WHITEN, THEN MIX: A TWO-PART ADAPTER FOR FEW-SHOT sEMG PERSONALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Personalizing a gesture decoder to a new sEMG band wearer is conventionally evaluated by the gain from that wearer's own calibration data, but this quantity does not distinguish transferable improvement from wearer-specific benefit. A correction that closes a between-wearer gap may also benefit other wearers. We introduce a cross-wearer substitution diagnostic: each correction is fitted on a different wearer and applied to the recipient. Using this diagnostic, the two axes show different transfer behavior: cross-wearer amplitude corrections transfer at every measured depth, whereas cross-wearer electrode-position corrections widen the gap. Under an invertible linear mixing model with identity source covariance, the population spatial covariance identifies the symmetric polar factor but not the orthogonal one. This distinction motivates our two-part few-shot adapter: Euclidean Alignment performs label-free, wearer-specific whitening, while ElectrodeMix adapts the residual on the electrode axis with $c^2 g$ learned parameters beyond the classifier head. Meta-pretraining uses a restricted inner loop to support this adaptation. Across the evaluated sEMG archives and both backbones, the method achieves the highest mean accuracy in every 10-shot cross-dataset setting. On EPN612-to-UCI-EMG transfer with ViM-Tiny, it exceeds the frozen probe for all 34 wearers, by 6.9 percentage points on average, using 1,024 adapter parameters versus 6.8M for full fine-tuning. This diagnostic requires only an additional cross-wearer fit and distinguishes transferable improvement from the advantage of recipient-specific fitting.

1 INTRODUCTION

Surface electromyography (sEMG) records muscle electrical activity through electrodes on the skin, but the observed signal is shaped by its transmission through the overlying tissue. Muscle thickness and position, subcutaneous fat thickness, and skin impedance differ across individuals, so the waveform differs from wearer to wearer even for the same gesture (Ozdemir et al., 2020; McManus et al., 2020). Moreover, such systems are typically deployed as forearm bands that are repeatedly put on and removed without assistance, resulting in variation across wearers in per-electrode amplitudes and the angle at which the band is seated (Jarque-Bou et al., 2024; Li et al., 2023b) (Fig. 1a).

Under a minimal adaptation budget, approaches differ in what they adapt and fall into three groups: input-stage alignment of wearer statistics to a reference distribution (Sun & Saenko, 2016; He & Wu, 2019), parameter-efficient fine-tuning with a frozen encoder (Hu et al., 2022; Zaken et al., 2022), and test-time adaptation that calibrates predictions without labels (Wang et al., 2021; Iwasawa & Matsuo, 2021). Meta-learning (Finn et al., 2017; Nichol & Schulman, 2018; Snell et al., 2017) designs the pretraining itself so that such adaptation can proceed from a few examples, and subsequent work (Raghu et al., 2019; Oh et al., 2020; Zintgraf et al., 2019) examines which parameters should be updated in the inner loop.

An own-calibration gain alone cannot establish wearer specificity. It can contain both a transferable benefit and an additional advantage from fitting the correction to its recipient. At the raw-signal level, inter-wearer variation separates into per-electrode gains and the band angle (Kyeong & Kim, 2020), and correcting either axis with the wearer's own calibration data substantially reduces the gap. Because this is the quantity usually reported, the two axes are indistinguishable at this level.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

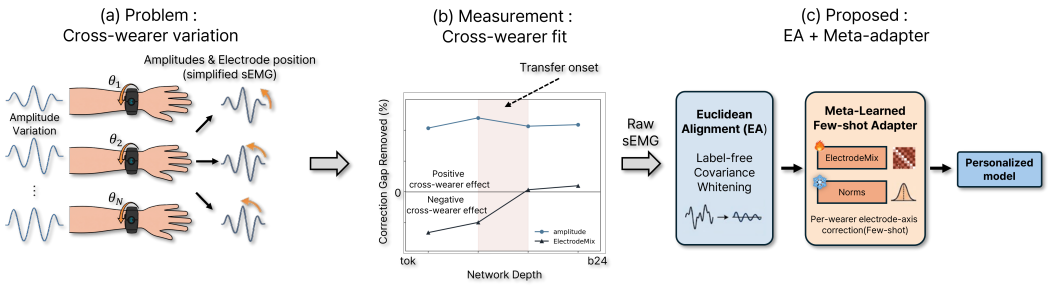

Figure 1: **Two-part few-shot personalization.** (a) Wearers differ in amplitude and electrode position. (b) Cross-wearer amplitude corrections reduce the gap at all measured depths, whereas the cross-wearer effect of ElectrodeMix becomes less negative with depth and turns positive in ViM-Tiny. (c) Wearer-specific EA whitens the input, while meta-pretrained ElectrodeMix adapts the residual at the tokenizer.

To distinguish these effects, we fit each correction on a different wearer and apply it to the recipient. A positive cross-wearer gain indicates transferable benefit, whereas the own-versus-cross difference measures the additional value of recipient-specific fitting (Fig. 1b). Amplitude corrections transfer across all measured depths, whereas position corrections widen the gap. Unconstrained electrode-axis maps acquire a transferable component deeper in the encoder.

An idealized linear model motivates this separation. For an invertible wearer transform $W = SQ$ and a source with identity covariance, the spatial covariance is $R = WW^\top = S^2$. Thus, R identifies S but not Q, and population-covariance whitening leaves Qs. This motivates separating covariance normalization from supervised residual adaptation. The polar argument motivates this separation rather than prescribing a particular constraint on the deployed residual adapter.

We therefore propose a two-part adapter (Fig. 1c). Euclidean Alignment (EA) provides wearer-specific whitening without labels or learned parameters, using the unique symmetric positive-definite whitener (Kessy et al., 2018). ElectrodeMix then adapts a free $c \times c$ map per embedding group at the tokenizer to handle residual wearer variation. Meta-pretraining supports this adaptation by updating ElectrodeMix and normalization affines in the inner loop. Downstream adaptation updates ElectrodeMix and a classifier head while keeping the encoder frozen.

Our contributions are as follows.

- We introduce a cross-wearer substitution diagnostic that separates transferable correction effects from the additional advantage of recipient-specific fitting, demonstrate its applicability across adapter families, and use it to analyze corrections across representation depths.
- We propose a two-part few-shot adapter combining label-free EA with supervised ElectrodeMix, using $c^2 g$ learned adapter parameters per wearer. Covariance identifiability under an idealized mixing model motivates separating normalization from residual adaptation, without prescribing the residual adapter's deployment constraint.
- We support the adapter with restricted meta-pretraining and evaluate few-shot personalization across sEMG archives and both backbones.

2 Related Work

Wearer variation in sEMG. Impedance, subcutaneous fat, muscle mass, and electrode placement contribute to inter-wearer variation (Okajima et al., 2019), and benchmarks report substantial unseen-wearer gaps (Kanoga et al., 2020; Nkurikiyeyezu et al., 2019; Gu et al., 2018). Existing methods estimate and compensate for electrode shifts (Xu et al., 2020), design shift-invariant features (Lin et al., 2024), or augment shifts during training (Zhang et al., 2025). We measure offsets on a continuous grid, resolving fractional displacements beyond discrete reindexing. Repeated unassisted placement causes session-to-session variation (Min et al., 2019), making personalization recurrent and constrained by the labels available for each calibration.

Correcting without labels. EA (He & Wu, 2019) whitens mean trial covariance to the identity (Junqueira et al., 2024), CORAL (Sun & Saenko, 2016) aligns second-order statistics, and Riemannian alignment uses the geometric mean (Zanini et al., 2017; Li & Zhang, 2025). EA and AdaBN (Li et al., 2016) operate at the input and internal normalization layers, respectively, but effectiveness at a given depth does not establish wearer specificity. Label-free adaptation can also update normalization affines through entropy minimization (TENT (Wang et al., 2021)), prototypes (T3A (Iwasawa & Matsuo, 2021)), or the feature extractor under a fixed source hypothesis (SHOT (Liang et al., 2021)). Our identifiability statement in Sec. 3 concerns a single population spatial covariance under the stated mixing model and does not characterize what all label-free methods can learn.

What to adapt. Parameter-efficient fine-tuning restricts adaptation to low-rank branches (LoRA (Hu et al., 2022)), biases (BitFit (Zaken et al., 2022)), bottleneck adapters (Houlsby et al., 2019), or scaling vectors ($(IA)^3$ (Liu et al., 2022)). Meta-learning supports rapid adaptation (Finn et al., 2017; Nichol & Schulman, 2018; Snell et al., 2017), with restricted inner loops that update only the head (ANIL (Raghu et al., 2019)), body (BOIL (Oh et al., 2020)), or context parameters (CAVIA (Zintgraf et al., 2019)). Recent biosignal personalization uses adapters, meta-learning, and prototype-based sEMG adaptation (Liu et al., 2025; Manneschi et al., 2026; Lee et al., 2026), with compact subject-specific updates. The closest channel-axis operators are the Spatial Adaptation Layer (SAL) (Pereira et al., 2024), which learns an affine map to the original session's spatial frame, and subject-specific EEG layers (Li et al., 2023a), which model electrode-position and impedance differences through linear scaling and mixing. Personalized federated learning likewise assumes shared and personal parameter subsets (Zhou et al., 2024; Yang et al., 2024). These methods typically evaluate gains from the target wearer's own calibration data, without testing whether a correction fitted on another wearer provides the same benefit. Our diagnostic makes this substitution the measurement.

3 PRELIMINARIES

Problem Statement. Let $\mathcal{W} = \{w_i\}_{i=1}^{N}$ be a set of N wearers, and let a recorded window be $x \in \mathbb{R}^{c\times T}$ over c electrodes and T samples. The gesture label space $\mathcal{Y}$ is shared: the desired output is the same, but the map from gesture to signal differs across wearers. The base model M_0, with parameters Θ_0, is trained once on a population and deployed unchanged. A wearer w_i supplies a calibration block $\mathcal{D}_i^{\mathrm{cal}}$ from a single session, of which only k windows per class carry labels. The model is evaluated on a later session $\mathcal{D}_i^{\mathrm{eval}}$, and the only state stored for that wearer is a personalization increment $\Delta\Theta_i$. Let $M_i[\Delta\Theta]$ denote the model for w_i carrying increment $\Delta\Theta$, which need not have been fitted on w_i, and let $L_i(\Delta\Theta)$ denote its held-out loss on $\mathcal{D}_i^{\mathrm{eval}}$. Personalization typically reports the gain from a wearer's own increment, but whether this gain is wearer-specific is revealed only by substituting another wearer's increment:

$$\Delta_i^{\mathrm{report}} = \mathcal{L}_i(0) - \mathcal{L}_i(\Delta\Theta_i), \qquad \Delta_i^{\mathrm{personal}} = \mathcal{L}_i(\Delta\Theta_j) - \mathcal{L}_i(\Delta\Theta_i), \quad j = \sigma(i), \tag{1}$$

where σ is a fixed permutation without fixed points. The cross-wearer gain is $\Delta_i^{\mathrm{report}} - \Delta_i^{\mathrm{personal}}$, so positive transfer can coexist with a positive $\Delta_i^{\mathrm{personal}}$, which represents the own-fit advantage for this pairing. Prior sEMG personalization studies have primarily evaluated gains from a wearer's own calibration data rather than explicitly applying a correction fitted on another wearer to the recipient.

Token Representation. We assume an encoder that preserves the electrode axis through tokenization: the tokenizer Tok emits a grid $Z \in \mathbb{R}^{c\times T'\times D}$ over the electrode index, T' time patches, and the embedding dimension. Writing $L = \{1, \ldots, |L|\}$ for the ordered index set of blocks, the base model is $M_0 = \left(\bigcirc_{\ell\in L} B^{\ell}\right) \circ \mathrm{Tok}$, with $\Theta_0 = \Theta^{\mathrm{Tok}} \cup \bigcup_{\ell\in L} \Theta^{B^{\ell}}$.

Wearer Transform. We model wearer-dependent gain and electrode mixing as $x = W_i s$, with $W_i \in \mathbb{R}^{c\times c}$ invertible and a shared source satisfying $\mathbb{E}[ss^\top] = I$. The unique polar decomposition (Higham, 1986) is $W_i = S_i Q_i$, where $S_i = (W_i W_i^\top)^{1/2}$ is symmetric positive definite and Q_i is orthogonal. Hence, the population spatial covariance is $R_i := \mathbb{E}[xx^\top] = S_i^2$. It determines S_i but not Q_i, since $S_i Q$ yields the same R_i for every orthogonal Q. These are algebraic factors rather than uniquely identified physiological causes. This identifiability statement concerns a single spatial covariance under the stated source model, not all unlabeled information.

Euclidean Alignment. A widely used prior approach to cross-subject alignment in EEG-based brain–computer interfaces whitens each subject's spatial covariance to the identity (He & Wu, 2019)

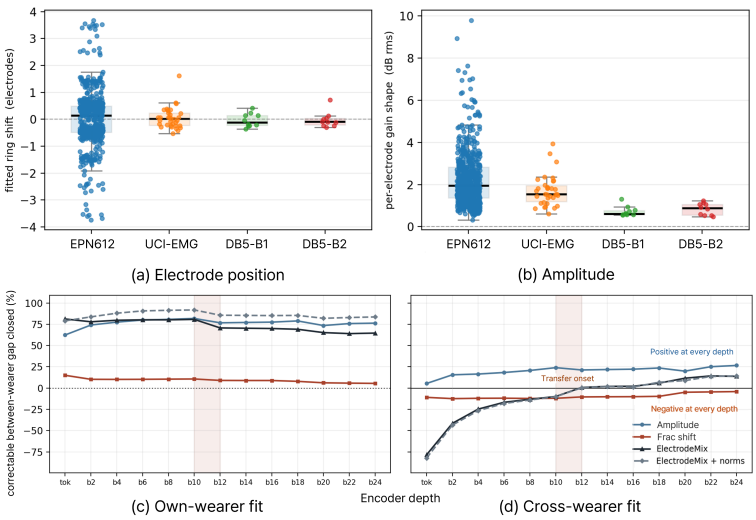

Figure 2: **Between-wearer variation and correction transfer across depth. Top**, raw-recording measurements across EPN612, UCI-EMG, and both DB5 rings: (a) wearer-level electrode-position offsets and (b) per-electrode gain differences. **Bottom**, EPN612 with ViM-Tiny: normalized gap closure across depth for corrections fitted on (c) the same wearer and (d) another wearer. Corrections are a fractional ring shift, a normalization affine, an unconstrained $c \times c$ electrode-axis map (ElectrodeMix), and ElectrodeMix+affine.

by averaging it over the wearer's unlabeled calibration windows and applying its inverse square root:

$$\bar{R}_i = \frac{1}{|\mathcal{D}_i^{\text{cal}}|} \sum_{x \in \mathcal{D}_i^{\text{cal}}} xx^\top, \qquad \tilde{x} = \bar{R}_i^{-1/2} x. \tag{2}$$

For positive-definite R_i, $R_i^{-1/2}$ is the unique symmetric positive-definite whitener. Under the population model, it equals S_i^{-1} and yields $Q_i s$. Eq. 2 estimates this operation from calibration windows. EA is wearer-specific but label-free, and its stored matrix contains no learned parameters.

4 Two axes

Between-wearer variation is measured directly from the raw recordings with no model involved. For each wearer, we compute the per-electrode RMS of their windows, fit a correction in closed form on one session, and evaluate it on the other against the leave-one-out population of their own archive (Fig. 2). We quantify the effect of a correction $\mathcal{C}$ by its normalized gap closure:

$$\text{closed}(\mathcal{C}) = \frac{U - D(\mathcal{C})}{U - A}, \qquad A = \sqrt{\tfrac{1}{2}F^2 + \tfrac{1}{N}U^2}, \tag{3}$$

where U, $D(\mathcal{C})$, and F are obtained by averaging wearer-level root-mean-square distances within each archive before correction, after correction, and between sessions, respectively. Appendix B.1 gives their exact construction and motivates A, which serves only as a normalization reference, not a lower bound.

Electrode position. We fit the circular shift that best aligns a wearer's per-class electrode pattern to the population on a continuous grid, using the exact band-limited shift induced by a DFT phase ramp. Most wearers have fitted offsets within one electrode (UCI-EMG 97%, both DB5 rings 100%), whereas EPN612 exhibits a wider distribution. An integer ring shift cannot capture fractional displacements: when fitted on the same wearer, it closes 5.2% of the correctable gap, compared with 18.4% for the fractional ring shift on UCI-EMG, and 6.6% compared with 10.9% on EPN612. The fractional ring shift yields higher normalized gap closure on all four archives.

Amplitude. We define the per-electrode gain shape in dB relative to the population after dividing out overall loudness and undoing the fitted rotation. This de-rotation is not cosmetic: without it, the amplitude axis attributes position variation to gain.

The across-wearer correlation between the absolute fitted shift and the RMS of the de-rotated per-electrode gain deviations in dB remains nonzero ($r = 0.24$ on EPN612 and $r = 0.52$ on UCI-EMG). We treat gain and position as separately parameterized corrections, not statistically independent factors. The two-part design follows the covariance-normalization and residual-adaptation distinction in Sec. 3 rather than an independence assumption.

5 Transferable and Wearer-Specific Corrections

An own-fitted correction can provide transferable benefit and a wearer-specific advantage. We evaluate these effects on encoder representations using Eq. 3, fitting corrections on the recipient's or another wearer's first session and evaluating them on the recipient's second session (Fig. 2d). Their difference, $[D(C_j) - D(C_i)]/(U - A)$, is a representation-level own-fit advantage rather than the task-loss quantity in Eq. 1. We test four corrections at thirteen depths, from the tokenizer to every second block of a 24-block encoder.

Transferability and wearer specificity. Positive cross-wearer closure indicates transferable benefit, whereas a positive own-minus-cross contrast indicates a wearer-specific advantage. Both can coexist when another wearer's correction helps but the recipient's own correction helps more. Fig. 2d shows the cross-wearer effect, while the difference from Fig. 2c measures the advantage of recipient-specific fitting. For electrode-axis maps, the zero crossing marks the onset of positive transfer, not loss of this advantage.

Amplitude. Cross-wearer affine corrections close a positive share of the gap at all thirteen depths, from $+5.4\%$ at the tokenizer to $+26.6\%$ after 24 blocks. The diagnostic uses normalization affines because EA whitening depends on each wearer's unlabeled covariance.

Electrode position. Cross-wearer shifts widen the gap at all thirteen depths (-10.9% to -4.1%), whereas own-fitted shifts reduce it (15.0% to 5.5%), preserving an own-fit advantage throughout.

The unconstrained map. For ViM-Tiny, the cross-wearer free map changes from -78.4% at the tokenizer to $+13.8\%$ after 24 blocks, crossing zero between blocks 10 and 12 (Appendix B.3). Own-fit closure remains above cross-wearer closure at all depths, with the largest separation at the tokenizer. This motivates placing ElectrodeMix before deeper mixing makes the correction increasingly transferable. A known one-electrode input shift yields a refitted correction of -0.940 at block 12 (Appendix B.4), showing that zero crossing does not imply loss of position information. Thus, tokenizer placement targets the strongest recipient-specific residual.

6 Method

We combine wearer-specific covariance normalization with supervised residual adaptation (Fig. 3). EA is reserved for deployment, where it performs explicit recipient-specific whitening using the target wearer's calibration covariance. During meta-pretraining, the inner loop instead adapts ElectrodeMix and the normalization affines, allowing task-specific normalization while reserving explicit covariance whitening for deployment. Under the model in Sec. 3, population whitening leaves an orthogonal residual, whereas empirical whitening need not (Appendix B.2).

ElectrodeMix. We apply $I + M$ along the token grid's electrode axis, with M zero-initialized. The embedding is partitioned into g groups, each with its own $c \times c$ map, enabling group-dependent electrode mixing with $c^2 g$ adapter parameters beyond the classifier head. The maps remain unconstrained because residuals after empirical whitening need not be orthogonal. The objective is task-driven adaptation, not recovery of a uniquely identified physical rotation.

Meta-pretraining. We pretrain the encoder to support few-shot adaptation through ElectrodeMix. Each wearer defines a label-free contrastive task: two augmentations of one window form a positive pair, while windows from other wearers serve as negatives. Augmenting n windows of wearer d twice gives $z_1, \ldots, z_{2n}$, with $z_i, z_{\pi(i)}$ forming a positive pair, while single views $u_1, \ldots, u_{K_{\text{neg}}}$

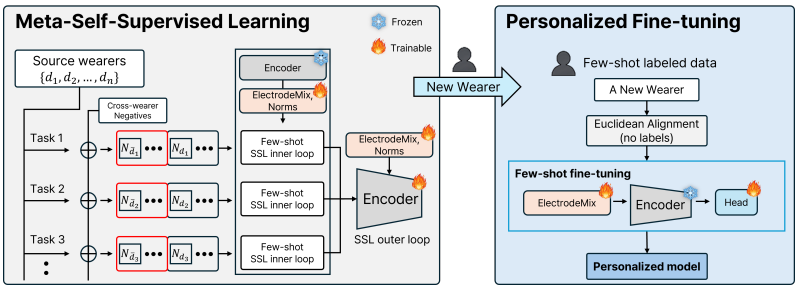

Figure 3: **The proposed method.** Left: each source wearer forms a pretraining task. The inner loop adapts ElectrodeMix and the normalization affines with a self-supervised objective while freezing the encoder and using other-wearer windows as negatives ($\mathcal{N}_{\bar{d}}$); the outer loop updates the encoder. Right: recipient calibration determines Euclidean Alignment, after which ElectrodeMix and a classifier head are adapted on labeled shots with the encoder frozen.

from the other wearers $\mathcal{N}_{\bar{d}}$ enter only as negatives:

$$\mathcal{L} = -\frac{1}{2n}\sum_{i=1}^{2n} \log \frac{\exp\left(s_{i\pi(i)}/\tau\right)}{\sum_{a\neq i}\exp\left(s_{ia}/\tau\right) + \sum_{m=1}^{K_{\text{neg}}}\exp\left(z_i^\top u_m/\tau\right)}, \qquad s_{ab} = z_a^\top z_b. \tag{4}$$

We use $\tau = 0.1$ and ℓ_2-normalized projection-head outputs z_i, u_m. The downstream classifier uses encoder features directly, without the projection head or this ℓ_2 normalization. At $K_{\text{neg}} = 0$, Eq. 4 contains only within-wearer comparisons; adding other-wearer negatives introduces explicit cross-wearer comparisons (Appendix C.2).

Augmentation. Among the pretraining augmentations, we use global gain and fractional ring shift to perturb the two wearer-variation axes. Global scaling preserves relative channel amplitudes, whereas the interpolated shift changes their spatial arrangement and attenuates high spatial frequencies. Unlike the phase-ramp shift used for measurement in Sec. 4, this interpolation is not a pure shift and cannot be reversed by a counter-shift alone. The augmentation is intended to discourage matching based solely on an unchanged electrode pattern, while ElectrodeMix is fitted to the recipient's calibration data. Exact operators and the full pipeline are provided in Appendix C.2.

The two loops. Let ψ denote the electrode map together with the layer-normalization affine. The inner loop takes K gradient steps on Eq. 4, updating ψ alone. The outer loop then updates every parameter through the adapted ψ under a first-order approximation (Appendix C.3). Downstream adaptation updates ElectrodeMix and the classifier head while keeping the encoder and normalization affines fixed. Meta-pretraining supports this restricted adaptation, but its inner-loop parameter set is not identical to the downstream update set.

7 Experiments

7.1 Experimental setups

Dataset. We use three sEMG archives recorded with 8-electrode forearm rings. EPN612 (Benalcázar et al., 2021) contains 612 wearers and six gesture classes, of whom 306 are used for adaptation evaluation. UCI-EMG (Krilova et al., 2018) contains 36 wearers with two sessions each, of whom 34 are used in our evaluation, with the second session held out from correction fitting. NinaPro DB5 (Pizzolato et al., 2017) includes 10 wearers and 53 classes recorded simultaneously with two Myo bands; we denote the two ring placements as DB5-B1 and DB5-B2. We use 200-sample windows at 200 Hz. UCI-EMG and EPN612 are scaled by a single factor shared across electrodes, preserving relative amplitudes across electrodes; DB5 is standardized per electrode within each recording, attenuating these relative differences and making its gain-shape values in Fig. 2b conservative. Protocols and splits are provided in Appendix A.1.

Metrics. We report evaluation-session accuracy at the end of the adaptation budget, averaging five draws within each wearer and then across wearers. Trial-seeded sampling provides all arms with

Table 1: **Intra-dataset 10-shot adaptation accuracy (%).** Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the Linear probe, which serves as the reference for all comparisons. Params denotes the per-wearer increment. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

				UCI-EMG	EPN612	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	84.54 ± 7.28 (0/34)	70.72 ± 9.38 (0/306)	30.60 ± 3.76 (0/10)	29.59 ± 4.92 (0/10)
		ViT-Tiny	0	83.75 ± 8.11 (0/34)	69.79 ± 9.61 (0/306)	38.23 ± 2.84 (0/10)	41.24 ± 4.50 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	84.61 ± 7.20 (15/34)	49.67 ± 11.02 (1/306)	**35.16 ± 4.98** (10/10)	**33.33 ± 5.88** (9/10)
		ViT-Tiny	5.37M	76.39 ± 8.28 (3/34)	57.15 ± 9.22 (1/306)	39.23 ± 2.70 (8/10)	41.05 ± 4.49 (4/10)
BitFit	PEFT	ViM-Tiny	41.9K	84.57 ± 7.06 (15/34)	68.37 ± 10.39 (66/306)	28.82 ± 3.78 (0/10)	27.90 ± 5.61 (2/10)
		ViT-Tiny	25.7K	83.74 ± 8.44 (19/34)	66.07 ± 10.20 (36/306)	42.15 ± 2.72 (10/10)	44.46 ± 4.44 (10/10)
LoRA	PEFT	ViM-Tiny	13.8K	84.56 ± 7.33 (12/34)	70.80 ± 9.38 (165/306)	30.36 ± 3.74 (3/10)	29.39 ± 4.85 (0/10)
		ViT-Tiny	9.2K	83.92 ± 8.00 (14/34)	70.10 ± 9.64 (192/306)	40.05 ± 2.84 (9/10)	43.77 ± 3.93 (10/10)
T3A	TTA	ViM-Tiny	0	84.43 ± 7.69 (16/34)	64.42 ± 9.77 (2/306)	34.08 ± 4.41 (10/10)	32.59 ± 6.18 (9/10)
		ViT-Tiny	0	83.15 ± 9.08 (13/34)	63.60 ± 9.73 (4/306)	36.56 ± 2.48 (2/10)	40.77 ± 4.12 (5/10)
CORAL	Align	ViM-Tiny	9.2K	61.30 ± 8.69 (0/34)	61.76 ± 9.80 (7/306)	8.25 ± 1.60 (0/10)	7.96 ± 2.04 (0/10)
		ViT-Tiny	9.2K	63.27 ± 9.82 (0/34)	60.55 ± 9.72 (4/306)	17.15 ± 1.91 (0/10)	14.30 ± 2.47 (0/10)
SAL	Align	ViM-Tiny	72	86.69 ± 6.13 (24/34)	68.24 ± 10.34 (86/306)	32.23 ± 3.85 (9/10)	29.78 ± 5.62 (6/10)
		ViT-Tiny	72	82.73 ± 7.34 (14/34)	64.74 ± 11.41 (65/306)	44.08 ± 3.63 (10/10)	44.37 ± 5.20 (10/10)
EA	Align	ViM-Tiny	0	86.43 ± 6.82 (27/34)	69.84 ± 9.25 (105/306)	32.79 ± 4.20 (8/10)	30.54 ± 5.32 (8/10)
		ViT-Tiny	0	85.24 ± 7.88 (21/34)	70.43 ± 9.55 (174/306)	39.13 ± 3.86 (8/10)	41.75 ± 4.88 (5/10)
AdaBN	Align	ViM-Tiny	0	84.96 ± 7.33 (21/34)	71.58 ± 9.37 (219/306)	29.60 ± 3.24 (0/10)	28.98 ± 5.05 (4/10)
		ViT-Tiny	0	83.37 ± 7.54 (14/34)	69.78 ± 9.75 (155/306)	42.14 ± 2.80 (10/10)	45.20 ± 4.24 (10/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	**90.13 ± 6.02** (32/34)	**72.91 ± 9.39** (237/306)	34.58 ± 4.42 (8/10)	32.89 ± 4.87 (9/10)
		ViT-Tiny	1.0K	**88.65 ± 6.21** (30/34)	**73.30 ± 9.64** (258/306)	**47.76 ± 2.85** (10/10)	**48.27 ± 4.84** (10/10)

identical windows for paired comparisons. Fold-based UCI-EMG and DB5 results pool 34 and 10 wearers, respectively. In each of five folds, the encoder is pretrained on four folds and evaluated on the held-out fold, so each wearer is evaluated once with an encoder not pretrained on that wearer. These are wearer counts, not independent encoder-training runs. We report mean accuracy, gains over the same encoder's frozen Linear probe, and adapted parameter counts. For accuracy, $\pm$ denotes the sample standard deviation across wearers. For Δ, paired 95% t confidence intervals are reported where shown; ablation tables report only the corresponding point differences. Sections 4 and 5 instead report Eq. 3 without a classifier. Full statistics are in Appendix A.3.

Implementation Details. We use a bidirectional Vision Mamba (Zhu et al., 2024) (192 channels, 24 blocks, state size 16) and a ViT-Tiny Transformer (Touvron et al., 2021) (192 channels, 12 blocks, 3 heads). Both divide $(8, 200)$ windows into $(1, 10)$ patches, yielding 160 tokens with the electrode index on the first axis. ElectrodeMix uses $g = 16$ and $c = 8$, giving $c^2 g = 1{,}024$ adapter parameters, excluding the classifier head and EA matrix. Downstream adaptation uses $1, 3, 5, 10$ labeled windows per class. The label-free fits (EA, CORAL) reuse the same calibration block without labels—60 windows per wearer at 10-shot and 530 on DB5—so the unlabeled budget equals the labeled one; only T3A reads evaluation windows. We report the 10-shot results here and the remaining results in Appendix D.1. Further details are provided in Appendices A.3 and C.2.

7.2 MAIN RESULTS

Intra-dataset. Tab. 1 reports adaptation within the same archive. Full FT is inconsistent, losing 21.05 pp on EPN612 with ViM-Tiny despite updating 6.81M parameters. PEFT and alignment baselines are also condition-dependent, with CORAL below the Linear probe in every cell and EA generally among the strongest alignment methods. Our method performs best on UCI-EMG and EPN612 for both backbones with only a 1.0K per-wearer increment, reaching 90.13% and 73.30%, respectively, and outperforming the Linear probe for 258/306 EPN612 wearers. On DB5-B1, it trails Full FT slightly with ViM-Tiny but ranks first with ViT-Tiny. Its gains over EA alone are 1.79 pp and 8.63 pp, supporting the two-part design in which covariance normalization is followed by supervised residual adaptation.

Cross-dataset. Tab. 2 reports transfer from EPN612 to unseen UCI-EMG and DB5 archives, which differ in recording setup, gesture set, and band placement. Full FT transfers inconsistently, whereas EA is the strongest alignment baseline overall. Our method performs best in every cell, reaching 86.93% on UCI-EMG with ViM-Tiny and 36.92%/37.15% on DB5-B1/B2 with ViT-Tiny. It outper-

Table 2: **Cross-dataset 10-shot adaptation accuracy (%).** All methods pretrain on EPN612 and deploy on archives unseen during pretraining. Win, in parentheses, counts wearers outperforming the Linear probe, the reference for all comparisons. Params denotes the per-wearer increment. **Bold** marks the best result in each cell, underline the second-best.

					Test set		
					UCI-EMG	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Train set**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	EPN612	80.03 ± 8.04 (0/34)	23.67 ± 3.43 (0/10)	24.38 ± 3.00 (0/10)
		ViT-Tiny	0		75.64 ± 7.95 (0/34)	23.72 ± 3.07 (0/10)	23.98 ± 3.05 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	EPN612	74.15 ± 8.60 (3/34)	28.46 ± 3.92 (10/10)	31.30 ± 5.54 (10/10)
		ViT-Tiny	5.37M		64.06 ± 7.10 (1/34)	26.90 ± 3.95 (9/10)	28.20 ± 4.62 (10/10)
BitFit	PEFT	ViM-Tiny	41.9K	EPN612	81.69 ± 7.67 (26/34)	25.26 ± 3.46 (9/10)	25.73 ± 3.83 (9/10)
		ViT-Tiny	25.7K		77.11 ± 7.59 (23/34)	25.50 ± 3.61 (10/10)	26.70 ± 3.62 (10/10)
LoRA	PEFT	ViM-Tiny	13.8K	EPN612	80.02 ± 8.07 (14/34)	23.11 ± 2.87 (2/10)	23.99 ± 2.81 (3/10)
		ViT-Tiny	9.2K		76.09 ± 7.60 (25/34)	23.82 ± 3.36 (6/10)	24.13 ± 2.90 (7/10)
T3A	TTA	ViM-Tiny	0	EPN612	79.80 ± 8.94 (16/34)	24.21 ± 3.78 (7/10)	25.25 ± 3.13 (6/10)
		ViT-Tiny	0		74.41 ± 8.93 (11/34)	21.54 ± 3.15 (0/10)	21.98 ± 2.48 (0/10)
CORAL	Align	ViM-Tiny	9.2K	EPN612	62.92 ± 8.38 (0/34)	18.91 ± 2.05 (1/10)	18.53 ± 2.54 (0/10)
		ViT-Tiny	9.2K		54.66 ± 9.45 (0/34)	22.02 ± 2.65 (4/10)	22.94 ± 3.42 (4/10)
SAL	Align	ViM-Tiny	72	EPN612	81.61 ± 7.55 (22/34)	25.72 ± 3.33 (10/10)	28.34 ± 4.26 (10/10)
		ViT-Tiny	72		74.75 ± 7.98 (13/34)	31.30 ± 2.75 (10/10)	32.23 ± 3.72 (10/10)
EA	Align	ViM-Tiny	0	EPN612	83.22 ± 8.01 (30/34)	31.54 ± 2.44 (10/10)	32.48 ± 3.34 (10/10)
		ViT-Tiny	0		78.36 ± 7.68 (30/34)	32.26 ± 1.83 (10/10)	33.16 ± 3.49 (10/10)
AdaBN	Align	ViM-Tiny	0	EPN612	81.06 ± 7.88 (23/34)	24.64 ± 3.16 (8/10)	25.27 ± 3.49 (8/10)
		ViT-Tiny	0		76.35 ± 7.41 (21/34)	24.83 ± 3.39 (8/10)	25.69 ± 3.20 (10/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	EPN612	**86.93 ± 7.42** (34/34)	**33.18 ± 2.36** (10/10)	**34.83 ± 4.00** (10/10)
		ViT-Tiny	1.0K		**84.26 ± 8.13** (33/34)	**36.92 ± 2.55** (10/10)	**37.15 ± 3.41** (10/10)

Table 3: **Task-level personal advantage.** We report 10-shot accuracy (%) on UCI-EMG. Cross and own denote donor- and recipient-fitted ElectrodeMix, respectively. Win, shown in parentheses, counts the wearers for whom the current row outperforms the preceding row; Δ denotes the mean paired gain (pp) over that same preceding row. **Bold** indicates the best result for each backbone.

	ViM-Tiny		ViT-Tiny	
Deployment	**Acc↑ (Win)**	**Δ vs. previous**	**Acc↑ (Win)**	**Δ vs. previous**
Linear probe	84.54 (0/34)	–	83.75 (0/34)	–
+ EA	86.43 (27/34)	+1.89 ± 0.96	85.24 (21/34)	+1.49 ± 1.38
+ EA + ElectrodeMix (cross)	86.50 (16/34)	+0.07 ± 1.13	86.35 (24/34)	+1.11 ± 1.13
+ EA + ElectrodeMix (own)	**90.13 (32/34)**	**+3.63 ± 1.09**	**88.65 (28/34)**	**+2.30 ± 1.01**

forms the Linear probe for 34/34 and 33/34 UCI-EMG wearers on ViM-Tiny and ViT-Tiny, respectively, and all DB5 wearers. Relative to EA, it gains 3.71–5.90 pp on UCI-EMG and 1.64–4.66 pp on DB5, showing that the benefit of residual adaptation persists beyond the pretraining archive.

Task-level personal advantage. Tab. 3 compares donor- and recipient-fitted ElectrodeMix on UCI-EMG. Both maps are fitted using the recipient's pretrained encoder and then frozen while the classifier is refitted on identical recipient shots under the same training protocol, with the recipient's EA fixed. The donor map provides approximately zero mean gain over EA with ViM-Tiny ($+0.07$ pp) and $+1.11$ pp with ViT-Tiny. Recipient-fitted maps reach 90.13% and 88.65%, exceeding donor-fitted maps by 3.63 and 2.30 pp, with positive differences for $32/34$ and $28/34$ wearers, respectively. For ViT-Tiny, transferable benefit therefore coexists with a personal advantage. The sign of the cross-wearer effect differs from that in Sec. 5, where closure is measured without front-stage whitening. The diagnostic is not specific to ElectrodeMix; applying it to other adaptation methods reveals that similar own-calibration gains can conceal distinct transfer structures (Appendix B.5).

7.3 ABLATION STUDY

Where the adapter goes. Sec. 5 shows that the free map's cross-wearer effect varies with depth. Tab. 4 evaluates adapter placement using a separately meta-pretrained encoder for each depth. Because the initial representations differ, we consider both final accuracy and the gain over each encoder's own frozen probe. Tokenizer placement (Ours) achieves the highest mean accuracy (88.65%)

Table 4: **Where the adapter goes.** 10-shot accuracy (%) on UCI-EMG with ViT-Tiny. Each placement is meta-pretrained separately. We report final accuracy and the gain over each encoder's own frozen probe. Win counts the wearers for whom the adapted model outperforms that probe.

Placement	Linear probe	Acc↑ (Win)	Δ over probe
tokenizer (proposed)	**83.75**	**88.65 ± 6.21** (30/34)	**+4.90**
block 2	85.81	87.84 ± 6.36 (28/34)	+2.03
block 4	86.32	87.84 ± 6.33 (26/34)	+1.52
block 6	86.26	87.87 ± 7.08 (27/34)	+1.61
block 8	85.70	87.68 ± 6.80 (27/34)	+1.98
block 10	86.01	87.77 ± 6.87 (24/34)	+1.76
block 12	85.46	86.74 ± 6.81 (21/34)	+1.28

Table 5: **Which alignment goes in front.** 10-shot accuracy (%) on EPN612 with ViM-Tiny. EA is replaced by an alternative front-stage alignment or removed entirely, while ElectrodeMix and the downstream optimization remain unchanged.

Front	Linear probe	Acc↑ (Win)	Δ over probe
EA (Proposed)	**70.72**	**72.91 ± 9.39** (237/306)	**+2.19**
None	70.72	71.89 ± 9.43 (217/306)	+1.17
Riemannian	70.72	71.86 ± 9.46 (188/306)	+1.14
CORAL	70.72	62.49 ± 9.67 (7/306)	-8.23

Table 6: **What the inner loop adapts.** 10-shot accuracy (%) on DB5-B1 with ViT-Tiny. Standard MAML adapts the full encoder; No inner loop omits inner-loop adaptation. The pretraining objective and downstream protocol are fixed.

Pretraining	Linear probe	Acc↑ (Win)	Δ over probe
Proposed	**38.23**	**47.76 ± 2.85** (10/10)	**+9.52**
Standard MAML	27.98	34.41 ± 3.41 (10/10)	+6.42
No inner loop	43.96	46.49 ± 2.53 (10/10)	+2.53

and the largest gain ($+4.90$ pp), outperforming its 83.75% probe for $30/34$ wearers. All block placements yield lower final accuracies (86.74–87.87%) and smaller gains over their respective probes ($+1.28$–$+2.03$ pp). These results support tokenizer placement in this setting.

Which alignment goes in front. Table 5 examines the choice of front-stage alignment by replacing only EA while keeping ElectrodeMix and the downstream optimization fixed. EA achieves the highest mean accuracy of 72.91%, gaining 2.19 pp over the common Linear probe and outperforming it for $237/306$ wearers. Removing the front stage reduces the gain to $+1.17$ pp, while Riemannian recentering reaches $+1.14$ pp and CORAL falls below the probe. These results support EA among the evaluated direct replacements for the front-stage alignment. The ViT-Tiny counterpart is reported in Appendix C.4.

What the inner loop adapts. Tab. 6 tests whether restricted meta-pretraining supports downstream use of the adapter. Opening the inner loop to the entire encoder reaches 34.41%, while removing the inner loop altogether reaches 46.49%, both below the 47.76% obtained when the inner loop is restricted to ElectrodeMix and the normalization affines. This comparison separates frozen-representation quality from final accuracy after adaptation.

8 Conclusion

We introduced a cross-wearer substitution diagnostic and a two-part few-shot sEMG adapter combining wearer-specific, label-free EA with supervised ElectrodeMix at the tokenizer. The diagnostic distinguishes transferable correction effects from the additional advantage of recipient-specific fitting. Under the stated linear mixing model, marginal spatial covariance identifies the symmetric polar factor but not the orthogonal one, providing structural motivation for separating covariance normalization from residual adaptation. ElectrodeMix uses $c^2 g$ learned adapter parameters beyond the classifier head and is supported by restricted meta-pretraining. Applying the substitution diagnostic to other adaptation methods also shows that similar own-calibration gains can correspond to distinct cross-wearer transfer structures. Cross-wearer substitution complements conventional own-calibration gains in evaluating few-shot personalization.

ETHICS STATEMENT

The authors have read and adhered to the ICLR Code of Ethics. This work uses only existing public sEMG datasets and involves no recruitment of new participants, prospective data collection, or intervention. Accordingly, the authors conducted no new human-subject study. Consent and research ethics procedures for the original data collections remain those reported by the respective dataset providers. Our study introduces no additional participant data beyond these archived recordings. For future real-world deployment, wearer-specific calibration data and personalized model parameters should be handled with appropriate privacy and security protections.

REPRODUCIBILITY STATEMENT

We provide the information needed to reproduce our experiments and analyses throughout the paper and appendix. Section 7.1 summarizes the experimental setup, datasets, evaluation protocol, and model configurations. Appendix A provides dataset preprocessing and splits, backbone and implementation details, and complete pretraining, deployment, and statistical protocols. Appendix B details the two-axis analysis, empirical whitening, cross-wearer depth analysis, donor-assignment sensitivity, and rotation-injection controls. Appendix C specifies the ElectrodeMix parameterization, contrastive objective, episode construction, augmentation operators, inner- and outer-loop optimization, and additional ablations. Additional few-shot and per-wearer results are reported in Appendix D. All datasets used in this work are publicly available and cited in the paper.

LLM USAGE STATEMENT

We used large language models (LLMs) solely for English grammar correction because the authors are non-native English speakers. LLMs were not used to develop research ideas, hypotheses, theoretical claims, methodology, experimental design, implementation, data processing, analysis or interpretation of results, tables, or other scientific content. The manuscript was written by the authors, and all LLM-assisted grammatical edits were reviewed and verified by the authors. We take full responsibility for the final content of this work.

REFERENCES

Marco E. Benalcázar, Lorena Isabel Barona López, Ángel Leonardo Valdivieso Caraguay, and J. Pablo Vásconez. Emg-epn-612 dataset, January 2021. URL https://doi.org/10.5281/zenodo.4421500.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Yikun Gu, Dapeng Yang, Qi Huang, Wei Yang, and Hong Liu. Robust emg pattern recognition in the presence of confounding factors: features, classifiers and adaptive learning. *Expert Systems with Applications*, 96:208–217, 2018.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019.

Nicholas J. Higham. Computing the polar decomposition—with applications. *SIAM Journal on Scientific and Statistical Computing*, 7(4):1160–1174, 1986. doi: 10.1137/0907079.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International conference on machine learning*, pp. 2790–2799. PMLR, 2019.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022.

Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. *Advances in Neural Information Processing Systems*, 34:2427–2440, 2021.

Néstor J. Jarque-Bou, Margarita Vergara, and Joaquín L. Sancho-Bru. Does exerting grasps involve a finite set of muscle patterns? a study of intra- and intersubject variability of forearm semg signals in seven grasp types. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, March 2024.

Bruna Junqueira, Bruno Aristimunha, Sylvain Chevallier, and Raphael Y de Camargo. A systematic evaluation of euclidean alignment with deep learning for eeg decoding. *Journal of Neural Engineering*, 21(3):036038, 2024.

Suguru Kanoga, Atsunori Kanemura, and Hideki Asoh. Are armband semg devices dense enough for long-term use?—sensor placement shifts cause significant reduction in recognition accuracy. *Biomedical Signal Processing and Control*, 60:101981, 2020.

Agnan Kessy, Alex Lewin, and Korbinian Strimmer. Optimal whitening and decorrelation. *The American Statistician*, 72(4):309–314, 2018. doi: 10.1080/00031305.2016.1277159.

N. Krilova, I. Kastalskiy, V. Kazantsev, V. A. Makarov, and S. Lobov. EMG Data for Gestures. UCI Machine Learning Repository, 2018. DOI: https://doi.org/10.24432/C5ZP5C.

Seulki Kyeong and Jung Kim. Enhanced torque estimation method from multi-channel surface electromyography compensating electrode location variation. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pp. 3965–3968, 2020. doi: 10.1109/EMBC44109.2020.9176498.

Hunmin Lee, Brian Lim, Ming Jiang, Zhi Yang, and Qi Zhao. Few-shot prototype adaptation for generalizable electromyography gesture recognition. *Scientific reports*, 16(1):12546, 2026.

Guofa Li, Delin Ouyang, Liu Yang, Qingkun Li, Kai Tian, Baiheng Wu, and Gang Guo. Cross-subject eeg linear domain adaption based on batch normalization and depthwise convolutional neural network. *Knowledge-Based Systems*, 280:111011, 2023a.

He Li, Shuxiang Guo, Dongdong Bu, and Hanze Wang. A two-stage ga-based semg feature selection method for user-independent continuous estimation of elbow angles. *IEEE Transactions on Instrumentation and Measurement*, January 2023b.

Yang Li and Jingyu Zhang. Utilizing statistical analysis for motion imagination classification in brain-computer interface systems. *Plos one*, 20(7):e0327121, 2025.

Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. *arXiv preprint arXiv:1603.04779*, 2016.

Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8602–8617, 2021.

Chuang Lin, Yuhao Wang, and Ming Dai. Robust gesture recognition based on attention-deep fast convolutional neural network and surface electromyographic signals. *Frontiers in Neuroscience*, 18:1306047, 2024.

Haiqi Liu, CL Chen, and Tong Zhang. Face: few-shot adapter with cross-view fusion for cross-subject eeg emotion recognition. *arXiv preprint arXiv:2503.18998*, 2025.

Haokun Liu, Derek Tam, Mohammed Muqeeth, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin A Raffel. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. *Advances in neural information processing systems*, 35:1950–1965, 2022.

Luca Manneschi, Matthew O. A. Ellis, and Elisa Donati. Beyond subject-specific models in dynamical human–machine interaction: Benchmarking and optimization strategies. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2026.

Lara McManus, Giuseppe De Vito, and Madeleine M Lowery. Analysis and biophysics of surface emg for physiotherapists and kinesiologists: Toward a common language with rehabilitation engineers. *Frontiers in neurology*, 11:576729, 2020.

Chulhong Min, Akhil Mathur, Alessandro Montanari, and Fahim Kawsar. An early characterisation of wearing variability on motion signals for wearables. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 166–168, 2019.

Alex Nichol and John Schulman. Reptile: a scalable metalearning algorithm. *arXiv preprint arXiv:1803.02999*, 2(3):4, 2018.

Kizito Nkurikiyeyezu, Anna Yokokubo, and Guillaume Lopez. Importance of individual differences in physiological-based stress recognition models. In *Intelligent Environments*, pp. 37–43, 2019.

Jaehoon Oh, Hyungjun Yoo, ChangHwan Kim, and Se-Young Yun. Boil: Towards representation change for few-shot learning. *arXiv preprint arXiv:2008.08882*, 2020.

Shotaro Okajima, Eduardo Iáñez, Hiroshi Yamasaki, Álvaro Costa García, Fady S Alnajjar, Noriaki Hattori, and Shingo Shimoda. Quantification of extent of muscle-skin shifting by traversal semg analysis using high-density semg sensor. In *2019 IEEE International Conference on Cyborg and Bionic Systems (CBS)*, pp. 272–277. IEEE, 2019.

Mehmet Akif Ozdemir, Deniz Hande Kisa, Onan Guren, Aytug Onan, and Aydin Akan. Emg based hand gesture recognition using deep learning. In *2020 Medical Technologies Congress (TIPTEKNO)*, pp. 1–4. IEEE, 2020.

Joao Pereira, Michael Alummoottil, Dimitrios Halatsis, and Dario Farina. Spatial adaptation layer: Interpretable domain adaptation for biosignal sensor array applications. *arXiv preprint arXiv:2409.08058*, 2024.

Stefano Pizzolato, Luca Tagliapietra, Matteo Cognolato, Monica Reggiani, Henning Müller, and Manfredo Atzori. Comparison of six electromyography acquisition setups on hand movement classification tasks. *PloS one*, 12(10):e0186132, 2017.

Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. Rapid learning or feature reuse? towards understanding the effectiveness of maml. *arXiv preprint arXiv:1909.09157*, 2019.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pp. 443–450. Springer, 2016.

Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pp. 10347–10357. PMLR, 2021.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021.

Zhenjin Xu, Linyong Shen, Jinwu Qian, and Zhen Zhang. Advanced hand gesture prediction robust to electrode shift with an arbitrary angle. *Sensors*, 20(4):1113, 2020.

Xiyuan Yang, Wenke Huang, and Mang Ye. Fedas: Bridging inconsistency in personalized federated learning. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11986–11995. IEEE, 2024.

Elad Ben Zaken, Yoav Goldberg, and Shauli Ravfogel. Bitfit: Simple parameter-efficient fine-tuning for transformer-based masked language-models. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 1–9, 2022.

Paolo Zanini, Marco Congedo, Christian Jutten, Salem Said, and Yannick Berthoumieu. Transfer learning: A riemannian geometry framework with applications to brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 65(5):1107–1116, 2017.

Anyuan Zhang, Ying He, Haowen Zheng, and Qunchao Lu. Data augmentation with fine-tuning strategy for enhanced myoelectric pattern recognition using deep learning. In *2025 18th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–6. IEEE, 2025.

Pengzhan Zhou, Yuepeng He, Yijun Zhai, Kaixin Gao, Chao Chen, Zhida Qin, Chong Zhang, and Songtao Guo. Fedah: Aggregated head for personalized federated learning. In *2024 IEEE Smart World Congress (SWC)*, pp. 250–257. IEEE, 2024.

Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*, 2024.

Luisa Zintgraf, Kyriacos Shiarli, Vitaly Kurin, Katja Hofmann, and Shimon Whiteson. Fast context adaptation via meta-learning. In *International conference on machine learning*, pp. 7693–7702. PMLR, 2019.

APPENDIX CONTENTS

A Experimental Details **15**

A.1 Datasets, preprocessing, and splits 15

A.2 Backbones and implementation 15

A.3 Pretraining, deployment, and statistics 16

B Analysis Details and Robustness **17**

B.1 Two-axis measurement 17

B.2 Empirical whitening and residual parameterization 18

B.3 Cross-wearer depth analysis and donor sensitivity 19

B.4 Rotation-injection and recovery controls 19

B.5 Cross-Adapter Generality of the Substitution Diagnostic 20

C Method Details and Ablations **21**

C.1 ElectrodeMix and group parameterization 21

C.2 Contrastive objective, episodes, and augmentation 21

C.3 Inner- and outer-loop optimization 23

C.4 Additional ablations 23

D Additional Results **25**

D.1 1-, 3-, and 5-shot adaptation results 25

D.2 Per-wearer adaptation results 26

A Experimental Details

A.1 Datasets, preprocessing, and splits

Table 7: **Datasets and evaluation protocols.** Wearers and Classes denote the numbers used for adaptation evaluation. Evaluation split specifies the calibration and held-out evaluation data, and Adapt. epoch denotes the downstream epoch at which accuracy is reported.

Condition	Wearers	Classes	Evaluation split	Adapt. epoch
UCI-EMG, within	34	6	session 1 → session 2	10
EPN612, within	306	6	session 0 → session 1	10
DB5-B1, within	10	53	repetitions 1–3 → 4–6	20
DB5-B2, within	10	53	repetitions 1–3 → 4–6	20
EPN612 → UCI-EMG	34	6	target archive held out	10
EPN612 → DB5-B1	10	53	target archive held out	20
EPN612 → DB5-B2	10	53	target archive held out	20

Evaluation protocols. We evaluate on three public sEMG archives under both within-dataset and cross-dataset adaptation protocols. The main experiments use 8-electrode ring inputs, allowing the same electrode-axis operator to be applied across conditions. As an additional scalability experiment, Tab. 23 combines the two DB5 rings into a 16-electrode array and evaluates ViT-Tiny. Tab. 7 summarizes the dataset-specific evaluation splits and downstream adaptation epochs. UCI-EMG and EPN612 use 10 adaptation epochs, whereas DB5 uses 20 because its 53-class classification task requires a longer downstream optimization budget.

Preprocessing and tokenization. All windows contain 200 samples recorded at 200 Hz over eight electrodes. For UCI-EMG and EPN612, each electrode is mean-centered and all electrodes share a single scale factor,

$$x_{j,t} \leftarrow \frac{x_{j,t} - \bar{x}_j}{\hat{\sigma}}, \qquad \hat{\sigma} = \sqrt{\frac{1}{cT-1}\sum_{j=1}^{c}\sum_{t=1}^{T}(x_{j,t} - \bar{x}_j)^2}, \tag{5}$$

which preserves relative amplitudes across electrodes. DB5 instead uses electrode-wise standardization over each recording. Both backbones use non-overlapping $(1, 10)$ patches, producing an 8×20 token grid with 160 tokens. The electrode axis is explicitly preserved through tokenization.

UCI-EMG splits. Wearers, rather than individual windows, define the pretraining splits. For UCI-EMG, wearers 11 and 30 are excluded because they are the only wearers containing gesture 7; removing them yields 34 wearers with a common six-class label set. The remaining wearers are permuted with `RandomState(0)` and divided into five folds of $7/7/7/7/6$ wearers. Each held-out wearer is evaluated using an encoder pretrained without that wearer, and a fixed subset of the remaining training wearers is reserved for validation.

EPN612 and DB5 splits. EPN612 provides two disjoint 306-wearer cohorts. Meta-pretraining uses the cohort derived from the archive's testing split, whereas adaptation evaluation uses the 306 wearers from the training split, with sessions 0 and 1 used for calibration and evaluation, respectively. For DB5, repetitions 1–3 are used for calibration and repetitions 4–6 for evaluation for both DB5-B1 and DB5-B2. The ten DB5 wearers are divided into five folds with held-out pairs $\{3, 9\}$, $\{5, 10\}$, $\{2, 7\}$, $\{4, 8\}$, and $\{1, 6\}$. Cross-dataset experiments pretrain on EPN612 and use UCI-EMG or DB5 only during deployment.

A.2 Backbones and implementation

Tab. 8 summarizes the configurations of the two encoder backbones. The two backbones use identical tokenization and embedding width; their principal architectural difference is the sequence-mixing operator.

Table 8: **Encoder configurations of ViM-Tiny and ViT-Tiny.**

	ViM-Tiny	ViT-Tiny
Embedding dimension	192	192
Blocks	24	12
Sequence mixing	bidirectional SSM	self-attention
State size / heads	16	3 heads
MLP ratio	–	4.0
Readout	mean	mean
ElectrodeMix groups	16	16
Encoder parameters	6.81M	5.37M

ElectrodeMix partitions the 192-dimensional embedding into $g = 16$ groups and applies one 8×8 electrode map to each group, giving $8^2 \times 16 = 1{,}024$ learned adapter parameters. The classifier head and the Euclidean Alignment matrix are excluded from this count. Parameter counts for the comparison methods are reported in Tables 1 and 2.

A.3 Pretraining, deployment, and statistics

Tab. 9 summarizes the episode composition and optimization settings used for meta-pretraining. The contrastive projection head consists of two layers, $192 \to 256 \to 128$, and its output is ℓ_2 normalized before Eq. (4). The projection head is discarded at deployment, where adaptation instead operates on the unnormalized 192-dimensional encoder representation.

Table 9: **Meta-pretraining configuration.** Support/query entries denote wearer windows used in each stage, while cross-wearer-negative entries denote the corresponding negative-window counts. Inner-loop settings apply to task adaptation, whereas outer-loop settings apply to meta-pretraining.

Setting	Value
Temperature τ	0.1
Projection head	$192 \to 256 \to 128$
Support / query windows	64 / 64
Cross-wearer negatives (support / query)	64 / 32
Inner-loop steps	10
Inner-loop learning rate	0.01
Inner-loop momentum	0.9
Weight decay	5×10^{-4}
Meta-gradient	first-order approximation
Outer optimizer	SGD
Outer learning rate	0.03
Outer schedule	cosine
Pretraining epochs	50

Stage-specific update sets. During meta-pretraining, the inner loop updates ElectrodeMix and the normalization affine parameters while keeping the encoder fixed within each inner step. The outer loop updates the encoder through the adapted parameters using a first-order approximation. At deployment, EA is recomputed from the recipient's unlabeled calibration windows, after which ElectrodeMix and a newly initialized classifier head are optimized using the recipient's labeled shots. The encoder and normalization affines remain fixed. Thus, the pretraining inner-loop parameter set is intentionally different from the downstream update set.

Few-shot deployment optimization. Table 10 reports the shared deployment settings. No method-specific learning-rate tuning is used in the main experiments. DB5 uses 20 epochs because its 53-class classifier head remains underfit after 10 epochs, whereas UCI-EMG and EPN612 use 10. Accuracy is reported at the fixed final epoch without early stopping or test-set model selection.

Table 10: **Few-shot deployment configuration.** Except for E2E scratch, all methods use the same pretrained checkpoint and calibration draws within each cell. Shared optimization settings apply to all applicable gradient-based methods. For paired values, the first corresponds to UCI-EMG and EPN612 and the second to DB5.

Setting	Value
Objective	Cross-entropy on labelled calibration windows
Shot budgets	$1, 3, 5, 10$ windows per class
Optimizer	SGD, momentum 0.9
Learning rate	0.05 for the head and adapted parameters
Method-specific LR tuning	None
Weight decay	5×10^{-4}
Learning-rate schedule	Constant
Epochs	10 / 20
Mini-batch size at $1/3/5/10$ shots	$1, 2, 4, 8$ / $7, 20, 34, 67$
Classifier head	Zero-initialized linear layer, $192 \to C$
Feature preprocessing	Calibration-mean centering; no variance scaling
Fine-tuning augmentation	None
Evaluation	Final epoch; no early stopping or test-set selection
Sampling	Five draws per wearer; seed $1000 + \mathrm{trial}$
EA-specific settings	
EA covariance	Mean of per-window spatial covariances
EA eigenvalue floor	$10^{-8}\lambda_{\max}$
EA shrinkage	0

Evaluation and statistics. We evaluate 1, 3, 5, and 10 labeled windows per class. The main paper reports the 10-shot setting. Each wearer is sampled five times, and the five draws are first averaged within each wearer. The wearer is therefore the statistical unit.

Accuracy is reported as the mean across wearers, with $\pm$ denoting the sample standard deviation across wearers. For paired differences Δ, $\pm$ denotes the half-width of the nominal 95% paired t confidence interval,

$$t_{0.975,n-1}\frac{\mathrm{sd}(d)}{\sqrt{n}}, \tag{6}$$

where d is the wearer-level paired difference. Fold-based results are pooled only after within-wearer averaging.

B Analysis Details and Robustness

B.1 Two-axis measurement

Table 11: **Raw-signal correction analysis on UCI-EMG.** Values denote normalized gap closure (%). Cross denotes a correction fitted on another wearer and applied to the recipient.

Correction	Own fit	Cross fit
Integer ring shift	5.2	-11.6
Fractional ring shift	18.4	-11.3
Normalization affine	64.9	+7.1
Free electrode map	75.6	-109.4

Raw-signal diagnostic. The raw-signal diagnostic is computed without an encoder. For each wearer, we construct class-conditional per-electrode RMS patterns, fit each correction on one session, and evaluate it on a separate session against the leave-one-out population of the same archive. Tab. 11 reports the corresponding own- and cross-wearer measurements on UCI-EMG. The fractional ring shift closes substantially more of the gap than the integer ring shift. Among the axis-

specific corrections, the normalization affine provides the largest closure, while the unconstrained free electrode map achieves the highest overall own-fit closure.

The correction families are evaluated separately to distinguish their transfer behavior. The normalization affine retains a positive cross-wearer component, whereas both ring-shift corrections and the free electrode map are harmful when applied from another wearer. The affine used here is a diagnostic correction for the amplitude axis, not a component combined with ElectrodeMix at deployment; the deployed method instead uses wearer-specific EA followed by ElectrodeMix. These measurements therefore characterize the transferability of each correction family rather than decomposing variation into mutually exclusive generic and wearer-specific factors.

Distance implementation. Each 200-sample window is divided into 25 non-overlapping eight-sample slices. Linear RMS is computed for each slice and electrode and then averaged across windows within each class, yielding a class$\times$slice$\times$electrode prototype. Distances are root-mean-square differences across all prototype entries. For each wearer, a correction is fitted from the calibration-session prototype to a leave-one-out reference formed from the evaluation-session prototypes of the other wearers, then applied to the wearer's evaluation-session prototype. U, $D(C)$, and F are the arithmetic means across wearers of the uncorrected, corrected, and between-session distances, respectively. At encoder depths, the same distance is applied to the corresponding class$\times 24\times$electrode representation. The $\frac{1}{2}$ term in A treats the between-session discrepancy as two equal session contributions, whereas $\frac{1}{N}$ is a heuristic finite-population correction. Thus, A is a normalization reference rather than a variance estimator or guaranteed lower bound.

B.2 Empirical whitening and residual parameterization

Residual after empirical whitening. The population model in Sec. 3 yields an exactly orthogonal residual after whitening. Deployment, however, uses a covariance estimated from a finite calibration set. Let

$$\bar{R}_i = R_i + E_i, \tag{7}$$

where R_i is the population covariance and E_i denotes the empirical covariance error. Define the residual transform after empirical whitening as

$$\widehat{T}_i = \bar{R}_i^{-1/2} W_i. \tag{8}$$

Under the source model in Section 3, $W_i W_i^\top = R_i$, and hence

$$\begin{aligned} \widehat{T}_i \widehat{T}_i^\top &= \bar{R}_i^{-1/2} R_i \bar{R}_i^{-1/2} \\ &= \bar{R}_i^{-1/2} (\bar{R}_i - E_i) \bar{R}_i^{-1/2} \\ &= I - \bar{R}_i^{-1/2} E_i \bar{R}_i^{-1/2}. \end{aligned} \tag{9}$$

Therefore, empirical whitening does not generally leave an exactly orthogonal residual. Its deviation from orthogonality is bounded by

$$\left\| \widehat{T}_i \widehat{T}_i^\top - I \right\|_2 \leq \frac{\|E_i\|_2}{\lambda_{\min}(\bar{R}_i)}. \tag{10}$$

This relation is not intended to attribute all residual variation to covariance-estimation error. Finite-sample estimation, deviations from the idealized linear source model, and task-specific variation can all contribute to the residual observed by the deployed model. We therefore use the polar decomposition to motivate separating whitening from residual adaptation, while leaving ElectrodeMix unconstrained rather than forcing the deployed map to be orthogonal.

Normalization of the representation diagnostic. For wearer i, the normalized closure is

$$\mathrm{closed}_i(C) = \frac{U_i - D_i(C)}{U_i - A_i}. \tag{11}$$

For two corrections C_a and C_b evaluated on the same wearer and at the same representation depth,

$$\mathrm{closed}_i(C_a) - \mathrm{closed}_i(C_b) = \frac{D_i(C_b) - D_i(C_a)}{U_i - A_i}. \tag{12}$$

Thus, when $U_i > A_i$, the common normalization preserves the ordering of the two corrections. Moreover,

$$\mathrm{closed}_i(C) = 0 \quad \Longleftrightarrow \quad D_i(C) = U_i, \tag{13}$$

so the zero crossing of a cross-wearer correction does not depend on the particular value of A_i. The normalization reference therefore rescales the reported magnitude but does not itself create the transfer-onset crossing. We do not interpret A_i as an oracle lower bound on $D_i(C)$.

B.3 Cross-wearer depth analysis and donor sensitivity

Table 12: **Donor-assignment sensitivity of the ViM-Tiny cross-wearer free map.** Each entry reports the minimum-to-maximum normalized gap closure (%) over twelve cyclic donor assignments. The reported ranges are not confidence intervals.

Depth	Block 10	Block 12	Block 14
Range of closure (%)	$[-12.0, -0.3]$	$[-0.3, +8.1]$	$[+0.9, +9.1]$

Donor-assignment sensitivity. Tab. 12 evaluates whether the cross-wearer free-map transition depends on the donor assignment. Across the twelve evaluated cyclic assignments, block 10 remains negative and block 14 remains positive, whereas block 12 spans zero. We therefore report $[\mathrm{b10}, \mathrm{b14}]$ as the observed sign-transition bracket across the evaluated assignments. The reported ranges characterize sensitivity to donor assignment rather than statistical uncertainty across wearers.

Depth-wise transfer. For ViM-Tiny, we evaluate the tokenizer and every second block, giving thirteen readout points. ViT-Tiny analogously provides seven points. The same correction parameterization is refitted at each depth.

For ViM-Tiny, the own-fitted fractional ring shift remains positive throughout the encoder, decreasing from 15.0% at the tokenizer to 5.5% after block 24. The cross-wearer ring shift remains negative at all thirteen depths. In contrast, the cross-wearer normalization affine remains positive throughout.

The unconstrained electrode map exhibits a different depth profile. Its cross-wearer closure changes from -78.4% at the tokenizer to $+13.8\%$ after block 24. The reference donor assignment crosses zero between blocks 10 and 12, while own-fit closure remains positive and higher than cross-wearer closure throughout. Thus, the zero crossing marks the onset of positive transfer for the transported map, not the disappearance of recipient-specific advantage.

The corresponding ViT-Tiny cross-wearer map remains negative at all seven measured depths. However, the own-minus-cross separation is largest at the tokenizer (172.8 pp versus 91.5 pp at the final block), and tokenizer placement yields the largest task-level gain in Tab. 4. Thus, the ViT-Tiny results also support placing ElectrodeMix at the tokenizer despite the absence of a zero crossing.

B.4 Rotation-injection and recovery controls

The depth analysis above estimates displacement from recorded signals. We additionally inject a known ring shift into the raw input and refit the inverse correction at each depth. This provides a ground-truth control for residual position sensitivity.

Table 13 reports recovery of the known one-electrode input shift across representation depths. Representation recovery measures the reduction in the injection-induced representation-space distance rather than the fraction of the shift magnitude recovered by $\hat{\delta}$. At the tokenizer, both backbones recover the injected shift exactly. For ViM-Tiny, the fitted inverse correction remains $\hat{\delta} = -0.940$ at block 12 while removing 42.2% of the induced representation-space gap.

B.5 Cross-Adapter Generality of the Substitution Diagnostic

The cross-wearer substitution diagnostic in Eq. 1 is defined for an arbitrary personalization increment $\Delta\Theta$ and is therefore not specific to ElectrodeMix. We apply the same own-versus-donor comparison to three additional adaptation methods on UCI-EMG with ViT-Tiny: SAL, BitFit, and

Table 13: **Recovery of an injected one-electrode shift.** A one-electrode circular shift is injected at the input, and an inverse ring correction is refitted at each depth. $\hat{\delta}$ denotes the fitted inverse shift. Representation recovery is the fraction of the injection-induced representation-space RMS distance removed on held-out scoring windows, $R^{\ell}_{\mathrm{ring}} = 100\left(1 - d^{\ell}_{\mathrm{corr}}/d^{\ell}_{0}\right)$, where fitting and scoring use disjoint halves of the windows. The final block corresponds to block 24 for ViM-Tiny and block 12 for ViT-Tiny.

Backbone	Tokenizer	Block 12	Block 18	Final block
ViM-Tiny, recovered $\hat{\delta}$	-1.000	-0.940	-0.680	-0.560
ViM-Tiny, representation recovery (%)	100.0	42.2	36.8	28.2
ViT-Tiny, recovered $\hat{\delta}$	-1.000	-0.980	–	-0.980
ViT-Tiny, representation recovery (%)	100.0	68.1	–	68.1

Table 14: **Cross-wearer substitution across adaptation methods.** 10-shot accuracy (%) on UCI-EMG with ViT-Tiny for 34 wearers. The frozen Linear probe is 83.75 ± 8.11. *Own* fits the adapter on the recipient's calibration shots, whereas *donor* fits it on another wearer's shots before freezing the adapter and refitting the classifier on the same recipient shots. Win counts wearers for whom the adapted model outperforms the frozen probe. Own–donor denotes the recipient-specific advantage.

	Own fit		Donor fit		
Adapter	Acc. (Win)	Δ probe	Acc. (Win)	Δ probe	Own–donor
ElectrodeMix	88.65 ± 6.21 (30/34)	$+4.90$	86.35 ± 7.71 (24/34)	$+2.60$	$+2.30$
SAL	82.73 ± 7.34 (14/34)	-1.02	79.77 ± 7.16 (4/34)	-3.98	$+2.96$
BitFit	83.74 ± 8.44 (19/34)	-0.01	85.24 ± 8.30 (23/34)	$+1.49$	-1.50
LoRA	83.92 ± 8.00 (14/34)	$+0.17$	83.96 ± 8.00 (21/34)	$+0.21$	-0.04

LoRA. All results use the same 34 wearers and 10-shot protocol as the corresponding main experiment. For each method, the adapter is fitted either on the recipient's own calibration shots or on another wearer's shots. The fitted adapter is then frozen, and the classifier head is refitted on identical recipient shots, so the own–donor contrast differs only in the wearer on which the adapter was fitted.

Own-versus-donor comparison. Tab. 14 shows that own-calibration gain and wearer specificity are distinct quantities. SAL and BitFit have similar own-calibration performance relative to the frozen probe, at -1.02 and -0.01 pp, respectively, yet their transfer structures differ substantially. Donor-fitted SAL decreases accuracy by 3.98 pp and improves over the probe for only 4/34 wearers, whereas donor-fitted BitFit gains 1.49 pp and improves for 23/34. Consequently, SAL exhibits a positive own–donor advantage of $+2.96$ pp, while BitFit exhibits a negative contrast of -1.50 pp. LoRA changes little under either fit, with an own–donor contrast of only -0.04 pp.

Table 15: **Paired contrasts between adaptation methods.** Differences are computed over the same 34 wearers. The left columns compare conventional own-calibration gains over the frozen probe, whereas the right columns compare the own–donor contrasts measured by the substitution diagnostic.

	Δ (own–probe)		Δ (own–donor)	
Pair	Difference	p	Difference	p
SAL vs. BitFit	-1.01	0.25	$+4.46$	3.1×10^{-6}
SAL vs. LoRA	-1.19	0.22	$+3.00$	7.8×10^{-4}
ElectrodeMix vs. SAL	$+5.92$	2.4×10^{-8}	-0.66	0.42
ElectrodeMix vs. BitFit	$+4.91$	2.4×10^{-7}	$+3.80$	2.7×10^{-7}

Paired contrasts across methods. The SAL–BitFit comparison makes this distinction explicit. Their own-calibration gains are not significantly different (-1.01 pp, $p = 0.25$), whereas their own–donor contrasts differ by 4.46 pp ($p = 3.1 \times 10^{-6}$; Tab. 15). Thus, two adapters that appear similar

under conventional own-calibration evaluation can exhibit significantly different own–donor contrasts. Conversely, wearer specificity alone does not imply usefulness: the own–donor contrasts of SAL and ElectrodeMix are not significantly different ($p = 0.42$), while SAL remains below the frozen probe in mean accuracy. These results show that cross-wearer substitution captures information not contained in conventional own-calibration gain and that this distinction extends beyond the proposed ElectrodeMix operator.

C Method Details and Ablations

C.1 ElectrodeMix and group parameterization

ElectrodeMix applies $I + M$ along the electrode axis of the token grid. The embedding dimension is partitioned into g groups, with an independent $c \times c$ map for each group. The group count g in this diagnostic is the same embedding-group parameter used by the deployed ElectrodeMix. For $c = 8$ and the default $g = 16$, the deployed adapter contains $8^2 \times 16 = 1{,}024$ trainable parameters.

Table 16: **Sensitivity to the ElectrodeMix group count.** Values denote normalized gap closure (%) as the shared group parameter g is varied.

Groups	Free map closure	Off-diagonal closure
1	91.3	41.0
2	84.6	36.7
4	83.1	35.5
8	75.0	27.8
16	75.6	26.8

Tab. 16 shows that both measurements change little between $g = 8$ and $g = 16$, indicating that the diagnostic has largely stabilized by the deployed setting. All downstream experiments fix $g = 16$, so this sweep evaluates the sensitivity of the representation diagnostic to the shared group parameterization rather than identifying the optimal group count for downstream accuracy.

C.2 Contrastive objective, episodes, and augmentation

The projection head maps the 192-dimensional encoder output through $192 \to 256 \to 128$, followed by ℓ_2 normalization. The normalized vectors are used only for the contrastive objective; the projection head is removed at deployment.

Cross-wearer negatives. Each episode contains 64 support and 64 query windows from the task wearer, while additional views from other wearers enter only as negative terms in the denominator. Tab. 17 reports a control in which this cross-wearer negative block is removed during meta-pretraining, with ViT-Tiny evaluated on UCI-EMG.

Table 17: **Ablation of cross-wearer negatives.** ViT-Tiny is evaluated on UCI-EMG with 34 wearers. The Linear probe is evaluated on the frozen encoder from the same pretraining configuration. Accuracy is reported as mean $\pm$ sample standard deviation across wearers. Win, shown in parentheses, denotes the number of wearers whose adapted accuracy exceeds the corresponding Linear probe, and Δ denotes the mean gain over that probe.

Cross-wearer negatives	Linear probe	Acc↑ (Win)	Δ over probe
Without	83.15	87.95 ± 5.94 (29/34)	+4.80
With (Proposed)	83.75	$\mathbf{88.65 \pm 6.21}$ (30/34)	+4.90

Tab. 17 compares meta-pretraining with and without cross-wearer negatives. Including the negative block increases the frozen Linear probe accuracy from 83.15% to 83.75% and the adapted accuracy from 87.95% to 88.65%. Relative to each configuration's own frozen Linear probe, however, the adaptation gains remain similar at $+4.80$ and $+4.90$ pp, with Win counts of $29/34$ and $30/34$, respectively. Thus, the cross-wearer negative block provides a modest overall improvement, while most of the few-shot adaptation gain remains without it.

Axis-specific augmentation. Among the pretraining augmentations, two operators explicitly perturb the wearer-variation axes analyzed in Sec. 4: global amplitude scaling and fractional ring shift along the electrode ring. Each stochastic operator is sampled independently for the two views of a positive pair.

Global gain. For a window $x \in \mathbb{R}^{c\times T}$, we sample a single multiplicative factor shared across all electrodes and time samples:

$$\log \kappa \sim \mathcal{U}\left(-\log(1+s), \log(1+s)\right), \qquad x'_{c,t} = \kappa x_{c,t}. \tag{14}$$

Equivalently, $\kappa \sim \mathrm{LogUniform}[(1+s)^{-1}, 1+s]$. We use $s = 0.2$ and apply this operator with probability 1, yielding $\kappa \in [0.833, 1.200]$. Because the same κ is applied to every electrode, relative amplitudes across channels are preserved:

$$\frac{x'_{c,t}}{x'_{c',t}} = \frac{x_{c,t}}{x_{c',t}}. \tag{15}$$

The augmentation therefore perturbs the overall signal level without introducing an artificial per-electrode gain pattern.

Fractional ring shift. To perturb electrode position, we sample a continuous displacement along the cyclic electrode axis:

$$\delta \sim \mathcal{U}(-\delta_{\max}, \delta_{\max}), \qquad k = \lfloor \delta \rfloor, \qquad f = \delta - k, \tag{16}$$

and linearly interpolate the two neighboring cyclic shifts:

$$x'_{c,t} = (1-f)x_{(c-k) \bmod c,t} + f x_{(c-k-1) \bmod c,t}. \tag{17}$$

We use $\delta_{\max} = 0.5$ electrodes and apply this operator with probability 0.5. For $f = 0$, Eq. 17 reduces to an integer cyclic roll; for fractional f, the operation mixes adjacent electrode positions.

Spatial-frequency response. Eq. 17 is a two-tap circular convolution along the electrode axis. For electrode-frequency index m, its frequency response is

$$H(m) = e^{-2\pi \mathrm{i} m k/c}\left[(1-f) + f e^{-2\pi \mathrm{i} m/c}\right], \tag{18}$$

with magnitude

$$|H(m)| = \sqrt{1 - 4f(1-f)\sin^2\left(\frac{\pi m}{c}\right)}. \tag{19}$$

Thus, fractional interpolation is not a pure circular shift. For the half-electrode case $f = 0.5$ and $c = 8$,

$$|H(m)| = \left|\cos\left(\frac{\pi m}{8}\right)\right|, \tag{20}$$

so higher electrode-axis frequencies are attenuated.

Augmentation versus measurement shift. The fractional interpolation above is used only for augmentation. The electrode-position diagnostic instead uses the exact DFT phase-ramp shift,

$$x' = \mathrm{Re}\,\mathcal{F}_c^{-1}\left[\mathcal{F}_c(x)_m e^{-2\pi \mathrm{i} \nu_m \delta/c}\right], \tag{21}$$

where ν_m denotes the signed electrode-axis DFT frequency. This operator satisfies $|H(m)| = 1$ for every m and therefore implements a pure, information-preserving circular shift. In contrast, Eq. 17 deliberately mixes neighboring electrodes. The measurement and augmentation operators therefore serve different purposes and are not interchangeable.

Additional augmentations. Impedance shift applies an independent zero-phase spectral tilt to each electrode. For window n and electrode j, $\beta_{n,j} \sim \mathcal{U}(-s, s)$ with $s = 0.4$, and

$$g_{n,j}(f) = \exp(\beta_{n,j}(2f-1)), \qquad \tilde{g}_{n,j}(f) = \frac{g_{n,j}(f)}{\frac{1}{F}\sum_{i=1}^{F} g_{n,j}(f_i)}, \tag{22}$$

for normalized frequency $f \in [0,1]$. The real-valued response $\tilde{g}_{n,j}$ is applied to the RFFT coefficients before the inverse transform, so the operation changes the per-electrode spectral shape without introducing phase delay.

Table 18: **Pretraining augmentation configuration.** p denotes the probability that each operator is applied to a view. The operators are listed in their application order.

Operation	p	Parameter
Random resized crop	0.8	scale $[0.4, 1.0]$
Impedance shift	0.8	$s = 0.4$
Time shift	1.0	± 4 samples
Global gain	1.0	$s = 0.2$
Jitter	1.0	$\sigma = 0.05$
Fractional ring shift	0.5	$\delta_{\max} = 0.5$

Random resized crop operates only along the temporal axis. For each window, a temporal fraction sampled from $\mathcal{U}(0.4, 1.0)$ is cropped and linearly resampled to the original length, with the same crop coordinates used for all electrodes. Jitter adds elementwise Gaussian noise,

$$x' = x + \sigma\epsilon, \qquad \epsilon \sim \mathcal{N}(0, 1), \qquad \sigma = 0.05, \tag{23}$$

where σ is expressed in the preprocessed signal units.

Tab. 18 summarizes the complete augmentation pipeline used during meta-pretraining. Global gain and fractional ring shift are the two operators that directly perturb the amplitude and electrode-position axes analyzed in the main paper; the remaining operators provide generic temporal, spectral, and noise perturbations.

C.3 Inner- and outer-loop optimization

Whitening protocol across stages. EA is not applied during meta-pretraining. Support, query, and cross-wearer-negative windows undergo the standard preprocessing and augmentation pipeline but are not whitened by Eq. 2. The inner loop adapts ElectrodeMix and the normalization affines; the latter provide task-specific normalization but neither estimate spatial covariance nor reproduce EA. At deployment, EA is computed once per recipient from the unlabeled calibration windows and applied to both calibration and evaluation windows before tokenization. The encoder and normalization affines are then frozen, while ElectrodeMix and a newly initialized classifier head are optimized using the recipient's labeled shots.

Meta-optimization and update sets. Let ψ denote ElectrodeMix together with the normalization affine parameters. Starting from the meta-pretrained initialization, the inner loop performs $K = 10$ momentum-SGD updates, for $k = 0, \ldots, K - 1$:

$$v^{(k+1)} = \mu v^{(k)} + \nabla_{\psi} \mathcal{L}_{\text{ssl}}\left(\psi^{(k)}\right), \tag{24}$$

$$\psi^{(k+1)} = \psi^{(k)} - \alpha v^{(k+1)}, \qquad v^{(0)} = 0, \tag{25}$$

where $\mu = 0.9$ is the inner-loop momentum. The encoder remains fixed throughout inner adaptation. The outer loop updates the encoder through the adapted ψ using a first-order approximation, without computing second-order meta-gradients.

At deployment, the update set changes: the normalization affine is fixed, while ElectrodeMix and a newly initialized classifier head are optimized using the recipient's labeled calibration windows. This difference is intentional; meta-pretraining is used to shape an encoder that supports adaptation through the proposed operator rather than to reproduce the downstream optimizer exactly.

C.4 Additional ablations

Front-stage alignment. We compare direct replacements for EA at the same front-stage position while keeping ElectrodeMix and the downstream adaptation procedure unchanged. Table 19 reports the ViT-Tiny results on EPN612; the corresponding ViM-Tiny comparison is reported in the main text.

Among the evaluated direct replacements, EA achieves the highest mean accuracy and the largest gain over the common Linear probe. Riemannian recentering remains close to the no-alignment

Table 19: **Which alignment goes in front.** 10-shot accuracy (%) on EPN612 with ViT-Tiny. EA is replaced by an alternative front-stage alignment or removed entirely, while ElectrodeMix and the downstream optimization remain unchanged. Δ denotes the mean gain over the common frozen Linear probe, and Win denotes the number of wearers for whom the method outperforms that probe.

Front	Linear probe	Acc↑ (Win)	Δ over probe
EA (Proposed)	**69.79**	**73.30 ± 9.64** (258/306)	**+3.51**
None	69.79	71.82 ± 9.82 (241/306)	+2.03
Riemannian	69.79	72.14 ± 9.65 (230/306)	+2.35
CORAL	69.79	63.38 ± 9.73 (17/306)	-6.41

condition, whereas CORAL substantially degrades performance. Together with the ViM-Tiny result in the main text, these results support EA as the front-stage alignment for the proposed method.

Table 20: **Orthogonal constraint on ElectrodeMix.** Results are 10-shot accuracies (%) on EPN612 with 306 wearers. Orthogonal + free composes an orthogonal correction with the free ElectrodeMix, whereas Orthogonal only replaces the free map. Accuracy is reported as mean ± sample standard deviation across wearers, Δ denotes the mean gain over the frozen probe, and Win denotes the number of wearers for whom the method outperforms that probe.

Backbone	Map	Acc↑	Δ over probe	Win
ViM-Tiny	Free (Proposed)	**72.91 ± 9.39**	**+2.19**	237/306
	Orthogonal + free	71.78 ± 9.46	+1.06	199/306
	Orthogonal only	69.37 ± 9.21	-1.35	117/306
ViT-Tiny	Free (Proposed)	**73.30 ± 9.64**	**+3.51**	258/306
	Orthogonal + free	72.80 ± 9.56	+3.01	255/306
	Orthogonal only	70.50 ± 9.37	+0.71	164/306

Orthogonal constraint. Tab. 20 compares the unconstrained ElectrodeMix with an orthogonal correction composed with the free map and an orthogonal map used in place of the free map. The orthogonal-only parameterization underperforms the free map on both backbones. Composing the orthogonal correction with the free map recovers part of this loss but remains below the unconstrained ElectrodeMix. This does not test the population identifiability statement itself; rather, it tests whether the idealized orthogonal residual should be imposed as a deployment constraint. These results indicate that the population model should be used as structural motivation for separating whitening from residual adaptation rather than as a literal constraint on the deployed map. Finite-sample covariance estimation, model mismatch, and task-specific variation can all make the residual non-orthogonal in practice.

Post-EA adapter comparison in the intra-dataset. Table 21 fixes EA and varies only the adapter fitted after it. EA + ElectrodeMix achieves the highest mean accuracy in all four conditions on both backbones. Relative to EA alone, it gains 1.79–3.70 pp on ViM-Tiny and 2.87–8.63 pp on ViT-Tiny. It also outperforms the strongest alternative post-EA adapter in every condition by 0.97–1.88 pp on ViM-Tiny and 2.48–2.95 pp on ViT-Tiny, while using only a 1.0K-parameter increment.

Post-EA adapter comparison in the cross-dataset. Table 22 applies the same control when the target archive is excluded from pretraining. Relative to EA alone, EA + ElectrodeMix improves accuracy by 1.64–3.71 pp on ViM-Tiny and 3.99–5.90 pp on ViT-Tiny. On ViM-Tiny, it outperforms EA + SAL on UCI-EMG and DB5-B2 but trails it by 0.17 pp on DB5-B1. On ViT-Tiny, it exceeds the strongest alternative in all three conditions by 0.48–3.44 pp. Thus, the benefit of ElectrodeMix after whitening persists across both backbones under cross-dataset deployment.

Extension to 16 electrodes. Because the main experiments use 8-electrode rings, Table 23 evaluates whether the same ElectrodeMix parameterization remains effective when the two DB5 rings are combined into a 16-electrode array. Doubling the electrode count increases the $c^2 g$ adapter size fourfold, from 1,024 to 4,096 parameters, which remains below 0.1% of the 5.41M parameters updated

Table 21: **Adapters behind the same whitener: intra-dataset 10-shot accuracy (%).** Every method applies EA to the input and differs only in the adapter fitted after it, so differences between rows are attributable to the adapter. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

			UCI-EMG	EPN612	DB5-B1	DB5-B2
Method	**Backbone**	**Params**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
EA	ViM-Tiny	0	86.43 ± 6.82 (27/34)	69.84 ± 9.25 (105/306)	32.79 ± 4.20 (8/10)	30.54 ± 5.32 (8/10)
	ViT-Tiny	0	85.24 ± 7.88 (21/34)	70.43 ± 9.55 (174/306)	39.13 ± 3.86 (8/10)	41.75 ± 4.88 (5/10)
EA + LoRA	ViM-Tiny	13.8K	86.54 ± 6.85 (27/34)	69.97 ± 9.31 (113/306)	32.00 ± 4.57 (6/10)	29.80 ± 5.21 (6/10)
	ViT-Tiny	9.2K	85.41 ± 7.74 (22/34)	70.82 ± 9.55 (197/306)	42.32 ± 3.00 (10/10)	44.48 ± 4.14 (10/10)
EA + BitFit	ViM-Tiny	41.9K	87.24 ± 6.29 (30/34)	70.72 ± 9.36 (158/306)	31.79 ± 3.15 (7/10)	29.68 ± 5.92 (7/10)
	ViT-Tiny	25.7K	85.90 ± 6.96 (24/34)	68.81 ± 9.76 (113/306)	44.06 ± 2.30 (10/10)	45.32 ± 4.32 (9/10)
EA + SAL	ViM-Tiny	72	88.41 ± 6.20 (31/34)	71.94 ± 9.34 (188/306)	33.44 ± 4.14 (9/10)	31.01 ± 5.84 (7/10)
	ViT-Tiny	72	85.94 ± 6.93 (23/34)	70.71 ± 9.86 (178/306)	44.83 ± 3.49 (10/10)	44.94 ± 5.14 (10/10)
EA + ElectrodeMix	ViM-Tiny	1.0K	**90.13 ± 6.02** (32/34)	**72.91 ± 9.39** (237/306)	**34.58 ± 4.42** (8/10)	**32.89 ± 4.87** (9/10)
	ViT-Tiny	1.0K	**88.65 ± 6.21** (30/34)	**73.30 ± 9.64** (258/306)	**47.76 ± 2.85** (10/10)	**48.27 ± 4.84** (10/10)

Table 22: **Adapters behind the same whitener: cross-dataset 10-shot accuracy (%).** Every encoder is pretrained on EPN612 and deployed on a dataset unseen during pretraining. Every method applies EA to the input and differs only in the adapter fitted after it, so differences between rows are attributable to the adapter. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

				Test set		
				UCI-EMG	DB5-B1	DB5-B2
Method	**Backbone**	**Params**	**Train set**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
EA	ViM-Tiny	0	EPN612	83.22 ± 8.01 (30/34)	31.54 ± 2.44 (10/10)	32.48 ± 3.34 (10/10)
	ViT-Tiny	0		78.36 ± 7.68 (30/34)	32.26 ± 1.83 (10/10)	33.16 ± 3.49 (10/10)
EA + LoRA	ViM-Tiny	13.8K	EPN612	83.24 ± 8.04 (30/34)	29.74 ± 2.56 (10/10)	30.61 ± 3.37 (10/10)
	ViT-Tiny	9.2K		79.24 ± 7.61 (32/34)	32.45 ± 2.23 (10/10)	33.70 ± 3.54 (10/10)
EA + BitFit	ViM-Tiny	41.9K	EPN612	84.22 ± 7.85 (32/34)	31.48 ± 2.81 (10/10)	32.92 ± 3.73 (10/10)
	ViT-Tiny	25.7K		80.82 ± 7.12 (31/34)	32.79 ± 2.12 (10/10)	34.45 ± 3.40 (10/10)
EA + SAL	ViM-Tiny	72	EPN612	84.29 ± 7.46 (28/34)	**33.35 ± 1.51** (10/10)	34.53 ± 4.05 (10/10)
	ViT-Tiny	72		77.80 ± 7.05 (20/34)	36.44 ± 2.47 (10/10)	36.52 ± 3.52 (10/10)
EA + ElectrodeMix	ViM-Tiny	1.0K	EPN612	**86.93 ± 7.42** (34/34)	33.18 ± 2.36 (10/10)	**34.83 ± 4.00** (10/10)
	ViT-Tiny	1.0K		**84.26 ± 8.13** (33/34)	**36.92 ± 2.55** (10/10)	**37.15 ± 3.41** (10/10)

by Full FT. The full method reaches $51.43 \pm 3.98\%$, outperforming the Linear probe by 11.40 pp and the second-best SAL by 3.84 pp, with gains for all ten wearers. These results show that ElectrodeMix remains effective at 16 electrodes while adding only a minor parameter increment relative to the full model.

D Additional Results

D.1 1-, 3-, and 5-shot adaptation results

We additionally evaluate the same comparison methods using 1, 3, and 5 labeled windows per class. Intra-dataset and cross-dataset results are reported separately to distinguish adaptation within the pretraining archive from deployment to a target archive excluded from encoder pretraining.

1-shot results. Tables 24 and 25 report the 1-shot results. No single method dominates the intra-dataset setting: T3A leads UCI-EMG with ViM-Tiny, EA leads both DB5 rings with ViM-Tiny, and SAL leads DB5-B2 with ViT-Tiny, while the proposed method leads the remaining four cells. In

Table 23: **16-electrode 10-shot adaptation accuracy (%).** Both DB5 armbands are treated as a single 16-electrode array with ViT-Tiny; the wearers, sessions, and folds are those of the DB5 columns in Table 1, and ElectrodeMix uses a single 16×16 map per embedding group. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment at 16 electrodes. $\pm$ denotes the sample standard deviation across wearers. **Bold** indicates the best result, and underline indicates the second-best result.

			DB5, 16 electrodes
Method	**Type**	**Params**	**Acc↑ (Win)**
Linear probe	No adapt.	0	40.03 ± 2.90 (0/10)
Full FT	Full adapt.	5.41M	44.52 ± 4.40 (9/10)
BitFit	PEFT	25.7K	45.60 ± 3.96 (10/10)
LoRA	PEFT	9.2K	42.93 ± 3.29 (10/10)
T3A	TTA	0	39.53 ± 2.10 (4/10)
CORAL	Align	9.2K	9.89 ± 0.99 (0/10)
SAL	Align	272	<u>47.59 ± 3.03</u> (10/10)
EA	Align	0	41.10 ± 2.69 (9/10)
AdaBN	Align	0	45.03 ± 3.38 (10/10)
Proposed	Align+PEFT	4.1K	**51.43 ± 3.98** (10/10)

the cross-dataset setting, the proposed method leads four of six cells, while EA leads DB5-B1 and DB5-B2 with ViT-Tiny.

3-shot results. Tables 26 and 27 report the 3-shot results. The proposed method leads every cross-dataset cell on both backbones and seven of eight intra-dataset cells. The exception is DB5-B1 with ViM-Tiny, where T3A achieves the highest mean accuracy. Among the remaining intra-dataset baselines, AdaBN ranks second on EPN612 with ViM-Tiny, while SAL ranks second on both DB5 rings with ViT-Tiny.

5-shot results. Tables 28 and 29 report the 5-shot results. The proposed method again leads every cross-dataset cell and seven of eight intra-dataset cells, with DB5-B1 on ViM-Tiny remaining the only exception. AdaBN ranks second on EPN612 with ViM-Tiny, and SAL ranks second on both DB5 rings with ViT-Tiny. Overall, alignment baselines remain particularly competitive in the 1-shot setting, whereas the benefit of supervised ElectrodeMix becomes more consistent as the labeled calibration budget increases.

D.2 Per-wearer adaptation results

To examine whether the average gain is concentrated in a small subset of wearers, we report 10-shot UCI-EMG results for individual wearers. Tab. 30 compares the frozen Linear probe, EA, and the full EA+ElectrodeMix configuration on ViM-Tiny. The table shows that EA+ElectrodeMix outperforms the frozen Linear probe for 32 of 34 wearers. The gains span a broad range of wearers rather than being driven by a few large improvements, although two wearers show negative mean differences.

Table 24: **Intra-dataset 1-shot adaptation accuracy (%).** Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the Linear probe, which serves as the reference for all comparisons. Params denotes the per-wearer increment. E2E scratch trains the encoder from random initialization using the same 1 shot per class. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

				UCI-EMG	EPN612	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	65.58 ± 9.49 (0/34)	52.82 ± 9.73 (0/306)	22.52 ± 3.60 (0/10)	21.65 ± 4.00 (0/10)
		ViT-Tiny	0	68.95 ± 9.55 (0/34)	52.19 ± 9.69 (0/306)	27.24 ± 2.37 (0/10)	29.67 ± 3.41 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	17.94 ± 1.52 (0/34)	24.33 ± 4.05 (0/306)	3.69 ± 0.67 (0/10)	3.45 ± 0.49 (0/10)
		ViT-Tiny	5.37M	16.84 ± 0.78 (0/34)	20.28 ± 3.72 (0/306)	3.25 ± 0.53 (0/10)	3.33 ± 0.59 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	56.12 ± 8.75 (1/34)	26.05 ± 4.78 (0/306)	13.50 ± 2.90 (0/10)	13.30 ± 3.23 (0/10)
		ViT-Tiny	5.37M	43.76 ± 8.66 (0/34)	38.63 ± 5.96 (9/306)	17.16 ± 2.31 (0/10)	18.76 ± 2.98 (0/10)
BitFit	PEFT	ViM-Tiny	41.9K	62.47 ± 9.21 (5/34)	47.97 ± 10.01 (50/306)	20.55 ± 3.66 (1/10)	19.82 ± 4.51 (0/10)
		ViT-Tiny	25.7K	67.64 ± 9.40 (16/34)	48.04 ± 9.67 (43/306)	25.24 ± 1.92 (1/10)	25.87 ± 3.54 (0/10)
LoRA	PEFT	ViM-Tiny	13.8K	65.59 ± 9.49 (11/34)	52.82 ± 9.74 (77/306)	21.90 ± 3.46 (2/10)	21.09 ± 4.00 (0/10)
		ViT-Tiny	9.2K	69.00 ± 9.52 (15/34)	52.20 ± 9.69 (135/306)	25.20 ± 1.72 (1/10)	27.23 ± 3.02 (0/10)
T3A	TTA	ViM-Tiny	0	**72.97 ± 11.11** (34/34)	52.91 ± 10.35 (152/306)	24.38 ± 3.73 (10/10)	22.62 ± 4.91 (9/10)
		ViT-Tiny	0	73.27 ± 10.54 (32/34)	52.29 ± 9.92 (154/306)	26.11 ± 2.17 (0/10)	29.44 ± 3.73 (5/10)
CORAL	Align	ViM-Tiny	9.2K	40.03 ± 7.23 (0/34)	46.05 ± 7.81 (29/306)	6.28 ± 1.33 (0/10)	6.18 ± 1.44 (0/10)
		ViT-Tiny	9.2K	37.75 ± 9.37 (0/34)	43.11 ± 7.15 (13/306)	9.58 ± 1.41 (0/10)	8.28 ± 1.25 (0/10)
SAL	Align	ViM-Tiny	72	65.18 ± 7.96 (16/34)	47.11 ± 9.38 (40/306)	20.78 ± 5.53 (2/10)	18.08 ± 5.55 (0/10)
		ViT-Tiny	72	62.72 ± 7.79 (6/34)	44.26 ± 8.80 (22/306)	28.47 ± 2.11 (8/10)	**30.73 ± 2.99** (9/10)
EA	Align	ViM-Tiny	0	68.10 ± 9.21 (29/34)	52.35 ± 9.20 (127/306)	**25.21 ± 1.80** (9/10)	**24.17 ± 4.18** (10/10)
		ViT-Tiny	0	70.71 ± 9.30 (21/34)	53.27 ± 9.12 (193/306)	28.16 ± 2.14 (8/10)	30.41 ± 3.85 (5/10)
AdaBN	Align	ViM-Tiny	0	64.66 ± 9.54 (3/34)	52.79 ± 9.56 (157/306)	21.81 ± 3.23 (3/10)	21.06 ± 4.30 (2/10)
		ViT-Tiny	0	67.32 ± 9.15 (9/34)	51.64 ± 9.53 (113/306)	25.07 ± 1.68 (1/10)	26.85 ± 3.65 (0/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	71.73 ± 8.94 (33/34)	**53.09 ± 9.27** (155/306)	24.59 ± 2.54 (10/10)	23.73 ± 4.32 (10/10)
		ViT-Tiny	1.0K	**73.31 ± 8.21** (26/34)	**53.66 ± 9.34** (208/306)	**28.75 ± 1.95** (9/10)	30.25 ± 3.95 (6/10)

Table 25: **Cross-dataset 1-shot adaptation accuracy (%).** Source-pretrained methods pretrain on EPN612 and deploy on an archive held out from pretraining by construction. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

					Test set		
					UCI-EMG	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Train set**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	EPN612	66.35 ± 7.50 (0/34)	16.60 ± 2.62 (0/10)	16.87 ± 2.51 (0/10)
		ViT-Tiny	0		59.61 ± 8.78 (0/34)	15.55 ± 2.53 (0/10)	15.29 ± 2.22 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	–	17.86 ± 1.36 (0/34)	3.78 ± 0.68 (0/10)	3.64 ± 0.42 (0/10)
		ViT-Tiny	5.37M		16.81 ± 0.53 (0/34)	3.36 ± 0.51 (0/10)	3.26 ± 0.41 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	EPN612	30.73 ± 5.36 (0/34)	11.14 ± 1.86 (0/10)	12.01 ± 2.37 (0/10)
		ViT-Tiny	5.37M		33.31 ± 6.35 (0/34)	9.03 ± 1.53 (0/10)	9.52 ± 2.12 (0/10)
BitFit	PEFT	ViM-Tiny	41.9K	EPN612	67.31 ± 7.99 (23/34)	16.77 ± 2.65 (5/10)	16.86 ± 2.32 (4/10)
		ViT-Tiny	25.7K		62.14 ± 8.40 (27/34)	13.84 ± 2.51 (0/10)	13.64 ± 2.35 (0/10)
LoRA	PEFT	ViM-Tiny	13.8K	EPN612	66.36 ± 7.54 (15/34)	16.07 ± 2.69 (1/10)	16.67 ± 2.25 (5/10)
		ViT-Tiny	9.2K		59.87 ± 8.63 (22/34)	14.55 ± 2.62 (0/10)	14.31 ± 1.99 (0/10)
T3A	TTA	ViM-Tiny	0	EPN612	70.15 ± 9.12 (28/34)	17.26 ± 2.88 (8/10)	17.41 ± 2.47 (9/10)
		ViT-Tiny	0		62.30 ± 9.53 (30/34)	15.28 ± 2.76 (4/10)	15.41 ± 2.19 (6/10)
CORAL	Align	ViM-Tiny	9.2K	EPN612	42.58 ± 6.25 (0/34)	12.54 ± 2.13 (0/10)	11.75 ± 1.84 (0/10)
		ViT-Tiny	9.2K		29.23 ± 4.37 (0/34)	9.36 ± 1.06 (0/10)	9.24 ± 1.32 (0/10)
SAL	Align	ViM-Tiny	72	EPN612	56.97 ± 6.93 (1/34)	15.83 ± 2.57 (1/10)	16.71 ± 3.29 (4/10)
		ViT-Tiny	72		45.49 ± 7.21 (1/34)	15.51 ± 1.78 (6/10)	15.87 ± 2.60 (7/10)
EA	Align	ViM-Tiny	0	EPN612	69.47 ± 7.41 (27/34)	22.56 ± 1.64 (10/10)	23.20 ± 2.73 (10/10)
		ViT-Tiny	0		60.52 ± 7.12 (20/34)	**20.81 ± 2.08** (10/10)	**21.31 ± 2.86** (10/10)
AdaBN	Align	ViM-Tiny	0	EPN612	66.94 ± 7.89 (24/34)	16.94 ± 2.68 (8/10)	17.52 ± 2.49 (8/10)
		ViT-Tiny	0		60.32 ± 8.00 (22/34)	13.23 ± 2.15 (0/10)	13.06 ± 1.98 (0/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	EPN612	**71.60 ± 7.97** (31/34)	**22.93 ± 1.62** (10/10)	**23.38 ± 2.99** (10/10)
		ViT-Tiny	1.0K		**64.10 ± 7.37** (30/34)	20.50 ± 2.21 (10/10)	21.26 ± 3.31 (10/10)

Table 26: **Intra-dataset 3-shot adaptation accuracy (%).** Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the Linear probe, which serves as the reference for all comparisons. Params denotes the per-wearer increment. E2E scratch trains the encoder from random initialization using the same 3 shots per class. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

				UCI-EMG	EPN612	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	78.40 ± 8.48 (0/34)	63.69 ± 9.77 (0/306)	27.66 ± 3.52 (0/10)	26.80 ± 5.09 (0/10)
		ViT-Tiny	0	78.67 ± 9.23 (0/34)	62.33 ± 10.01 (0/306)	34.53 ± 2.27 (0/10)	37.02 ± 4.04 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	20.69 ± 2.53 (0/34)	26.11 ± 4.08 (0/306)	4.93 ± 0.95 (0/10)	4.75 ± 0.79 (0/10)
		ViT-Tiny	5.37M	16.80 ± 0.85 (0/34)	26.02 ± 5.51 (0/306)	4.30 ± 0.90 (0/10)	4.49 ± 0.62 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	70.13 ± 8.95 (3/34)	27.31 ± 5.01 (0/306)	23.60 ± 3.43 (0/10)	23.23 ± 5.04 (1/10)
		ViT-Tiny	5.37M	52.07 ± 7.54 (0/34)	43.24 ± 6.43 (0/306)	27.45 ± 2.39 (0/10)	28.46 ± 3.93 (0/10)
BitFit	PEFT	ViM-Tiny	41.9K	78.81 ± 7.66 (16/34)	55.12 ± 11.67 (32/306)	25.72 ± 3.90 (0/10)	24.80 ± 5.54 (0/10)
		ViT-Tiny	25.7K	76.24 ± 9.47 (12/34)	55.78 ± 10.39 (22/306)	35.20 ± 2.53 (7/10)	37.09 ± 3.89 (5/10)
LoRA	PEFT	ViM-Tiny	13.8K	78.43 ± 8.48 (17/34)	63.75 ± 9.77 (151/306)	27.28 ± 3.48 (1/10)	26.42 ± 4.99 (1/10)
		ViT-Tiny	9.2K	78.80 ± 9.14 (19/34)	62.24 ± 10.09 (153/306)	35.12 ± 2.24 (6/10)	37.97 ± 3.68 (9/10)
T3A	TTA	ViM-Tiny	0	81.19 ± 9.44 (28/34)	60.38 ± 10.10 (26/306)	**30.57 ± 4.33** (10/10)	28.68 ± 6.05 (9/10)
		ViT-Tiny	0	80.07 ± 9.71 (23/34)	59.27 ± 10.13 (40/306)	32.69 ± 2.11 (1/10)	36.22 ± 4.01 (3/10)
CORAL	Align	ViM-Tiny	9.2K	52.38 ± 7.48 (0/34)	54.31 ± 9.04 (1/306)	7.69 ± 1.32 (0/10)	7.16 ± 1.65 (0/10)
		ViT-Tiny	9.2K	47.65 ± 11.30 (0/34)	49.93 ± 8.02 (4/306)	13.60 ± 2.12 (0/10)	11.89 ± 1.96 (0/10)
SAL	Align	ViM-Tiny	72	77.79 ± 7.23 (15/34)	57.48 ± 10.90 (28/306)	28.39 ± 4.36 (8/10)	26.40 ± 5.96 (4/10)
		ViT-Tiny	72	73.96 ± 8.66 (8/34)	54.56 ± 10.73 (26/306)	38.37 ± 3.03 (10/10)	39.37 ± 4.30 (10/10)
EA	Align	ViM-Tiny	0	81.13 ± 8.15 (28/34)	62.75 ± 9.83 (104/306)	30.18 ± 2.92 (9/10)	28.39 ± 4.88 (9/10)
		ViT-Tiny	0	79.91 ± 8.98 (22/34)	62.78 ± 9.75 (180/306)	35.60 ± 3.10 (8/10)	37.79 ± 4.86 (6/10)
AdaBN	Align	ViM-Tiny	0	78.27 ± 8.30 (10/34)	64.06 ± 9.76 (174/306)	26.61 ± 3.14 (0/10)	26.10 ± 5.17 (3/10)
		ViT-Tiny	0	76.89 ± 8.37 (10/34)	60.95 ± 10.13 (84/306)	34.94 ± 2.21 (8/10)	37.60 ± 3.93 (8/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	**85.19 ± 7.19** (33/34)	**64.72 ± 10.05** (190/306)	29.70 ± 3.56 (9/10)	**28.90 ± 5.01** (10/10)
		ViT-Tiny	1.0K	**83.85 ± 6.90** (28/34)	**64.56 ± 9.96** (231/306)	**40.25 ± 2.74** (10/10)	**41.12 ± 4.45** (10/10)

Table 27: **Cross-dataset 3-shot adaptation accuracy (%).** Source-pretrained methods pretrain on EPN612 and deploy on an archive held out from pretraining by construction. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

					Test set		
					UCI-EMG	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Train set**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	EPN612	75.62 ± 7.94 (0/34)	20.98 ± 3.20 (0/10)	21.15 ± 2.71 (0/10)
		ViT-Tiny	0		69.72 ± 7.95 (0/34)	20.70 ± 2.73 (0/10)	20.44 ± 2.85 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	–	20.88 ± 2.48 (0/34)	4.96 ± 1.01 (0/10)	4.60 ± 0.78 (0/10)
		ViT-Tiny	5.37M		17.03 ± 0.97 (0/34)	4.35 ± 1.11 (0/10)	4.32 ± 0.86 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	EPN612	31.86 ± 6.78 (0/34)	19.46 ± 3.42 (1/10)	20.50 ± 4.44 (3/10)
		ViT-Tiny	5.37M		36.88 ± 6.07 (0/34)	15.10 ± 2.65 (0/10)	15.75 ± 3.19 (0/10)
BitFit	PEFT	ViM-Tiny	41.9K	EPN612	76.17 ± 8.04 (22/34)	22.18 ± 3.27 (10/10)	22.31 ± 3.67 (8/10)
		ViT-Tiny	25.7K		70.85 ± 7.47 (19/34)	20.86 ± 3.07 (5/10)	21.18 ± 3.65 (7/10)
LoRA	PEFT	ViM-Tiny	13.8K	EPN612	75.55 ± 7.93 (17/34)	20.70 ± 2.96 (4/10)	20.84 ± 2.55 (4/10)
		ViT-Tiny	9.2K		70.35 ± 7.77 (25/34)	20.50 ± 2.58 (3/10)	20.25 ± 2.81 (4/10)
T3A	TTA	ViM-Tiny	0	EPN612	77.24 ± 8.84 (24/34)	21.48 ± 3.21 (8/10)	22.14 ± 2.91 (8/10)
		ViT-Tiny	0		71.00 ± 9.14 (24/34)	19.09 ± 2.80 (0/10)	19.29 ± 2.68 (0/10)
CORAL	Align	ViM-Tiny	9.2K	EPN612	54.24 ± 7.54 (0/34)	16.19 ± 2.39 (1/10)	15.85 ± 2.26 (0/10)
		ViT-Tiny	9.2K		40.23 ± 8.37 (0/34)	14.29 ± 2.07 (0/10)	14.71 ± 2.53 (0/10)
SAL	Align	ViM-Tiny	72	EPN612	71.32 ± 7.54 (8/34)	22.35 ± 3.02 (8/10)	23.32 ± 4.21 (9/10)
		ViT-Tiny	72		58.93 ± 7.81 (2/34)	23.85 ± 1.72 (10/10)	23.59 ± 3.73 (10/10)
EA	Align	ViM-Tiny	0	EPN612	79.80 ± 7.41 (32/34)	27.67 ± 2.47 (10/10)	28.59 ± 3.23 (10/10)
		ViT-Tiny	0		72.36 ± 7.28 (30/34)	27.39 ± 2.33 (10/10)	27.99 ± 3.63 (10/10)
AdaBN	Align	ViM-Tiny	0	EPN612	76.49 ± 7.99 (25/34)	21.85 ± 3.29 (8/10)	22.00 ± 3.16 (7/10)
		ViT-Tiny	0		69.36 ± 7.98 (15/34)	19.84 ± 2.85 (1/10)	19.79 ± 3.13 (1/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	EPN612	**83.27 ± 6.99** (33/34)	**29.55 ± 2.19** (10/10)	**30.18 ± 3.66** (10/10)
		ViT-Tiny	1.0K		**77.70 ± 7.94** (33/34)	**29.11 ± 2.20** (10/10)	**29.75 ± 3.92** (10/10)

Table 28: **Intra-dataset 5-shot adaptation accuracy (%).** Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the Linear probe, which serves as the reference for all comparisons. Params denotes the per-wearer increment. E2E scratch trains the encoder from random initialization using the same 5 shots per class. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

				UCI-EMG	EPN612	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	81.48 ± 7.92 (0/34)	67.18 ± 9.65 (0/306)	29.25 ± 3.76 (0/10)	28.09 ± 5.01 (0/10)
		ViT-Tiny	0	81.38 ± 8.50 (0/34)	65.92 ± 9.91 (0/306)	36.69 ± 2.35 (0/10)	39.30 ± 4.35 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	25.81 ± 3.94 (0/34)	31.50 ± 5.78 (0/306)	5.49 ± 1.29 (0/10)	5.25 ± 0.71 (0/10)
		ViT-Tiny	5.37M	16.72 ± 0.78 (0/34)	29.04 ± 5.57 (0/306)	4.82 ± 0.69 (0/10)	4.89 ± 0.80 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	79.52 ± 7.17 (8/34)	36.50 ± 7.80 (0/306)	29.15 ± 4.46 (5/10)	28.31 ± 5.24 (6/10)
		ViT-Tiny	5.37M	65.54 ± 8.00 (0/34)	49.79 ± 8.09 (0/306)	32.52 ± 2.46 (0/10)	34.08 ± 3.94 (0/10)
BitFit	PEFT	ViM-Tiny	41.9K	81.55 ± 7.49 (14/34)	62.44 ± 11.08 (51/306)	27.53 ± 3.74 (1/10)	26.28 ± 5.45 (1/10)
		ViT-Tiny	25.7K	80.71 ± 8.07 (18/34)	60.91 ± 10.78 (28/306)	39.19 ± 2.73 (9/10)	40.83 ± 4.30 (7/10)
LoRA	PEFT	ViM-Tiny	13.8K	81.52 ± 7.91 (16/34)	67.28 ± 9.65 (152/306)	28.94 ± 3.79 (2/10)	27.81 ± 4.91 (1/10)
		ViT-Tiny	9.2K	81.46 ± 8.44 (19/34)	66.12 ± 9.95 (173/306)	37.95 ± 2.20 (7/10)	41.11 ± 4.15 (9/10)
T3A	TTA	ViM-Tiny	0	82.99 ± 8.37 (24/34)	62.36 ± 10.04 (9/306)	**32.37 ± 4.99** (9/10)	30.45 ± 6.31 (8/10)
		ViT-Tiny	0	81.98 ± 8.74 (19/34)	61.51 ± 10.07 (10/306)	35.07 ± 2.20 (1/10)	38.82 ± 4.43 (5/10)
CORAL	Align	ViM-Tiny	9.2K	56.85 ± 8.46 (0/34)	58.08 ± 9.54 (4/306)	8.30 ± 1.69 (0/10)	7.45 ± 1.80 (0/10)
		ViT-Tiny	9.2K	55.29 ± 10.73 (0/34)	54.98 ± 9.17 (4/306)	15.63 ± 1.76 (0/10)	13.08 ± 2.39 (0/10)
SAL	Align	ViM-Tiny	72	82.93 ± 6.28 (20/34)	62.60 ± 10.89 (61/306)	30.64 ± 3.90 (9/10)	28.40 ± 5.80 (5/10)
		ViT-Tiny	72	79.05 ± 7.87 (10/34)	58.81 ± 11.16 (28/306)	41.66 ± 3.20 (10/10)	42.26 ± 4.89 (10/10)
EA	Align	ViM-Tiny	0	84.27 ± 7.12 (27/34)	66.15 ± 9.49 (102/306)	31.81 ± 3.32 (9/10)	29.65 ± 5.32 (7/10)
		ViT-Tiny	0	82.90 ± 7.88 (21/34)	66.38 ± 9.78 (172/306)	37.69 ± 3.23 (7/10)	40.03 ± 5.12 (7/10)
AdaBN	Align	ViM-Tiny	0	81.74 ± 7.83 (20/34)	67.85 ± 9.59 (202/306)	28.12 ± 3.51 (0/10)	27.48 ± 5.31 (2/10)
		ViT-Tiny	0	80.40 ± 8.13 (12/34)	65.32 ± 10.08 (118/306)	38.90 ± 2.66 (9/10)	41.54 ± 4.10 (10/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	**88.02 ± 6.50** (33/34)	**68.62 ± 9.82** (204/306)	31.71 ± 3.60 (9/10)	**30.59 ± 5.27** (10/10)
		ViT-Tiny	1.0K	**86.71 ± 6.32** (30/34)	**68.50 ± 9.95** (241/306)	**43.86 ± 2.37** (10/10)	**44.76 ± 4.79** (10/10)

Table 29: **Cross-dataset 5-shot adaptation accuracy (%).** Source-pretrained methods pretrain on EPN612 and deploy on an archive held out from pretraining by construction. Win, shown in parentheses, denotes the number of wearers for whom a method outperforms the frozen Linear probe of the same encoder. Params denotes the per-wearer increment. ± denotes the sample standard deviation across wearers. **Bold** indicates the best result in each cell, and underline indicates the second-best result.

					Test set		
					UCI-EMG	DB5-B1	DB5-B2
Method	**Type**	**Backbone**	**Params**	**Train set**	**Acc↑ (Win)**	**Acc↑ (Win)**	**Acc↑ (Win)**
Linear probe	No adapt.	ViM-Tiny	0	EPN612	78.14 ± 7.87 (0/34)	22.19 ± 3.24 (0/10)	22.63 ± 2.75 (0/10)
		ViT-Tiny	0		72.77 ± 8.07 (0/34)	22.16 ± 2.79 (0/10)	22.29 ± 3.02 (0/10)
E2E scratch	Floor	ViM-Tiny	6.81M	–	25.51 ± 3.51 (0/34)	5.49 ± 1.21 (0/10)	5.22 ± 0.90 (0/10)
		ViT-Tiny	5.37M		16.83 ± 0.79 (0/34)	5.08 ± 0.90 (0/10)	4.84 ± 0.76 (0/10)
Full FT	Full adapt.	ViM-Tiny	6.81M	EPN612	55.34 ± 8.24 (0/34)	23.13 ± 3.40 (8/10)	24.51 ± 4.71 (9/10)
		ViT-Tiny	5.37M		51.17 ± 6.46 (0/34)	19.50 ± 3.48 (2/10)	20.56 ± 4.11 (2/10)
BitFit	PEFT	ViM-Tiny	41.9K	EPN612	79.47 ± 7.73 (25/34)	23.52 ± 3.35 (9/10)	23.52 ± 3.46 (8/10)
		ViT-Tiny	25.7K		74.46 ± 7.84 (24/34)	22.95 ± 3.24 (8/10)	24.09 ± 3.62 (10/10)
LoRA	PEFT	ViM-Tiny	13.8K	EPN612	78.11 ± 7.85 (13/34)	21.91 ± 2.97 (2/10)	22.34 ± 2.74 (3/10)
		ViT-Tiny	9.2K		73.22 ± 7.72 (21/34)	22.14 ± 3.03 (5/10)	22.41 ± 2.99 (6/10)
T3A	TTA	ViM-Tiny	0	EPN612	78.45 ± 8.78 (18/34)	22.72 ± 3.17 (6/10)	23.57 ± 2.95 (9/10)
		ViT-Tiny	0		72.86 ± 8.72 (15/34)	20.48 ± 2.81 (0/10)	20.59 ± 2.89 (1/10)
CORAL	Align	ViM-Tiny	9.2K	EPN612	58.62 ± 7.39 (0/34)	17.49 ± 2.15 (1/10)	17.34 ± 2.97 (0/10)
		ViT-Tiny	9.2K		46.43 ± 9.25 (0/34)	17.91 ± 2.34 (0/10)	18.85 ± 3.12 (0/10)
SAL	Align	ViM-Tiny	72	EPN612	77.42 ± 7.57 (14/34)	24.16 ± 3.16 (10/10)	25.50 ± 4.41 (9/10)
		ViT-Tiny	72		67.29 ± 7.53 (2/34)	26.83 ± 2.55 (10/10)	27.97 ± 3.70 (10/10)
EA	Align	ViM-Tiny	0	EPN612	81.78 ± 7.66 (32/34)	29.53 ± 2.30 (10/10)	30.61 ± 3.15 (10/10)
		ViT-Tiny	0		75.60 ± 7.67 (28/34)	29.64 ± 1.78 (10/10)	30.53 ± 3.76 (10/10)
AdaBN	Align	ViM-Tiny	0	EPN612	79.01 ± 7.52 (23/34)	23.26 ± 3.11 (10/10)	23.42 ± 3.22 (9/10)
		ViT-Tiny	0		73.34 ± 7.67 (22/34)	22.12 ± 3.10 (5/10)	22.90 ± 3.25 (8/10)
Proposed	Align+PEFT	ViM-Tiny	1.0K	EPN612	**85.31 ± 7.39** (34/34)	**31.27 ± 2.13** (10/10)	**32.37 ± 3.46** (10/10)
		ViT-Tiny	1.0K		**81.67 ± 7.95** (33/34)	**32.72 ± 2.11** (10/10)	**33.36 ± 3.88** (10/10)

Table 30: **Per-wearer 10-shot adaptation results on UCI-EMG with ViM-Tiny.** Acc denotes accuracy (%). Wearers are sorted by the gain of EA+ElectrodeMix over the frozen Linear probe. Each entry reports the mean ± standard deviation over five calibration draws for that wearer. Gain denotes the paired EA+ElectrodeMix minus Linear probe difference in percentage points.

Wearer	Acc↑			Gain
	Linear Probe	EA	EA+ElectrodeMix	
17	77.37 ± 3.61	83.45 ± 1.94	90.25 ± 2.90	**+12.88 ± 2.40**
1	71.11 ± 4.70	75.23 ± 4.08	83.96 ± 4.07	**+12.85 ± 4.25**
22	73.45 ± 1.66	80.35 ± 1.78	84.06 ± 0.74	**+10.61 ± 1.40**
24	86.39 ± 4.39	90.75 ± 2.23	96.71 ± 1.63	**+10.31 ± 2.95**
4	81.77 ± 2.45	84.48 ± 2.40	90.91 ± 1.50	**+9.14 ± 3.37**
27	72.85 ± 3.29	78.75 ± 3.20	81.67 ± 3.42	**+8.83 ± 2.86**
36	84.35 ± 1.98	86.45 ± 1.87	92.20 ± 2.39	**+7.86 ± 3.22**
28	78.61 ± 1.05	80.13 ± 1.31	86.27 ± 1.41	**+7.66 ± 2.23**
5	84.29 ± 1.68	81.45 ± 4.27	91.81 ± 3.45	**+7.52 ± 2.74**
7	87.74 ± 1.80	86.04 ± 1.70	94.81 ± 2.04	**+7.06 ± 1.53**
2	83.50 ± 0.83	90.82 ± 1.25	90.40 ± 3.28	**+6.90 ± 3.21**
13	87.88 ± 0.60	94.13 ± 1.94	94.77 ± 2.34	**+6.89 ± 1.88**
14	79.50 ± 1.36	77.13 ± 2.72	85.90 ± 1.24	**+6.39 ± 1.77**
15	89.78 ± 2.30	93.11 ± 1.53	96.06 ± 1.61	**+6.28 ± 1.27**
16	88.95 ± 1.41	92.46 ± 0.88	95.18 ± 1.72	**+6.23 ± 1.71**
31	86.48 ± 1.55	88.79 ± 2.24	92.34 ± 1.50	**+5.86 ± 2.15**
6	76.31 ± 5.64	78.06 ± 5.92	81.88 ± 2.91	**+5.56 ± 4.27**
26	71.54 ± 3.84	75.15 ± 3.29	76.92 ± 4.79	**+5.38 ± 3.49**
25	74.69 ± 3.91	75.13 ± 6.18	80.06 ± 4.12	**+5.38 ± 3.59**
21	84.00 ± 2.13	85.16 ± 1.18	89.33 ± 0.83	**+5.33 ± 2.74**
19	93.21 ± 1.86	96.66 ± 1.08	98.22 ± 0.60	**+5.01 ± 1.53**
10	92.63 ± 1.94	93.39 ± 0.98	96.88 ± 0.70	**+4.25 ± 1.75**
8	83.54 ± 1.54	82.18 ± 3.48	87.67 ± 2.65	**+4.13 ± 3.68**
34	95.73 ± 1.09	97.28 ± 0.72	99.76 ± 0.17	**+4.03 ± 0.93**
32	90.06 ± 2.14	91.51 ± 2.81	93.99 ± 1.65	**+3.92 ± 0.89**
29	87.05 ± 2.48	88.24 ± 2.02	90.67 ± 1.89	**+3.63 ± 2.89**
12	79.34 ± 1.17	81.12 ± 0.96	82.60 ± 0.43	**+3.26 ± 0.91**
35	79.58 ± 1.76	80.69 ± 1.54	82.23 ± 0.53	**+2.65 ± 1.67**
33	89.69 ± 1.31	92.41 ± 1.59	92.07 ± 1.87	**+2.38 ± 2.32**
20	97.32 ± 2.29	97.68 ± 1.68	99.11 ± 0.30	**+1.79 ± 2.29**
9	96.47 ± 1.08	95.65 ± 1.09	97.65 ± 0.91	**+1.18 ± 0.69**
18	92.15 ± 1.45	88.56 ± 1.51	93.28 ± 1.35	**+1.13 ± 0.82**
3	87.32 ± 1.83	87.37 ± 2.57	86.78 ± 1.71	**-0.53 ± 1.58**
23	89.71 ± 0.76	88.86 ± 1.57	88.02 ± 4.67	**-1.69 ± 4.49**