a large pretrained model should be read on total pipeline cost rather than on decoder size.

Relative to the shared-private and shared-latent literature (Table 1), the approach collapses two stages those methods keep separate: where they first learn a shared space and then learn to decode from it, we let the decode target define the shared space in one step, so the space cannot drift away from what is decodable [27, 28, 30]. The shared axes are then tied to a specific behavior and do not transfer to an unrelated task, but within the task they are correct by construction and require no adversarial or reconstructive training to discover.

6. Limitations and Future Work

We report both the seven-subject set (0.566) and the full nine-subject library (0.550), the latter on a matched subject count to CORTEG’s nine. Both bracket CORTEG’s 0.554, and the 0.012 gap is smaller than the decoder’s seed-to-seed spread (0.013), so the comparison should be read as reaching the same accuracy band rather than as a beat in either direction. Section 4.2 removes the protocol difference by running our decoder under theirs, and the reading does not change.

The external comparison is made against CORTEG’s published figure under their own protocol, and the ridge baseline reported alongside it is what makes that comparison meaningful: a plain ridge regression on the inputs we rebuilt from their pipeline reaches 0.380 against the 0.336 their paper reports for the same baseline, so our reconstruction of their inputs reproduces their own

published baseline.

The method assumes producible target behavior and cross-subject axis homology, and both bound where it applies. The alignment φ_t is fit from labeled target movement, so a target who cannot produce the calibration movements has no way to define the shared axes; our claims are scoped to able-bodied cross-subject transfer, as in the epilepsy-monitoring cohort studied here, not to patients who have lost the movement. The deeper assumption is that flexing a given finger engages homologous cortex across subjects, which is what makes a behavior-defined axis mean the same thing in two brains. Cortical reorganization can break that homology: after a stroke or amputation, the cortical representation of a movement can shift or reweight, so the axis that a source population defines may no longer coincide with the target’s, and the transfer premise weakens.

Two consequences follow, and the alignment-head experiment makes both concrete. First, the alignment we publish is linear and fixed, a per-channel correlation map, and a linear map cannot express a reorganization that redistributes a movement’s representation nonlinearly across the grid. Second, a map estimated once cannot adapt to the target beyond that estimate. The frozen-decoder column of Table 3 measures what this costs even in an unreorganized cohort: a nonlinear head re-estimated on the target lifts transfer from 0.3529 to 0.5173, better on 7 of 7 subjects. The concrete improvement scheme is therefore not speculative. Replace the fixed correlation filter with a nonlinear head re-estimated on the target rather than reused from a source definition, keep the behavioral supervision so that the axes retain their meaning, and hold the shared decoder frozen so that the burden of accommodating the target’s cortex falls entirely on the head that has the capacity for it. Our measurement shows that capacity is usable in precisely the regime where reorganization would demand it. What it does not show, and what remains open, is that the scheme carries to genuinely reorganized cortex, which requires a cohort containing such patients.

The method also carries a coverage precondition that bounds where the

construction can be applied at all: a behavior-defined axis can be estimated only if some electrode observes the cortex that the calibration behavior drives. We measured this on a second, twelve-subject ECoG cohort recorded during clinical epilepsy monitoring rather than for motor research (Supplementary Section S7). Wrist speed rises after the annotated reach onsets in 11 of 12 subjects, so the behavior is present and the recordings are time-aligned; but only 6 of 12 subjects have any electrode with a reach-locked high-gamma response above threshold, three of them strongly (peak z of 10.8 on 7 electrodes, 20.7 on 5, and 4.6 on 14) and three marginally, at one to three electrodes. In a cohort implanted for clinical rather than motor-research reasons, the precondition therefore holds for about half the subjects. It is a scope condition of the method, and it should be checked by exactly this measurement before the method is applied to a new cohort.

On generality, we now report two ECoG cohorts, the Stanford fingerflex library [52] and the independent cursor-tracking cohort of Section 4.4, so the evidence is no longer from a single dataset. What it is not is broad: the replication rests on four subjects and one alternative behavior, it is a replication of the size and shape of the effect rather than a significance claim, and one of the four subjects goes the other way. The behavior-supervised alignment principle is dataset-independent by design, so extending it to additional ECoG datasets and to other motor tasks remains the natural way to demonstrate generality.

The within-subject FingerFlex comparison is cited from the original paper [38] rather than re-run under our exact protocol, and Table 2 marks it as cross-paper for that reason. The same-protocol ablation supplies the controlled within-pipeline comparison that anchors the cross-subject claims, so the headline does not rest on cross-paper number matching.

The behavior-supervised alignment is not specific to ECoG: any modality with a decodable behavioral target, including EEG, MEG, and intracortical signals, admits the same construction [47, 30, 29], and scaling beyond five degrees of freedom to richer hand kinematics is a direct generalization, since a larger behavioral target set yields a higher-dimensional shared space defined the same

way.

The alignment heads are a first step into learning the alignment rather than computing it, and they show the direction is live: a nonlinear behavior-supervised head is the strongest frozen-decoder front end we measured, and it costs the interpretability of the rank-1 map, which the fixed filter retains. The open question is how to keep a jointly trained head anchored to the behavioral axes across many subjects, because the property that makes the whole scheme work is that the shared axes are defined by the task. The principle to preserve through all of these extensions is the one the ablation isolates: the shared space must be supervised by behavior, because that is what makes it common across grids that share nothing else.

7. Conclusion

Casting cross-subject neural decoding as transfer across non-corresponding input spaces clarifies why the usual tools fail and points to what works. When two brains are recorded by sensor arrays that share no feature index and no geometry, statistical alignment has nothing common to equalize, and our ablation shows it stays near chance. Behavior-supervised alignment manufactures the missing correspondence: it defines the shared latent axes by the decode target itself, so the axes mean the same thing across subjects whose grids do not. On this shared space a single TCN decoder, trained once on the source subjects, transfers to a new subject with minimal calibration, reaching a cross-subject mean r of 0.566 on seven subjects and 0.550 across the full nine, comparable to the CORTEG foundation model’s 0.554 (a gap smaller than the method’s own seed-to-seed spread), at about four to five minutes of calibration instead of ten to thirty, and with cortical maps one can read against known somatotopy. Under CORTEG’s own protocol on all nine subjects the same decoder reaches 0.530, so the comparison survives being made on one protocol; the pattern replicates on an independent four-patient cohort with a different behavior; and a randomly initialized alignment head of matched capacity, trained identically, falls to 0.319,

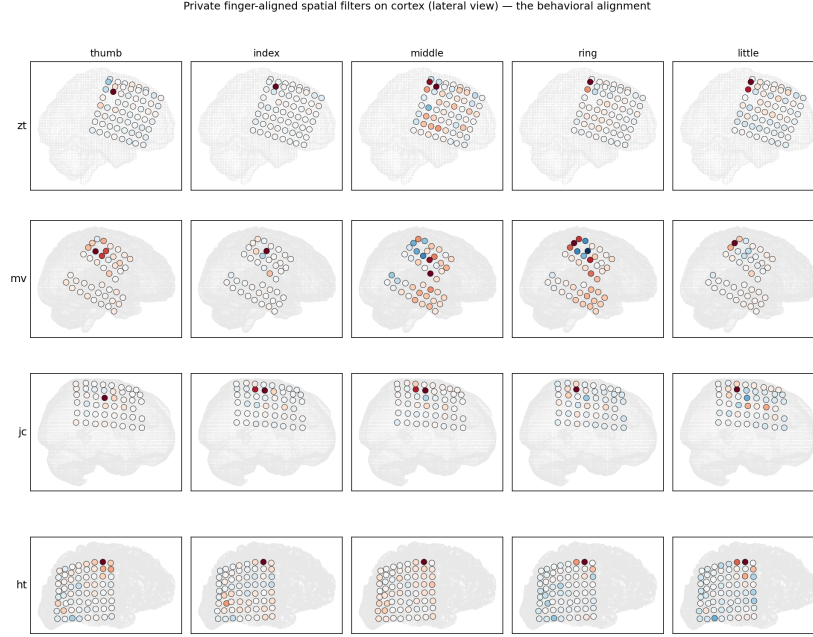


Figure 3: The rank-1 behavior-aligned forward maps (one per finger) rendered on the cortical surface (lateral view). These leading maps localize focally over sensorimotor cortex and follow a gross finger-ordered somatotopic organization, recovering known motor organization from the decoding objective alone. This auditable front-end stands in contrast to the opaque internal representations of foundation-model decoders.

which places the gain on the behavioral correspondence itself. The decoder is ordinary; the alignment is the contribution. It is the alignment, not the decoder, that crosses the grid gap.

Data Availability Statement

All three electrocorticography cohorts analysed in this study are publicly available and none required a data-use agreement. The fingerflex recordings are the nine-patient set of the Miller library of human electrocorticographic data [52], distributed by the Stanford Digital Repository under PURL [zk881ps0522](https://purl.org/ps0522) (CC BY-SA) as a single **fingerflex** archive with one MATLAB file per patient. The independent cohort of Section 4.4 is the cursor-tracking half of the same

library and the same PURL, tasks `mouse_track` and `joystick_track`, from which we use the four patients fp, gf, rh and rr; each task is distributed as its own archive under that PURL. The twelve-patient clinical cohort used for the coverage audit (Supplementary Section S7) is AJILE12 [54], available from the DANDI archive as dandiset 000055, from which we use one session per patient. All code for the alignment, the shared decoder, and the reported experiments will be deposited in a public repository (`behavior-aligned-ecog-transfer`) with an archival DOI at acceptance.

CRedit Author Contributions

Ward Ceysens: Methodology, Software, Investigation, Writing – original draft. **Axel Faes:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this work the authors used a large language model to assist with drafting and editing. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- [1] F. R. Willett, D. T. Avansino, L. R. Hochberg, J. M. Henderson, K. V. Shenoy, High-performance brain-to-text communication via handwriting, *Nature* 593 (2021) 249–254. doi:10.1038/s41586-021-03506-9.
- [2] V. Gilja, P. Nuyujukian, C. A. Chestek, J. P. Cunningham, B. M. Yu, J. M. Fan, M. M. Churchland, M. T. Kaufman, J. C. Kao, S. I. Ryu, K. V. Shenoy, A high-performance neural prosthesis enabled by control algorithm design, *Nature Neuroscience* 15 (12) (2012) 1752–1757. doi:10.1038/nn.3265.
- [3] S. J. Pan, Q. Yang, A survey on transfer learning, *IEEE Transactions on Knowledge and Data Engineering* 22 (10) (2010) 1345–1359. doi:10.1109/TKDE.2009.191.
- [4] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, *Proceedings of the IEEE* 109 (1) (2021) 43–76. doi:10.1109/JPROC.2020.3004555.
- [5] G. Wilson, D. J. Cook, A survey of unsupervised deep domain adaptation, *ACM Transactions on Intelligent Systems and Technology* 11 (5) (2020) 51. doi:10.1145/3400066.
- [6] B. Sun, J. Feng, K. Saenko, Return of frustratingly easy domain adaptation, in: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 2058–2065. arXiv:1511.05547.
- [7] 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.
- [8] 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.

- [9] G. Chau, et al., Population transformer: Learning population-level representations of neural activity, in: International Conference on Learning Representations (ICLR), 2025. arXiv:2406.03044.
- [10] D. Zhang, Z. Yuan, Y. Yang, J. Chen, J. Wang, Y. Li, Brant: Foundation model for intracranial neural signal, in: Advances in Neural Information Processing Systems 36 (NeurIPS), 2023.
- [11] W. Jiang, L. Zhao, B.-L. Lu, Large brain model for learning generic representations with tremendous EEG data in BCI, in: International Conference on Learning Representations (ICLR), 2024. arXiv:2405.18765.
- [12] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, F. Bießmann, On the interpretation of weight vectors of linear models in multivariate neuroimaging, *NeuroImage* 87 (2014) 96–110. doi:10.1016/j.neuroimage.2013.10.067.
- [13] K. J. Miller, S. Zanos, E. E. Fetz, M. den Nijs, J. G. Ojemann, Decoupling the cortical power spectrum reveals real-time representation of individual finger movements in humans, *Journal of Neuroscience* 29 (10) (2009) 3132–3137. doi:10.1523/JNEUROSCI.5506-08.2009.
- [14] C. Wang, S. Mahadevan, Heterogeneous domain adaptation using manifold alignment, in: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011), 2011, pp. 1541–1546. doi:10.5555/2283516.2283652.
- [15] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (8) (2013) 1798–1828. doi:10.1109/TPAMI.2013.50.
- [16] M. Long, Y. Cao, J. Wang, M. I. Jordan, Learning transferable features with deep adaptation networks, in: Proceedings of the 32nd International Conference on Machine Learning (ICML), Vol. 37 of Proceedings of Machine Learning Research, 2015, pp. 97–105. arXiv:1502.02791.

- [17] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, Adversarial discriminative domain adaptation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2962–2971. doi:10.1109/CVPR.2017.316.
- [18] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, V. Lempitsky, Domain-adversarial training of neural networks, *Journal of Machine Learning Research* 17 (59) (2016) 1–35. arXiv:1505.07818.
- [19] Y. Ganin, V. Lempitsky, Unsupervised domain adaptation by backpropagation, in: Proceedings of the 32nd International Conference on Machine Learning (ICML), Vol. 37 of Proceedings of Machine Learning Research, 2015, pp. 1180–1189. arXiv:1409.7495.
- [20] C. Finn, P. Abbeel, S. Levine, Model-agnostic meta-learning for fast adaptation of deep networks, in: Proceedings of the 34th International Conference on Machine Learning (ICML), Vol. 70 of Proceedings of Machine Learning Research, 2017, pp. 1126–1135. arXiv:1703.03400.
- [21] A. van den Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding (2018). arXiv:1807.03748.
- [22] B. Sun, K. Saenko, Deep CORAL: Correlation alignment for deep domain adaptation, in: Computer Vision – ECCV 2016 Workshops, Vol. 9915 of Lecture Notes in Computer Science, Springer, 2016, pp. 443–450. arXiv:1607.01719, doi:10.1007/978-3-319-49409-8_35.
- [23] P. Zanini, M. Congedo, C. Jutten, S. Said, Y. Berthoumieu, Transfer learning: A Riemannian geometry framework with applications to brain-computer interfaces, *IEEE Transactions on Biomedical Engineering* 65 (5) (2018) 1107–1116. doi:10.1109/TBME.2017.2742541.
- [24] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Multiclass brain-computer interface classification by Riemannian geometry, *IEEE*

Transactions on Biomedical Engineering 59 (4) (2012) 920–928.
doi:10.1109/TBME.2011.2172210.

- [25] A. Barachant, S. Bonnet, M. Congedo, C. Jutten, Classification of covariance matrices using a Riemannian-based kernel for BCI applications, *Neurocomputing* 112 (2013) 172–178. doi:10.1016/j.neucom.2012.12.039.
- [26] A. de Cheveigné, G. M. Di Liberto, D. Arzounian, D. D. E. Wong, J. Hjortkjær, S. A. Fuglsang, L. C. Parra, Multiway canonical correlation analysis of brain data, *NeuroImage* 186 (2019) 728–740. doi:10.1016/j.neuroimage.2018.11.026.
- [27] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, D. Erhan, Domain separation networks, in: *Advances in Neural Information Processing Systems* 29 (NeurIPS), 2016, pp. 343–351. arXiv:1608.06019.
- [28] C. Jing, et al., PTSM: Physiology-aware and task-invariant spatio-temporal modeling for cross-subject EEG decoding (2025). arXiv:2508.11357.
- [29] K. N. Spalding, et al., Shared latent representations of speech production for cross-patient speech decoding, *bioRxivPreprint* 2025.08.21.671516 (2025). doi:10.1101/2025.08.21.671516.
- [30] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli, J.-R. King, Decoding speech perception from non-invasive brain recordings, *Nature Machine Intelligence* 5 (2023) 1097–1107. doi:10.1038/s42256-023-00714-5.
- [31] 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.
- [32] 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.

- [33] M. Azabou, V. Arora, V. Ganesh, X. Mao, S. Nachimuthu, M. J. Mendelson, B. A. Richards, M. G. Perich, G. Lajoie, E. L. Dyer, A unified, scalable framework for neural population decoding, in: *Advances in Neural Information Processing Systems* 36 (NeurIPS), 2023. arXiv:2310.16046.
- [34] C. Wang, V. Subramaniam, A. U. Yaari, G. Kreiman, B. Katz, I. Cases, A. Barbu, BrainBERT: Self-supervised representation learning for intracranial recordings, in: *International Conference on Learning Representations (ICLR)*, 2023. arXiv:2302.14367.
- [35] A. Zahorodnii, et al., Neuroprobe: Evaluating intracranial brain responses to naturalistic stimuli (2025). arXiv:2509.21671.
- [36] N. Liang, L. Bougrain, Decoding finger flexion from band-specific ECoG signals in humans, *Frontiers in Neuroscience* 6 (2012) 91. doi:10.3389/fnins.2012.00091.
- [37] R. Flamary, A. Rakotomamonjy, Decoding finger movements from ECoG signals using switching linear models, *Frontiers in Neuroscience* 6 (2012) 29. doi:10.3389/fnins.2012.00029.
- [38] V. Lomtev, A. Kovalev, A. Timchenko, FingerFlex: Inferring finger trajectories from ECoG signals (2022). arXiv:2211.01960.
- [39] A. Petrosyan, M. Sinkin, M. Lebedev, A. Ossadtchi, Decoding and interpreting cortical signals with a compact convolutional neural network, *Journal of Neural Engineering* 18 (2) (2021) 026019. doi:10.1088/1741-2552/abe20e.
- [40] S. Bai, J. Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling (2018). arXiv:1803.01271.
- [41] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, G. D. Hager, Temporal convolutional networks for action segmentation and detection, in: *2017 IEEE*

- Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1003–1012. doi:10.1109/CVPR.2017.113.
- [42] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, WaveNet: A generative model for raw audio (2016). arXiv:1609.03499.
 - [43] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Vol. 9351 of Lecture Notes in Computer Science, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4_28.
 - [44] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance, EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces, *Journal of Neural Engineering* 15 (5) (2018) 056013. doi:10.1088/1741-2552/aace8c.
 - [45] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggersperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, *Human Brain Mapping* 38 (11) (2017) 5391–5420. doi:10.1002/hbm.23730.
 - [46] D. Kostas, F. Rudzicz, Thinker invariance: Enabling deep neural networks for BCI across more people, *Journal of Neural Engineering* 17 (5) (2020) 056008. doi:10.1088/1741-2552/abb7a7.
 - [47] D. Wu, Y. Xu, B.-L. Lu, Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016, *Neural Networks* 153 (2022) 235–253. doi:10.1016/j.neunet.2022.06.008.
 - [48] K. J. Miller, G. Schalk, E. E. Fetz, M. den Nijs, J. G. Ojemann, R. P. N. Rao, Cortical activity during motor execution, motor imagery,

and imagery-based online feedback, *Proceedings of the National Academy of Sciences* 107 (9) (2010) 4430–4435. doi:10.1073/pnas.0913697107.

- [49] G. Schalk, J. Kubánek, K. J. Miller, N. R. Anderson, E. C. Leuthardt, J. G. Ojemann, D. Limbrick, D. Moran, L. A. Gerhardt, J. R. Wolpaw, Decoding two-dimensional movement trajectories using electrocorticographic signals in humans, *Journal of Neural Engineering* 4 (3) (2007) 264–275. doi:10.1088/1741-2560/4/3/012.
- [50] S. Dähne, F. C. Meinecke, S. Haufe, J. Höhne, M. Tangermann, K.-R. Müller, V. V. Nikulin, SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviourally relevant parameters, *NeuroImage* 86 (2014) 111–122. doi:10.1016/j.neuroimage.2013.07.079.
- [51] I. Sturm, S. Lapuschkin, W. Samek, K.-R. Müller, Interpretable deep neural networks for single-trial EEG classification, *Journal of Neuroscience Methods* 274 (2016) 141–145. doi:10.1016/j.jneumeth.2016.10.008.
- [52] K. J. Miller, A library of human electrocorticographic data and analyses, *Nature Human Behaviour* 3 (11) (2019) 1225–1235. doi:10.1038/s41562-019-0678-3.
- [53] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. R. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, B. Blankertz, Review of the BCI competition IV, *Frontiers in Neuroscience* 6 (2012) 55. doi:10.3389/fnins.2012.00055.
- [54] 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.