CREWS: CORTICAL REGION ALIGNMENT FOR ENROLLMENT WITHOUT PATIENT-SPECIFIC PARAMETERS IN SPEECH BCIS

Anonymous authors
Paper under double-blind review

ABSTRACT

Intracortical speech brain-computer interfaces (BCIs) offer a promising pathway to restore communication for individuals with severe speech disorders. However, conventional systems are predominantly patient-dependent and suffer from the scarcity of large-scale paired training data for each individual. Although transferring a decoder trained on previous patients could reduce this burden, cross-patient generalization remains challenging because electrode array placements cover distinct electrode indices and lack spatial alignment across subjects. Existing approaches typically mitigate this spatial mismatch by learning target-patient-specific input mappings, which fundamentally precludes zero-shot decoding. To overcome this limitation, we propose CREWS, a speech BCI framework that aligns neural recordings by common cortical regions rather than discrete electrode indices. CREWS reallocates each array index to its corresponding anatomical region slot and processes these inputs using shared, region-specific encoders. This design enables neural recordings from a novel patient to be decoded by parameters pre-trained on identical cortical regions, eliminating the need for patient-specific calibration parameters. In a transfer learning setup where a target patient shares only a single cortical region with the training cohort, CREWS pioneers viable zero-shot decoding, successfully demonstrating that cross-patient alignment without adaptation is attainable. Furthermore, with as few as 5 labeled adaptation sentences, CREWS achieves a phoneme error rate (PER) of 68.9%, outperforming baseline cross-patient methods that rely on patient-specific input layers. These results demonstrate that anatomical cortical-region alignment enables cross-patient transfer without patient-specific parameters and can substantially reduce the calibration data required to deploy intracortical speech BCIs.

1 INTRODUCTION

An intracortical speech brain-computer interface (BCI) captures neural activity through microelectrode arrays implanted in cortical regions associated with speech and decodes the words that an individual with paralysis aims to articulate. These BCIs have advanced rapidly, progressing from decoding unconstrained sentences from a large vocabulary (Willett et al., 2023) to enabling accurate communication in naturalistic, everyday settings (Card et al., 2024).

A major barrier to the clinical deployment of these systems, however, is the extensive data required from each new user. Existing high-performance speech BCIs are typically trained from scratch for each individual, relying on dataset scales that reach thousands of attempted sentences. Although transferring knowledge from existing cohort data to a novel patient can alleviate calibration burdens, current paradigms predominantly evaluate decoders within single-subject regimes. The established benchmarks, such as the Brain-to-Text evaluation suite, evaluate models strictly on an intra-patient basis (Willett et al., 2024). Cross-patient transfer remains exceptionally challenging because the same electrode index records different neurons in each patient (Azabou et al., 2023).

Recent studies of cross-patient speech decoding attempt to resolve these spatial disparities by learning patient-specific input transformations. Decoders based on stereo-EEG and micro-electrocorticography utilize patient-specific convolutional or alignment layers to harmonize neural

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.

signals (Singh et al., 2025; Spalding et al., 2026). Similarly, in intracortical speech BCIs, Levin et al. (2026) transferred source-patient models to novel target subjects by appending a patient-specific input mapping. While this adaptation strategy outperformed from-scratch training when calibration data was extremely limited (< 200 sentences), it still relies on optimizing a dedicated target layer using scarce data. Moreover, because these frameworks require explicitly learning new input parameters for each subject, they inherently lack zero-shot decoding capabilities.

Simply eliminating patient-specific parameters is also insufficient: Levin et al. (2026) report that a naive shared input layer yields worse performance than training from scratch, suggesting patient-specific adaptation is indispensable. However, a standard shared input layer applies the same weights to the i-th electrode channel across all subjects, despite these channels recording functionally disparate neural ensembles. We therefore attribute the failure of shared input architecture to the lack of correspondence between the inputs of different patients rather than to sharing itself.

Although electrode indices lack spatial correspondence across patients, their underlying anatomical cortical regions remain functionally conserved. In this paper, we introduce CREWS, an input framework that aligns cross-patient neural recordings by anatomical region rather than channel index. CREWS assigns each array to a designated patient-shared regional slot, then averages features exclusively across the slots occupied by a given subject. The cross-patient transfer primarily relies on the ventral area 6v (v6v) slot, from which attempted speech is decoded with the greatest accuracy (Willett et al., 2023; Kunz et al., 2025). Because all three subjects in our cohort share a v6v implant, an unseen target patient's v6v array can be directly processed by shared encoders pre-trained on source cohort data. By eliminating the need for patient-specific input layers, CREWS achieves zero-shot decoding on unseen target subjects and enables rapid adaptation.

To the best of our knowledge, CREWS is the first intracortical speech BCI capable of zero-shot speech decoding on a novel subject, achieving a phoneme error rate (PER) below chance level. On a novel target subject (T16), CREWS achieves a zero-shot PER of 77.7%, compared with the chance level of 91.4%. With as few as 5 labeled calibration sentences from T16, CREWS attains a PER of 68.9%, outperforming all baseline models by at least 12.6 points. Crucially, replacing the v6v array with any other array of T16, or passing the v6v array to any other slot, raises the zero-shot PER above the chance level, indicating that anatomical alignment drives successful cross-patient transfer. These findings demonstrate that anatomical cortical-region alignment can substantially reduce calibration data requirements for deploying intracortical speech BCIs to novel users.

Our contributions are as follows:

- We introduce CREWS, an input framework that aligns neural recordings across subjects by anatomical cortical region rather than electrode channel indices, enabling end-to-end parameter sharing across a single unified speech BCI.
- We demonstrate that CREWS achieves zero-shot speech decoding on a novel target subject, yielding a PER below chance level, and outperforms baseline models when adapted on just a few labeled calibration sentences.
- Through ablations, we show that zero-shot transfer depends on cortical-region alignment, a projection specific to each slot, averaging the present slots, and dropping whole regions during training, whereas a label-free electrode order mainly benefits adaptation.

2 Related Work

Intracortical speech BCIs. Conventional intracortical speech BCIs typically train patient-specific decoders exclusively on individual user recordings, frequently incorporating dedicated day-, week-, or dataset-specific input layers to accommodate non-stationary signal shifts across sessions (Willett et al., 2023; Card et al., 2024; Wairagkar et al., 2025; Kunz et al., 2025). In contrast, CREWS dispenses with day-specific input layers entirely, a design choice supported by our empirical finding that removing these session-specific adaptation layers from published decoders results in negligible changes to phoneme error rates (Appendix A.1).

Cross-patient transfer. In motor BCIs, NDT2 and POYO accommodate a new patient through context tokens and new unit embeddings, respectively (Ye et al., 2023; Azabou et al., 2023). For

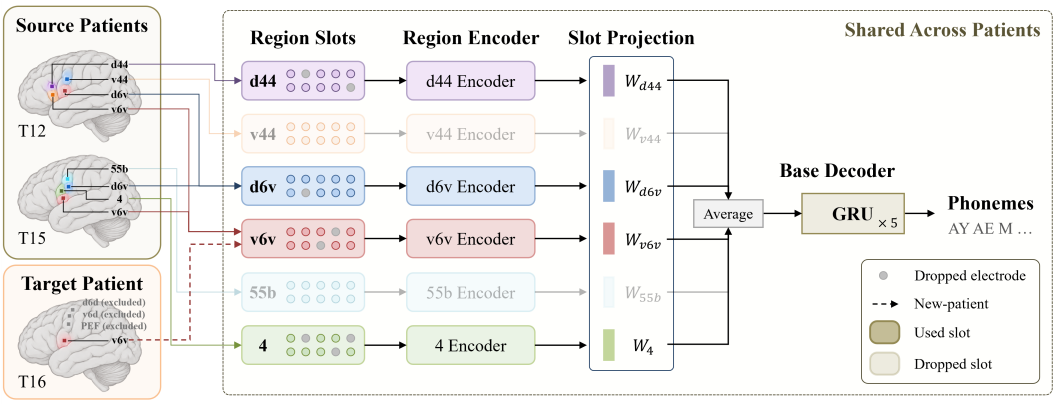


Figure 1: Overview of CREWS. Each array is assigned to the slot of its cortical region among the regions of the source patients, and only the present slots are averaged before the base decoder. During training, coverage dropout removes whole slots and single electrodes. A target subject is decoded from its arrays in these regions only (v6v for T16), zero-shot and adaptation, without new parameters. For simplicity, the source patients are drawn together, although each training sample comes from one patient.

speech decoding, Singh et al. (2025) learn a convolution for each patient from stereo-EEG, whereas Spalding et al. (2026) align the micro-electrocorticography of different patients using canonical correlation analysis. Without this alignment, a model pooled across patients decodes phonemes less accurately than a model trained on each patient alone (Spalding et al., 2026). Using the same BrainGate patients, Levin et al. (2026) transfer an intracortical speech BCI through a separate input layer for each recording session, outperforming training from scratch below 200 sentences of the new patient. Boccato et al. (2026) train jointly on T12 and T15 with dataset-specific affine layers and adapt to new patients by learning a new layer from isolated words. All of these methods learn parameters specific to a new patient.

Anatomical common spaces. Anatomical common spaces have been used in a different setting, the pretraining of foundation models for intracranial recordings. The Population Transformer and BaRISTA embed electrodes by their coordinates or regions and fine-tune the pretrained models on new patients (Chau et al., 2025; Oganesian et al., 2025). Our contribution is not the anatomical common space itself. Rather, CREWS applies this space to phoneme sequence decoding from Utah arrays and aligns patients by cortical region rather than by electrode index. Region alignment removes the patient-specific input layer and allows a new patient to be decoded zero-shot. Because Utah arrays provide no electrode coordinates that correspond across patients, CREWS orders the electrodes within each region using a label-free signal statistic instead of coordinate embeddings.

3 Method

Electrode indices do not correspond across patients, whereas cortical regions do. CREWS therefore aligns patients by cortical region rather than by electrode index, allowing all patients to share every parameter of one speech BCI (Figure 1). Each Utah array is assigned to the slot of its cortical region, where an encoder shared by all patients reads the electrodes in an order derived from a label-free signal statistic. The outputs of the covered slots pass through slot projections and are averaged, yielding one input sequence for any number of covered regions. The base decoder, the published GRU decoder for intracortical speech without day-specific input layers, maps this sequence to phonemes. During training, coverage dropout randomly removes regions other than v6v, enabling the decoder to read the subset of regions that a target subject shares. Because no parameter is patient-specific, CREWS decodes a target subject zero-shot and adapts to the patient by training on its sentences together with data from the source patients.

3.1 Problem Setting

Let $\mathcal{P}$ denote the set of patients in the training cohort, $\mathcal{A}_p$ the set of arrays implanted in patient p, and $r(a)$ the cortical region in which array a lies. A trial of patient p consists of neural features $\boldsymbol{X}^{(p)}$ and a phoneme sequence $\boldsymbol{y}$, where each electrode contributes the threshold-crossing count and the spike-band power in 20 ms bins (Willett et al., 2023). Existing cross-patient decoders decompose the parameters into a shared network and an input layer θ_p for each patient (Levin et al., 2026; Boccato et al., 2026). The number of parameters of these decoders grows with the number of patients. CREWS instead learns a single parameter set θ by minimizing the CTC loss over all patients,

$$\mathcal{L}(\theta) = \sum_{p \in \mathcal{P}} \mathbb{E}_{(\boldsymbol{X}^{(p)}, \boldsymbol{y}) \sim \mathcal{D}_p} \big[\ell_{\mathrm{CTC}}\big(f_\theta(\boldsymbol{X}^{(p)}), \boldsymbol{y}\big) \big], \tag{1}$$

where $\mathcal{D}_p$ denotes the labeled trials of patient p and f_θ the decoder. For a target subject $q \notin \mathcal{P}$, zero-shot decoding applies f_θ to $\boldsymbol{X}^{(q)}$ without any labeled trial of q, and k-shot adaptation minimizes Eq. 1 over $\mathcal{P} \cup \{q\}$ with k labeled trials in $\mathcal{D}_q$. Neither setting introduces a parameter specific to q.

3.2 Region Slots

The input stage has nine region slots, v6v, d6v, 4, 55b, v44, d44, PEF, v6d, and d6d, named after the multi-modal parcellation of Glasser et al. (2016). Each slot contains the 64 electrodes of one Utah array, and an array is assigned only to the slot of its region $r(a)$. The region of each array is the area of this parcellation in which the array was implanted, as reported by the original studies (Willett et al., 2023; Card et al., 2024; Kunz et al., 2025). Two patients with arrays in the same region therefore feed the same slot, although their electrodes record different neurons. A slot belongs to a region rather than to a patient. A slot covered by a single source patient, such as 55b in T15, becomes shared as soon as another patient with an array in that region is added to the cohort. A slot that no source patient covers, such as PEF for T16, has an encoder that has not been trained. The array of a target subject in such a slot is excluded until a later patient covers the region. T16 is therefore decoded from its v6v array alone.

3.3 Region Encoders

Let g_r denote the encoder of slot r, shared by all patients: a linear map of the 2×64 features of one time bin to 256 dimensions followed by layer normalization and a GELU,

$$\boldsymbol{h}_r(t) = g_r\big(\boldsymbol{X}^{(p)}_{\pi_{p,r}}(t)\big), \tag{2}$$

where $\pi_{p,r}$ denotes the order of the electrodes of patient p in slot r. Because the encoder assigns one weight to each input position, the electrodes of a slot require an order that is comparable across patients. A permutation-invariant encoder would avoid the order but would discard the identity of each electrode, which the linear map retains. For every electrode, the autocorrelation of its threshold crossings and spike-band power at lags of 2-5 bins (40-100 ms) is computed from unlabeled recordings and converted to ranks within the patient and region. The order $\pi_{p,r}$ is obtained by sorting the electrodes by the mean of these ranks. The mean rank predicts an electrode's importance from one source patient to another, with a Spearman correlation of 0.57 (Appendix A.2). Consequently, the order is defined without electrode-level anatomical information and without labeled data from a new patient. Because the encoders are shared, the v6v array of a target subject is read by the encoder trained on the v6v arrays of the source patients.

3.4 Averaging the Present Slots

Each slot applies a slot projection $\boldsymbol{W}_r$ to the encoder output, and only the slots that the current sample covers are averaged:

$$\boldsymbol{u}_r(t) = \mathrm{GELU}\big(\boldsymbol{W}_r \boldsymbol{h}_r(t)\big), \qquad \boldsymbol{z}(t) = \frac{\sum_r m_r \boldsymbol{u}_r(t)}{\sum_r m_r}, \tag{3}$$

where $m_r \in \{0, 1\}$ denotes whether slot r is present in the sample. At test time, $m_r = 1$ for every region that the patient covers and at least one source patient covers; during training, coverage

dropout samples m_r (Section 3.5). The regions present in the input differ from patient to patient. Averaging the projected slots gives the decoder an input of the same form for any combination of regions, without a position that records which slots are present. The decoder therefore cannot identify the patient from the set of present slots. Unlike a sum, the average does not grow with the number of present slots, which ranges from one for T16 to four for the source patients.

3.5 Coverage Dropout

A target subject covers only part of the cohort regions: T16 shares only v6v with the source patients (Appendix A.3). Among the covered regions, v6v contributes the most to the decoding of attempted speech (Appendix A.4). Willett et al. (2023) decoded speech most accurately from the ventral 6v array of T12, and Kunz et al. (2025) found significant speech decoding in T16 only from the v6v array. Consistent with this role, all three patients in our cohort were implanted with an array in v6v. Coverage dropout therefore samples the mask of Eq. equation 3 for every training sample while retaining v6v: with probability 0.15, only v6v is kept; otherwise, each other region is dropped independently with probability 0.15. In expectation, these probabilities remove 21% of the regions of a patient with four arrays. Each electrode is additionally set to zero with probability 0.10, and the remaining electrodes are rescaled as in standard dropout. Because Eq. equation 3 averages only the present slots, the dropped regions leave no trace in $\boldsymbol{z}(t)$. The decoder is thereby trained on subsets of regions, including v6v alone, the single region that T16 shares with the source patients.

3.6 Base Decoder and Adaptation

For a fair comparison, CREWS and all baselines share the published GRU temporal body for intracortical speech (Willett et al., 2023; Card et al., 2024) without its day-specific input layers, referred to as the base decoder. The base decoder accepts 14 consecutive fused feature vectors with a stride of 4 bins per input token. A five-layer unidirectional GRU (Cho et al., 2014) with 768 units maps these tokens to posteriors over 39 phonemes, a word-boundary token, and a connectionist temporal classification (CTC) blank token (Graves et al., 2006). To adapt CREWS to a target subject, the user's k sentences are appended to source cohort data, and training continues without introducing new parameters or freezing any layers. Batches are sampled round-robin across patients, weighting each patient, including target subjects with k sentences, equally in Eq. 1. As a result, the adapted model preserves performance on source subjects.

The input stage holds 0.90M of CREWS's 25.13 M parameters: 99k for the encoder and the projection of each slot. Unlike a patient-specific input layer, which adds 131k parameters for every patient with four arrays, the input stage grows with the number of regions rather than that of patients.

4 Experimental Setup

Patients and data. Neural recordings from three BrainGate2 clinical trial patients were evaluated (Appendix A.3). Implanted array locations were: T15 (v6v, 4, 55b, d6v), T12 (v6v, d6v, v44, d44), and T16 (PEF, v6v, v6d, d6d), leaving v6v as the sole anatomical region shared across all three subjects. T15 and T12 served as source cohort subjects with 9,022 and 10,800 open-vocabulary training sentences, and T16 served as the novel target subject. T16 attempted 120 sentences drawn from a 50-word vocabulary; 60 trials were held out for evaluation, and the remaining 60 were used for k-sentence adaptation. To verify that findings generalize across subjects, T15 or T12 was alternately held out as the target subject ($k \leq 60$), with the remaining two patients forming the source cohort. T16 contributed 108 of its sentences to each of these source cohorts through its v6v array alone. These cohorts therefore held about half the training sentences available when T16 served as the target subject (10,908 or 9,130 against 19,822). Threshold crossings and spike-band power were block-wise z-scored per electrode and then smoothed over time using a Gaussian kernel with a standard deviation of 40 ms.

Zero-shot and few-shot evaluation. Zero-shot decoding denotes decoding a target subject without training on labeled sentences of that patient ($k = 0$). The unlabeled recordings of the target subject are used only for the normalization and the electrode order of Section 3.3. Few-shot adaptation denotes training on k labeled sentences of the target subject.

Table 1: Zero-shot decoding and few-shot adaptation on the target subject T16. PER (%); k denotes the number of adaptation sentences; the chance level of zero-shot CREWS is 91.4; lowest PER in bold; – denotes a decoder without output. All baselines are our implementations on the base decoder. CREWS attains the lowest PER at every k (paired bootstrap over trials, $p < 0.002$).

Method	New params	Zero-shot	k=5	k=10	k=20	k=60
No source data						
Base decoder from scratch	all	–	91.7	92.9	91.8	92.9
Patient-specific input layer						
Frozen decoder	131k	–	88.7	87.1	84.8	83.6
Levin et al. (2026) style	131k	–	83.6	68.1	55.4	24.2
Shared input stage						
Shared input layer	0	91.6	81.5	64.2	48.6	26.1
CREWS	**0**	**77.7**	**68.9**	**61.5**	**45.2**	**23.5**

Metrics and chance level. The phoneme error rate (PER) was computed as the edit distance between decoded and reference phoneme sequences divided by reference length, averaged across sentences. Chance level was estimated by trial shuffling (Ojala & Garriga, 2010; Combrisson & Jerbi, 2015): decoded sequences were held fixed, while reference sentences were randomly permuted across trials for the same subject (averaged over 30 permutations). This shuffling preserves underlying recording conditions and subject-specific sentence distributions. Because chance levels depend on decoded sequence length and vary by model, table captions report chance level of zero-shot CREWS as the reference. All metrics reflect averages across three independent training seeds.

Baselines. All baselines utilized the base decoder architecture (Section 3.6) and were trained on identical source data within our pipeline, mirroring the cited methodologies rather than reusing published metrics. Consequently, the models differ solely in the input stage and the adaptation schedule. The base decoder trained from scratch on the k target sentences follows the single-patient recipe of Willett et al. (2023) and Card et al. (2024). The patient-specific input layer adds a linear layer with a softsign nonlinearity (131k parameters per patient) (Card et al., 2024). For the new patient, this layer was trained with the base decoder frozen, as in the linear adaptation of Boccato et al. (2026). The Levin et al. (2026) style trained this layer together with all parameters on the new sentences and the source data, following their replay of source data during fine-tuning. Batches were drawn from the patients in turn, as for CREWS, rather than 70% from the new patient. Finally, the shared input baseline replaces patient-specific layers with a single 256-channel linear layer trained jointly, replicating the transfer failure observed by Levin et al. (2026).

Input of the target subject. The PEF, v6d, and d6d slots are covered by no source patient, and training them on the sentences of T16 alone would create parameters specific to T16. CREWS therefore received only the v6v array of T16, whereas the baselines received all four arrays. The comparison thus gives the baselines more neural input than CREWS.

Training details. Source models and the base decoder trained from scratch were trained for 300 epochs using AdamW (Loshchilov & Hutter, 2019) with a learning rate of 0.02, a cosine schedule with 1,000 warm-up steps, and a weight decay of 0.001. Each batch contained 64 trials, and gradient norms were clipped at 10. Dropout of 0.2 was applied to the input tokens of the GRU and of 0.4 between its layers. Following Willett et al. (2023), white noise with a standard deviation of 1.0 and a constant offset with a standard deviation of 0.2 were added to the features during training. All transfer methods were adapted for 100 epochs with a learning rate of 0.005.

Table 2: Adaptation schedules at $k = 10$ and forgetting of the source patients. PER (%); the chance level of zero-shot CREWS on T16 is 91.4. Source columns give the adapted model; the frozen decoder leaves the patient-specific source models unchanged, and no adaptation gives CREWS before adaptation, zero-shot on T16. Bold marks CREWS trained with the source data.

		Target subject	Source subjects	
Adaptation	New params	T16	T15	T12
Patient-specific input layer				
Frozen decoder	131k	87.1	16.8	22.1
Levin et al. (2026) style	131k	68.1	17.1	23.4
CREWS				
No adaptation	0	77.7	17.0	21.7
Trained with the source data	**0**	**61.5**	19.1	24.6

5 Results

5.1 Transfer to a New Patient

Table 1 evaluates transfer to T16, whose arrays share only the v6v region with the source cohort. Although CREWS receives only the v6v array of T16 while baselines access all four arrays, CREWS is the sole model enabling zero-shot transfer. CREWS decodes T16 at a PER of 77.7, significantly below chance level ($p < 0.002$), whereas models with patient-specific input layers fail to produce output due to untrained parameters. Replacing T16's neural signals with Gaussian noise or zeros raises PER above 100 (Figure 2), confirming that zero-shot decoding relies directly on T16's neural activity. The shared input baseline remains chance level, replicating the transfer failure noted by Levin et al. (2026) and demonstrating that sharing inputs without anatomical correspondence yields no transfer to unseen subjects.

With few-shot adaptation, CREWS achieves the lowest PER among all baselines across every sentence count k (Table 1). With 5 sentences, CREWS attains lower PER than every baseline, including the base decoder trained from scratch. With 10 sentences, it reduces PER to 61.5, compared with 68.1 for the Levin et al. (2026) style. Jointly training the shared input layer on the new sentences improves over the patient-specific input layers up to 20 sentences, but it remains above CREWS at every k and provides no zero-shot transfer. These results demonstrate that anatomical region alignment enables effective parameter-free transfer across both zero-shot and few-shot adaptation, yielding the largest margin over baselines at 5 sentences.

5.2 Source Patients after Adaptation

The source columns of Table 2 examine whether removing patient-specific parameters reduces accuracy on the patients seen during training. Before adaptation, CREWS attains a PER of 17.0 on T15 and 21.7 on T12, matching the patient-specific input layer (16.8 and 22.1). These results indicate that a decoder without patient-specific parameters serves the source patients without loss of accuracy.

The table also compares the adaptation schedules at $k = 10$ on T16 and on the source patients. With patient-specific input layers, freezing the base decoder keeps the source models unchanged but attains only 87.1 on T16, and the Levin et al. (2026) style attains 68.1 while adding one patient-specific input layer. Adapting CREWS with the source data attains 61.5 on T16 without new parameters, while PER on T15 and T12 rises by 2.1 and 2.9 points. These results indicate that one model without new parameters achieves the lowest PER on the new patient while keeping the source patients within 3 points of their pre-adaptation accuracy.

5.3 Other Held-Out Patients

To test whether these results generalize beyond T16, Table 3 holds out T15 or T12 as the target subject instead, using an open vocabulary and up to 60 adaptation sentences. CREWS again achieves

Table 3: Rotation of the target subject. T15 or T12 serves as the target subject, and the other two patients form the source cohort. PER (%); the chance level of zero-shot CREWS is 94.2 for T15 and 95.0 for T12; lowest PER in bold; – denotes a decoder without output.

			Few-shot, k sentences			
Target subject	Method	Zero-shot	k=5	k=10	k=20	k=60
T15	Base decoder from scratch	–	95.8	96.3	96.5	94.0
	Levin et al. (2026) style	–	87.1	85.0	78.2	66.4
	Shared input layer	83.5	78.6	75.9	71.2	57.3
	CREWS	**77.5**	**76.9**	**73.8**	**66.3**	**52.7**
T12	Base decoder from scratch	–	95.3	96.7	93.0	92.3
	Levin et al. (2026) style	–	81.1	79.3	76.7	73.4
	Shared input layer	84.8	79.3	77.6	76.9	74.3
	CREWS	**78.9**	**77.1**	**76.2**	**70.2**	**68.5**

Table 4: Ablations of CREWS on the target subject T16. Each row below the first changes one component and trains the model under the protocol of the first row. PER (%); the chance level of zero-shot CREWS on T16 is 91.4.

Variant	Zero-shot	k=5	k=10	k=20	k=60
CREWS	77.7	68.9	61.5	45.2	23.5
Region alignment					
Shared input layer, no region slots	91.6	81.5	64.2	48.6	26.1
v6v slot only	79.0	77.3	69.1	57.4	28.4
Input stage					
No electrode autocorrelation ordering	80.7	77.4	72.3	67.0	50.1
No slot projection	84.8	74.2	68.9	46.9	26.4
Slot concatenation instead of average	86.7	79.4	70.6	49.9	32.5
Coverage dropout					
None	133.7	76.7	63.3	59.5	26.7
Region only dropped	80.6	70.5	64.2	58.4	24.9
Electrode only dropped	122.3	79.0	62.1	56.3	25.2

below-chance zero-shot decoding, attaining 77.5 for T15 and 78.9 for T12, outperforming the shared input baseline. Although these source cohorts contain roughly half the training sentences of the main cohort, CREWS achieves the lowest PER for both target subjects across all k, outperforming the Levin et al. (2026) approach. These results indicate that the advantage of CREWS holds for all three target subjects, with both closed and open vocabularies.

5.4 Ablations

In Table 4, each variant changes one component of CREWS. Replacing the region slots with a single input layer over electrode indices raises the zero-shot PER on T16 from 77.7 to 91.6, the chance level. Restricting every patient to the v6v slot keeps the zero-shot PER at 79.0. These results indicate that zero-shot transfer arises from aligning the arrays by cortical region.

Without the autocorrelation ordering, each slot reads the electrodes in its own recording order, so the same input position of a region encoder receives unrelated electrodes in different patients. This variant raises zero-shot PER and further hinders adaptation. These results indicate that an electrode order comparable across patients supports zero-shot transfer and is required for adaptation.

Within the input stage, removing the slot projections or concatenating the slots instead of averaging them raises the zero-shot PER to 84.8 and 86.7, respectively. One explanation is that, without a slot-specific layer after the layer normalization of each region encoder, the average cannot weight v6v, the most informative region (Appendix A.4), above the other slots. Concatenating the region encoder outputs assigns each slot a fixed position, so the set of filled positions identifies the source patient from its regions. Its PER exceeds that of the average at every k, consistent with a decoder

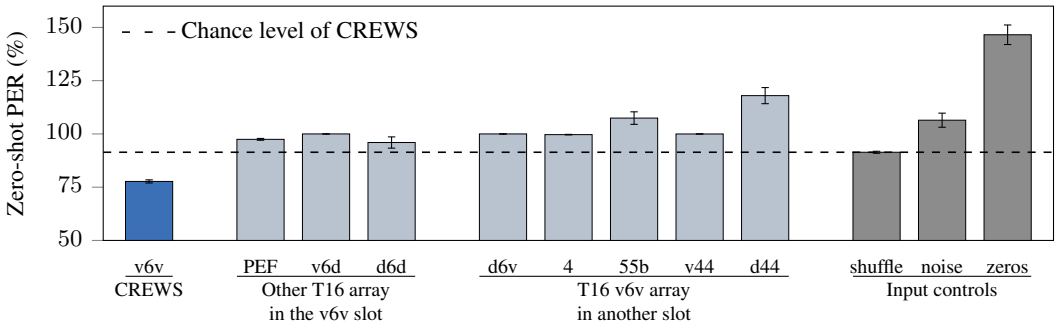


Figure 2: Zero-shot PER of CREWS on the new patient T16 when the input of a slot is changed at test time. The v6v slot receives the v6v array of T16, as in CREWS, or another array of T16; the slot of another region reads the v6v array of T16; or the v6v input is replaced by a control (electrodes shuffled within the array, Gaussian noise, or zeros). The dashed line marks the chance level of CREWS. Error bars denote the standard error over trials.

that relies on the region set to identify the patient. These results indicate that the slot projections and the average over the present slots are both required for zero-shot transfer.

Removing coverage dropout causes the largest zero-shot degradation in Table 4, raising the PER on T16 from 77.7 to 133.7. Dropping regions alone restores a zero-shot PER of 80.6, whereas dropping electrodes alone leaves it at 122.3. Region dropout is therefore the component that enables zero-shot decoding of a new patient who shares only v6v with the source patients.

5.5 Analysis of the Region Slots

Figure 2 evaluates whether slot transfer depends on matching input source to trained cortical region. Passing the v6v array of T16 to another slot, or another array of T16 to the v6v slot, raises the zero-shot PER to between 96 and 118. Because speech information in T16 is localized to v6v (Kunz et al., 2025), the failure stems from slot mismatch rather than array quality. Furthermore, intra-array electrode shuffling degrades PER to chance level, while replacing inputs with noise or zeros pushes PER above 100. Only the v6v array-slot pairing yields below chance PER. For source patients, removing individual regions at test time reproduces known regional performance rankings (Appendix A.4). Thus, cross-patient slot transfer requires strict anatomical concordance between the input array and its designated slot.

6 Conclusion

We introduced CREWS, an intracortical speech BCI that aligns patients by cortical region rather than by electrode index and contains no patient-specific parameters. Although prior work found that transfer fails without patient-specific input layers, establishing a correspondence between patients through cortical regions makes transfer possible without them. One model serves source patients as accurately as patient-specific input layers and decodes a new patient who shares only v6v with the source patients at zero-shot below the chance level. With few-shot adaptation, CREWS attains the lowest PER among the compared methods at every number of sentences, for T16 as well as for T15 and T12 as new patients. A slot transfers to a new patient only when it reads the array of its own cortical region. While the autocorrelation order of electrodes matters mainly for adaptation, ablations attribute zero-shot transfer to the slot projections, the average over the present slots, and region dropout during training. These results suggest that aligning patients by cortical region can reduce the labeled data that each new patient must provide to an intracortical speech BCI.

Limitations. Because zero-shot decoding remains far from practical use, a new patient still needs to provide a few labeled sentences. With three patients who share only v6v, the cohort cannot test transfer through the other regions or separate the effect of adding source patients from the amount of source data. Adaptation with source data also requires access to the source patients' recordings. Finally, the assignment of arrays to regions relies on the array locations reported at implantation.

AI USE STATEMENT

The authors used Claude (Anthropic) as a coding assistant to implement the methods and data-processing code, format tables, search for and format references, and edit the manuscript for readability. No data were generated with generative AI tools. We reviewed, edited, and verified all code, tables, and text, and every reported number against the experimental outputs. The authors take full responsibility for the content of this paper.

ETHICS STATEMENT

This work is a secondary analysis of intracortical recordings released with prior publications (Willett et al., 2023; Card et al., 2024; Kunz et al., 2025). The recordings were collected in the BrainGate2 pilot clinical trial under an Investigational Device Exemption and with approval of the institutional review boards of the participating sites, and every patient provided informed consent, as described in the original publications. The released data contain no personally identifying information; the data were used in accordance with the terms of their release, and no attempt was made to re-identify any patient.

REPRODUCIBILITY STATEMENT

Section 3 specifies the model, and Section 4 specifies the data, metrics, and compared methods. Section 4 lists the training settings, Appendix A.2 describes the label-free electrode order, and Appendix A.6 lists the language-model decoding settings.

REFERENCES

Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael J. Mendelson, Blake Richards, Matthew G. Perich, Guillaume Lajoie, and Eva L. Dyer. A unified, scalable framework for neural population decoding. In *Advances in Neural Information Processing Systems*, volume 36, pp. 44937–44956, 2023. doi: 10.52202/075280-1948.

Tommaso Boccato, Michal Olak, and Matteo Ferrante. Cross-subject decoding of human neural data for speech brain computer interfaces. *Journal of Neural Engineering*, 23(4):046018, 2026. doi: 10.1088/1741-2552/ae8576.

Nicholas S. Card, Maitreyee Wairagkar, Carrina Iacobacci, Xianda Hou, Tyler Singer-Clark, Francis R. Willett, Erin M. Kunz, Chaofei Fan, Maryam Vahdati Nia, Darrel R. Deo, et al. An accurate and rapidly calibrating speech neuroprosthesis. *New England Journal of Medicine*, 391(7):609–618, 2024. doi: 10.1056/NEJMoa2314132.

Geeling Chau, Christopher Wang, Sabera Talukder, Vighnesh Subramaniam, Saraswati Soedarmadji, Yisong Yue, Boris Katz, and Andrei Barbu. Population Transformer: Learning population-level representations of neural activity. In *The Thirteenth International Conference on Learning Representations*, pp. 69100–69121, 2025.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014. doi: 10.3115/v1/D14-1179.

Etienne Combrisson and Karim Jerbi. Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250:126–136, 2015. doi: 10.1016/j.jneumeth.2015.01.010.

Matthew F. Glasser, Timothy S. Coalson, Emma C. Robinson, Carl D. Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F. Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016. doi: 10.1038/nature18933.

Alex Graves, Santiago Fernández, Faustino Gomez, and Jürgen Schmidhuber. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 369–376, 2006. doi: 10.1145/1143844.1143891.

Erin M. Kunz, Benyamin Abramovich Krasa, Foram Kamdar, Donald T. Avansino, Nick Hahn, Seonghyun Yoon, Akansha Singh, Samuel R. Nason-Tomaszewski, Nicholas S. Card, Justin J. Jude, et al. Inner speech in motor cortex and implications for speech neuroprostheses. *Cell*, 188 (17):4658–4673.e17, 2025. doi: 10.1016/j.cell.2025.06.015.

Alisa D. Levin, Donald T. Avansino, Foram B. Kamdar, Nicholas S. Card, Maitreyee Wairagkar, Brandon G. Jacques, Justin J. Jude, Carrina Iacobacci, Bayardo E. Lacayo, Payton H. Bechefsky, et al. Cross-brain transfer of high-performance intracortical speech and handwriting BCIs. *bioRxiv*, 2026. doi: 10.64898/2026.01.12.699110.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *The Seventh International Conference on Learning Representations*, 2019.

Lucine L. Oganesian, Saba Hashemi, and Maryam M. Shanechi. BaRISTA: Brain scale informed spatiotemporal representation of human intracranial neural activity. In *Advances in Neural Information Processing Systems*, volume 38, 2025. doi: 10.52202/085713-5573.

Markus Ojala and Gemma C. Garriga. Permutation tests for studying classifier performance. *Journal of Machine Learning Research*, 11(62):1833–1863, 2010.

Daniel Povey, Arnab Ghoshal, Gilles Boulianne, Lukáš Burget, Ondřej Glembek, Nagendra Goel, Mirko Hannemann, Petr Motlíček, Yanmin Qian, Petr Schwarz, et al. The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, 2011.

Aditya Singh, Tessy Thomas, Jinlong Li, Greg Hickok, Xaq Pitkow, and Nitin Tandon. Transfer learning via distributed brain recordings enables reliable speech decoding. *Nature Communications*, 16(1):8749, 2025. doi: 10.1038/s41467-025-63825-0.

Z. Spalding, S. Duraivel, S. Rahimpour, C. Wang, K. Barth, C. Schmitz, S. P. Lad, A. H. Friedman, D. G. Southwell, J. Viventi, et al. Shared latent representations of speech production for cross-patient speech decoding. *Nature Communications*, 17(1):8716, 2026. doi: 10.1038/s41467-026-75455-1.

Maitreyee Wairagkar, Nicholas S. Card, Tyler Singer-Clark, Xianda Hou, Carrina Iacobacci, Lee M. Miller, Leigh R. Hochberg, David M. Brandman, and Sergey D. Stavisky. An instantaneous voice-synthesis neuroprosthesis. *Nature*, 644(8075):145–152, 2025. doi: 10.1038/s41586-025-09127-3.

Francis R. Willett, Erin M. Kunz, Chaofei Fan, Donald T. Avansino, Guy H. Wilson, Eun Young Choi, Foram Kamdar, Matthew F. Glasser, Leigh R. Hochberg, Shaul Druckmann, et al. A high-performance speech neuroprosthesis. *Nature*, 620(7976):1031–1036, 2023. doi: 10.1038/s41586-023-06377-x.

Francis R. Willett, Jingyuan Li, Trung Le, Chaofei Fan, Mingfei Chen, Eli Shlizerman, Yue Chen, Xin Zheng, Tatsuo S. Okubo, Tyler Benster, et al. Brain-to-Text Benchmark '24: Lessons learned. *arXiv preprint arXiv:2412.17227*, 2024. doi: 10.48550/arXiv.2412.17227.

Joel Ye, Jennifer L. Collinger, Leila Wehbe, and Robert Gaunt. Neural Data Transformer 2: Multi-context pretraining for neural spiking activity. In *Advances in Neural Information Processing Systems*, volume 36, pp. 80352–80374, 2023. doi: 10.52202/075280-3523.

A Appendix

A.1 Day-Specific Input Layers

Table A.1 evaluates whether session-specific input layers utilized in prior decoding architectures are essential within a single-subject framework. The reference decoders of Card et al. (2024) and Willett et al. (2023) were retrained using a unified input layer shared across all recording days, keeping all other architectural parameters identical. On subject T15, this shared input representation achieves virtually identical performance to the day-specific baseline (PER of 9.55 vs. 9.59, respectively). On subject T12, removing session-specific parameters results in a modest PER increase of 1.46 points. These empirical findings align with the effect of block-wise z-score normalization applied to raw neural signals, which effectively mitigates inter-session baseline drifts in electrode mean and variance; consequently, CREWS dispenses with day-specific parameterization entirely.

Table A.1: Day-specific input layers in the published GRU decoders. PER (%) with the official code and data splits, averaged over three training seeds; the shared variant applies one input layer to every recording day.

Input layer	T15	T12
Day-specific input layers	9.59	20.35
One shared input layer	9.55	21.81

A.2 Electrode Order within a Region

For every electrode channel, label-free summary statistics of the threshold-crossing and spike-band-power features were computed directly from unannotated neural recordings, with each statistic subsequently rank-normalized within individual subjects and anatomical regions. When fitted on one patient and evaluated on another, a ridge regression model trained on these statistics predicts electrode feature importance within a trained patient-specific decoder with a Spearman rank correlation coefficient of 0.58 (from T15 to T12) and 0.69 (from T12 to T15). Notably, feature autocorrelation across temporal lags of 40–100 ms alone attains a correlation of 0.57; consequently, CREWS sorts the electrode channels within each regional slot according to the mean rank of these autocorrelation values. Table 4 compares this proposed channel sorting strategy against the default recording sequence of the electrode arrays.

A.3 Patients and Data

Table A.2 lists the arrays and sentences of each patient.

Table A.2: Patients and data. Each patient has four 64-electrode Utah arrays; bold marks v6v, the only region shared by all three patients. Sentences are given as training / evaluation.

Patient	Array regions	Role	Sentences	Vocabulary
T15 (Card et al., 2024)	**v6v**, 4, 55b, d6v	source	9,022 / 100	open
T12 (Willett et al., 2023)	**v6v**, d6v, v44, d44	source	10,800 / 99	open
T16 (Kunz et al., 2025)	PEF, **v6v**, v6d, d6d	new patient	$k \in \{5, 10, 20, 60\}$ / 60	50 words

A.4 Region Removal on the Source Patients

Evaluating models after systematically removing individual cortical regions of a source patient at test time reproduces the regional importance rankings previously established for each dataset (Figure A.1). For subject T15, omitting ventral area 6v (v6v) induces the largest increase in PER, followed sequentially by area 55b and area 4, whereas excluding dorsal area 6v (d6v) leaves decoding performance virtually unchanged. This hierarchy directly mirrors the findings of Card et al.

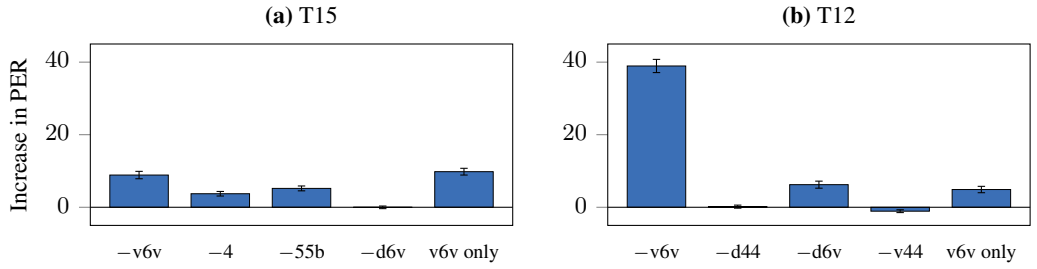


Figure A.1: Region removal on the source patients. Increase in PER of CREWS on T15 (a) and T12 (b) when one region is removed at test time or only v6v is retained. Error bars denote the standard error over trials.

(2024), who reported identical relative array contributions for T15 and observed no performance drop upon removing the d6v array. For subject T12, removing the v6v region raises PER by 38.9 points, whereas omitting either array located in area 44 shifts PER by at most 1.1 points. This observation aligns with the results reported by Willett et al. (2023), which demonstrated that area v6v provides the primary information for speech decoding while area 44 contributes minimally to speech production signaling. Consequently, although every regional slot is shared across the patient cohort, CREWS dynamically prioritizes functional cortical regions in accordance with the physiological rankings identified by single-subject decoders in prior literature.

A.5 Adaptation Schedules

Table A.3: Adaptation schedules of CREWS on T16 at $k = 5$. PER (%); the chance level of zero-shot CREWS on T16 is 91.4.

Schedules at k=5	T16
CREWS, patients sampled in turn	68.9
Input stage only, with the source data	75.7
70% of each batch from T16, with the source data	74.1
Fine-tuned on T16 only	74.6

Table A.3 compares adaptation schedules of CREWS on T16 at $k = 5$. Sampling 70% of each batch from T16, as in Levin et al. (2026), attains 74.1, and fine-tuning all parameters on the T16 sentences alone attains 74.6. Adapting only the input stage with the source data attains 75.7, compared with 68.9 for CREWS, which samples the patients in turn. These results indicate that sampling the patients in turn attains the lowest PER among the four schedules, by at least 5.2 points.

A.6 Language-Model Decoding

The phoneme posteriors of CREWS were decoded using the 3-gram language model with a 125,000-word vocabulary released by Willett et al. (2023), compiled into a weighted finite-state transducer and searched using the Kaldi decoder (Povey et al., 2011). The beam width, acoustic scale, and blank penalty were 17, 0.5, and $\log 9$ for T15, and 18, 0.8, and $\log 2$ for T12. The resulting word error rate is 14.2% on T15 and 21.9% on T12. Word error rate was not computed for T16, whose sentences are drawn from a closed 50-word vocabulary.