

Rapid parameter-efficient calibration for cross-session intracortical speech decoding across multiple datasets

Dibakar Roy

dibakarroy0804@gmail.com

Rajshahi University of Engineering and Technology

Fariya Tabassum

Rajshahi University of Engineering and Technology

Rana Mozumder

Bangladesh University of Engineering and Technology

Article

Keywords: intracortical brain–computer interface, speech decoding, cross-session calibration, neural adaptation, neural signal processing

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11038178/v1>

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

Additional Declarations: No competing interests reported.

Rapid parameter-efficient calibration for cross-session intracortical speech decoding across multiple datasets

Dibakar Roy^{1*}, Fariya Tabassum¹ and Rana Mozumder²

¹*Department of Electrical & Computer Engineering, Rajshahi
University of Engineering & Technology, Rajshahi, 6204, Bangladesh.

²Department of Biomedical Engineering, Bangladesh University of
Engineering & Technology, Dhaka, 1000, Bangladesh.

*Corresponding author(s). E-mail(s): dibakarroy0804@gmail.com;
Contributing authors: fariya@ece.ruet.ac.bd;
ranamozumder123@gmail.com;

Abstract

Intracortical speech brain–computer interfaces (BCIs) can restore communication for people who cannot speak, but longitudinal neural variability can degrade decoding across sessions. We implemented a parameter-efficient cross-session calibration framework using separate recurrent connectionist temporal classification (CTC) decoders for two public intracortical speech datasets (T12 and T15). Adaptation was restricted to a 256×256 affine transformation and bias at the neural input, while the recurrent encoder and CTC classifier remained frozen. Sessions were split chronologically; the calibration protocol was selected only on T12 validation sessions, frozen before later T12 testing, and transferred unchanged to T15. Only 65,792 parameters ($\approx 0.05\%$ of either model) were updated using 40 or 80 labeled trials. On T12, macro phoneme error rate improved by 11.37% and 15.51%, with all six held-out sessions improving under all three seeds. On T15, improvements were 2.72% and 10.05%, revealing greater session dependence. Extending prior work on high-performance and rapidly calibrating speech neuroprostheses, this study tests a fixed lightweight calibration rule across two independent datasets under a validation-locked protocol. These findings provide a leakage-resistant cross-dataset evaluation of rapid longitudinal recalibration and show that input-space adaptation can recover a substantial, but not universal, component of cross-session mismatch.

Keywords: intracortical brain–computer interface, speech decoding, cross-session calibration, neural adaptation, neural signal processing

1 Introduction

Loss of natural speech can substantially restrict communication, independence, and social participation in people with paralysis and neurological disorders that impair voluntary speech. Brain–computer interfaces (BCIs) offer a potential communication pathway by translating neural activity associated with intended or attempted actions into an external output. Cortical recordings have been used to synthesize intelligible speech directly from neural activity [1], and sentence-level communication was subsequently demonstrated in a person with anarthria [2]. Sequence-to-sequence modeling has also been used to translate electrocorticographic activity into text [3], while intracortical decoding of attempted handwriting demonstrated high-performance brain-to-text communication [4]. More recent intracortical speech systems have achieved substantially higher neural-to-text performance using attempted speech [5], and multimodal cortical interfaces have enabled simultaneous decoding of text, speech audio, and facial-avatar movements [6]. Accurate speech decoding after comparatively short initial training has further been demonstrated in a participant with amyotrophic lateral sclerosis [7]. Complementary intracranial studies have demonstrated online speech synthesis using a chronically implanted BCI in a person with ALS [8] and investigated the feasibility of decoding covert speech from electrocorticography using a Transformer trained on overt speech [9]. Together, these developments establish invasive speech BCIs as a promising approach for restoring communication.

High decoding accuracy during one recording period, however, does not ensure that a fixed decoder will remain reliable during later sessions. Longitudinal changes in recorded cortical signals and neural prosthetic decoder performance have been documented over extended recording periods [10], while substantial within-day neural signal instabilities have also been observed in human intracortical interfaces [11]. Such variability can arise from changes in the electrode–tissue interface, the recorded neural population, behavioral state, physiological modulation, and other sources of neural or measurement drift. Repeated-session adaptation has therefore been a longstanding problem in intracortical BCIs. Self-calibrating operation has been demonstrated for communication BCIs [12], robustness to future neural variability has been investigated explicitly [13], and rapid recalibration has been developed to reduce repeated setup requirements [14]. Alignment of low-dimensional neural representations can further stabilize decoding across recording instabilities [15]. Adaptive updating of an ECoG-based BCI using neural correlates of task performance has similarly shown the potential value of reducing repeated dedicated decoder-training requirements [16]. These findings indicate that temporal distribution shift and repeated calibration remain central challenges for neural interfaces intended for longitudinal use.

Cross-session neural decoding can be viewed as a domain-shift problem in which a decoder trained on earlier recording sessions is applied to a later target session. General neural domain-adaptation methods can reduce distribution mismatch [17, 18], but unrestricted model fine-tuning introduces many target-specific degrees of freedom and can be undesirable when only a small amount of labeled neural data is available. The broader BCI literature has therefore investigated transfer-learning strategies that reduce the amount of calibration data or retraining required for a new operating condition. Riemannian geometry-based transfer learning has, for example, been used to reduce training time in code-modulated visual evoked potential BCIs [19], while transfer of symmetric feature representations has been investigated for reducing training requirements in functional-near-infrared-spectroscopy BCIs [20]. Although these recording modalities and decoding tasks differ from intracortical speech decoding, they illustrate the broader need to minimize target-specific calibration burden while preserving decoder performance.

A more constrained approach for a high-capacity intracortical decoder is to preserve the temporal and phonetic sequence model and estimate only a compact session-specific transformation. Parameter-efficient adaptation follows a related principle in other machine-learning settings by restricting target-specific optimization to a small subset of model parameters [21, 22]. Applied to intracortical speech decoding, this restriction provides a controlled scientific question: how much decoder-relevant mismatch in a future recording session can be corrected near the neural input without relearning the recurrent temporal representation or phoneme classifier?

This question was examined using two independently recorded public intracortical speech datasets, referred to as T12 [5] and T15 [7]. Complete recording sessions were separated chronologically into earlier training sessions, intermediate validation sessions, and later held-out sessions. T12 was used for protocol development and chronological held-out evaluation. Adaptation scope and optimization duration were selected exclusively from T12 validation sessions before the final test sessions were examined. A separate high-capacity source decoder was trained for T15, but the target-session calibration rule selected during T12 development was transferred without retuning in response to T15 held-out performance.

The final procedure updates only a 256×256 affine transformation and bias for a future recording session. This corresponds to 65,792 trainable parameters, less than 0.05% of either complete decoder. Two small labeled calibration budgets, $K = 40$ and $K = 80$ trials, were evaluated. Performance was assessed using phoneme error rate (PER), session-level paired comparisons, repeated calibration seeds, substitution/deletion/insertion decomposition, sequence-length behavior, and direct computational measurements. Accordingly, the study contributes a compact affine-only calibration mechanism that updates less than 0.05% of the full decoder, a chronology-preserving validation and test design that prevents held-out-session tuning, and an independent T15 evaluation in which the calibration protocol selected on T12 is transferred without test-driven retuning. Error decomposition and direct computational measurements further characterize how the adaptation changes sequence errors, how strongly its benefit depends on calibration budget, and the computational cost

required to obtain that recovery. The overall decoding and future-session calibration framework is summarized in Fig. 1.

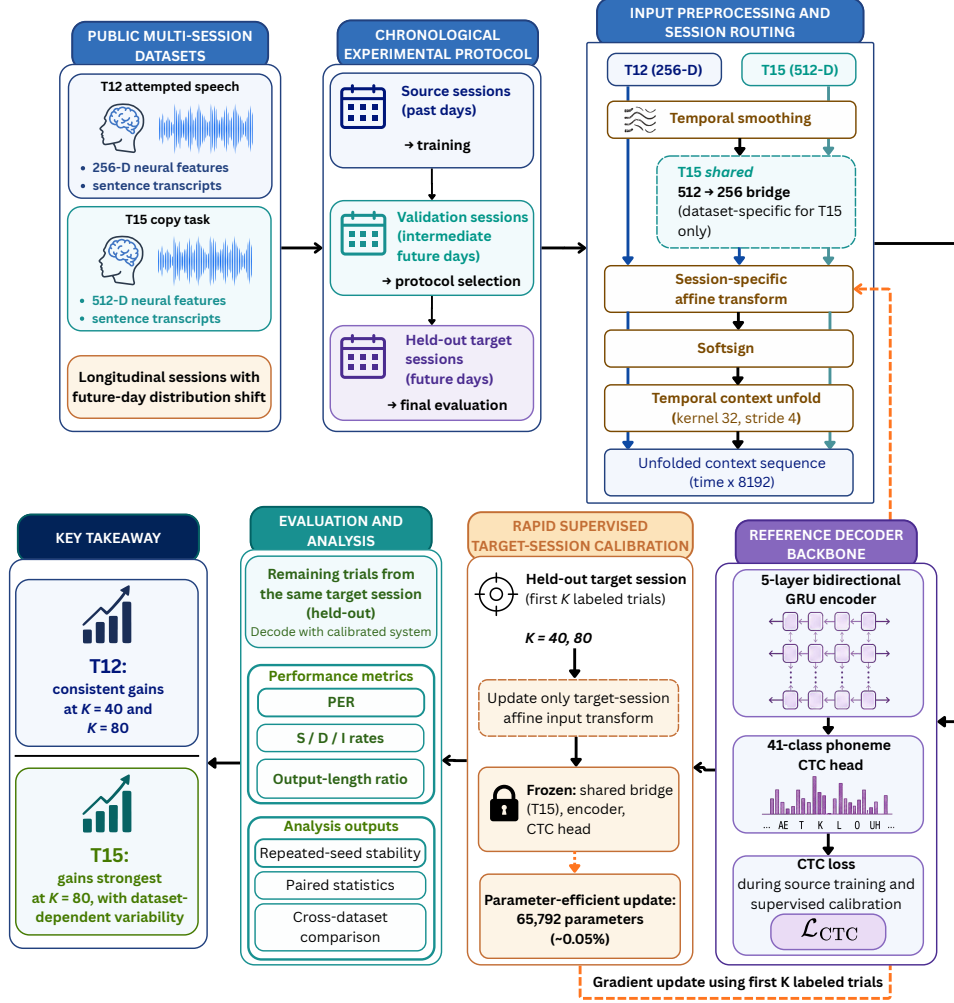


Fig. 1 Overall architecture for cross-session intracortical speech decoding and rapid calibration. Native neural features from T12 or T15 are temporally smoothed and mapped to a 256-dimensional decoder space, with T15 additionally using a learned 512-to-256 input bridge. A session-specific affine transformation precedes the Softsign activation, temporal unfolding, five-layer bidirectional GRU encoder, and phoneme-level CTC classifier. During future-session calibration, only the 65,792 parameters of the target-session affine transformation are updated; the input bridge, recurrent encoder, and CTC classifier remain frozen.

2 Results

2.1 Validation-only selection of the calibration protocol

Independent source decoders were established before any held-out target-session calibration was performed. The T12 source decoder reached its selected validation minimum at optimizer step 7,750, with validation PER 0.314116. The independently trained T15 source decoder reached its selected minimum at step 11,750, with validation PER 0.395386. All subsequent held-out comparisons used these fixed source checkpoints.

The target-session adaptation mechanism was selected using only the chronological T12 validation sessions. Three restricted parameter scopes were compared: residual-adapter-only, affine-only, and joint affine-plus-adapter adaptation. Aggregate validation phoneme error rate (PER) was 0.307066 for adapter-only adaptation, 0.295796 for affine-only adaptation, and 0.295989 for affine-plus-adapter adaptation. The difference between the two best configurations was small; therefore, a material decoding advantage of affine-only adaptation over the combined configuration is not implied. Affine-only calibration was retained because it achieved the lowest evaluated validation PER while requiring fewer trainable parameters.

Calibration duration was then refined while keeping the affine-only scope fixed. Validation PER decreased from 0.300810 after 25 optimization steps to 0.295796 after 50 steps, followed by values of 0.292836, 0.290487, 0.289866, 0.288768, and 0.286564 after 75, 100, 125, 150, and 200 steps, respectively. The 200-step configuration was consequently fixed before held-out evaluation. Figure 2 summarizes the validation-only protocol selection.

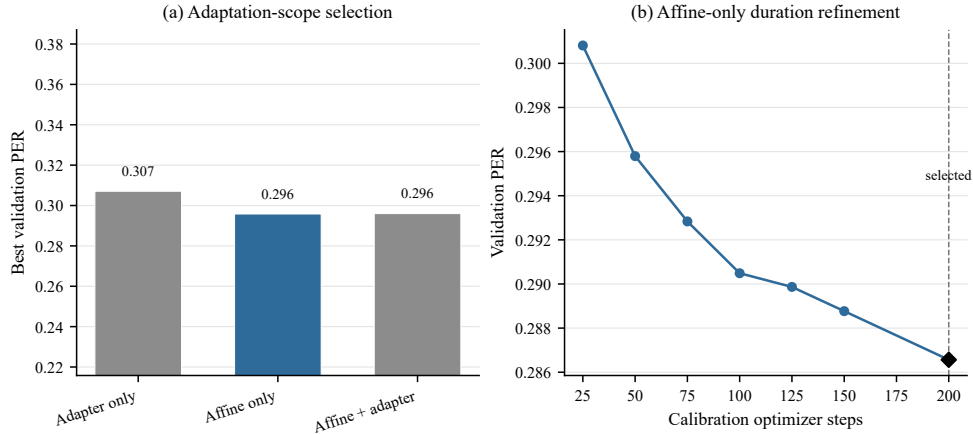


Fig. 2 Validation-based selection of the calibration protocol. Panel (a) compares restricted adaptation scopes using chronological T12 validation sessions. Panel (b) shows validation PER as the duration of affine-only calibration is increased. Affine-only adaptation with 200 optimization steps was fixed before held-out evaluation.

2.2 Calibration improves later-session T12 decoding

The frozen affine-only procedure was first evaluated on the six chronologically held-out T12 sessions. With $K = 40$ labeled trials, macro PER decreased from 0.4211 before calibration to 0.3732 ± 0.0004 after calibration, where the uncertainty denotes the standard deviation across three calibration seeds. This corresponds to an 11.37% relative reduction. Pooled PER decreased from 0.4279 to 0.3793. All six held-out sessions improved under every calibration seed.

Increasing the calibration budget to $K = 80$ produced a larger effect. Macro PER decreased from 0.4219 to 0.3565 ± 0.0006 , corresponding to a 15.51% relative reduction, while pooled PER decreased from 0.4293 to 0.3629. Again, all six sessions improved under all three calibration seeds. Session-level trajectories are shown in Fig. 3, and the aggregate results are summarized in Table 1.

For statistical comparison, the post-calibration PER for each session was first averaged across the three calibration seeds and paired with the corresponding before-calibration PER. An exact two-sided Wilcoxon signed-rank test yielded $p = 0.03125$ for both $K = 40$ and $K = 80$. Thus, the post-calibration PER was lower than the corresponding pre-calibration PER for every held-out T12 session, with both comparisons meeting the nominal $\alpha = 0.05$ reference threshold. Because four dataset-budget comparisons were evaluated without multiplicity adjustment, these p values are interpreted as nominal inferential evidence together with the effect magnitude and complete directional consistency rather than as multiplicity-adjusted confirmatory evidence.

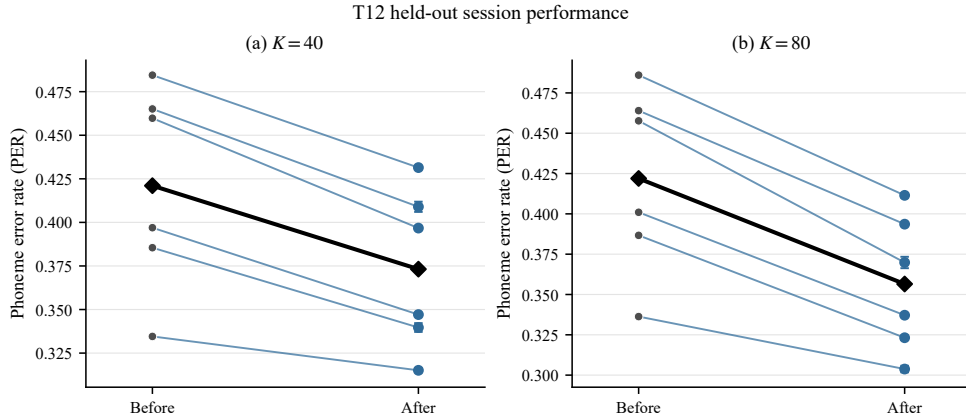


Fig. 3 Held-out T12 session performance. Panels show phoneme error rate before and after affine-only calibration for $K = 40$ and $K = 80$ labeled target-session trials. Each thin line represents one chronologically held-out recording session, and the thicker summary line denotes the macro session average. For both calibration budgets, all six sessions improved and the nominal exact two-sided Wilcoxon signed-rank test yielded $p = 0.03125$.

2.3 T12 calibration reduces substitution and deletion errors

Error decomposition provided additional information about the mechanisms underlying the T12 improvement. At $K = 40$, the pooled substitution rate decreased from 0.2754 to 0.2445 and the deletion rate decreased from 0.1057 to 0.0857, whereas the insertion rate changed only slightly from 0.0469 to 0.0491. The pooled decoded-to-reference phoneme-length ratio increased from 0.9412 to 0.9634.

At $K = 80$, the substitution rate decreased from 0.2766 to 0.2335 and the deletion rate from 0.1048 to 0.0810, while the insertion rate remained similar, changing from 0.0479 to 0.0485. The corresponding decoded-to-reference length ratio increased from 0.9432 to 0.9675.

These pooled diagnostics are descriptive and were used to characterize the composition of the PER reduction rather than as additional independent hypothesis tests. The simultaneous reduction in deletion errors and movement of the decoded-to-reference length ratio toward one indicate that calibration reduced under-decoding rather than merely exchanging one edit-error type for another (Fig. 4).

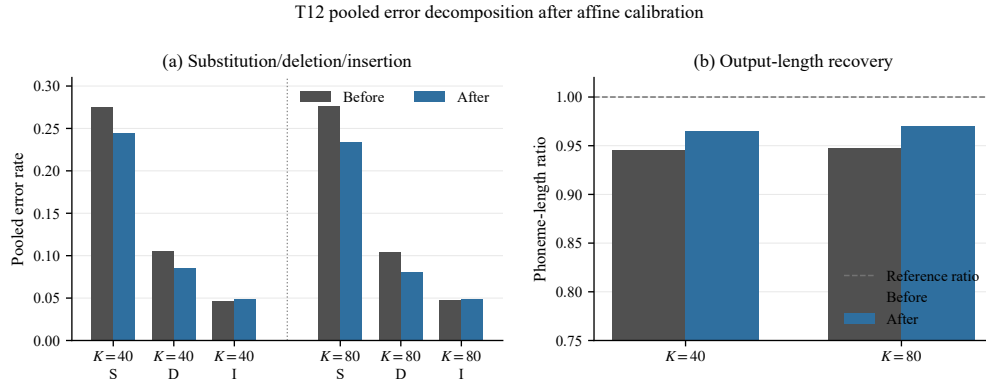


Fig. 4 T12 pooled error decomposition after affine-only calibration. Panel (a) shows pooled substitution, deletion, and insertion rates before and after calibration at $K = 40$ and $K = 80$. Panel (b) shows the corresponding pooled decoded-to-reference phoneme-length ratios; the reference value of one represents equal decoded and reference sequence lengths. These pooled diagnostics are descriptive; formal paired inference was performed on session-level PER as reported in Fig. 3 and Table 1.

2.4 Independent evaluation on T15

The same target-session calibration rule was next evaluated on T15 using a separately trained T15 source decoder. The learning rate, adaptation scope, number of optimization steps, calibration budgets, and repeated seed set were not retuned in response to the held-out T15 results.

At $K = 40$, all eight held-out T15 sessions satisfied the predeclared minimum evaluation-size requirement. Macro PER decreased from 0.6324 to 0.6152 ± 0.0041 , corresponding to a relative reduction of 2.72%. Pooled PER decreased from 0.6136 to

0.5835. Four of the eight sessions improved under every calibration seed. The nominal exact paired Wilcoxon value was $p = 0.546875$.

At $K = 80$, six of the eight T15 sessions remained eligible under the predefined requirement that at least 40 trials remain for evaluation after calibration. Macro PER decreased from 0.6123 to 0.5508 ± 0.0004 , corresponding to a 10.05% relative reduction, and pooled PER decreased from 0.6006 to 0.5260. Four of the six eligible sessions improved under all three seeds. The nominal exact paired Wilcoxon value was $p = 0.15625$.

Thus, neither T15 condition fell below the nominal 0.05 reference threshold, and the session-level response was more heterogeneous than on T12. Figure 5 shows the paired T15 results.

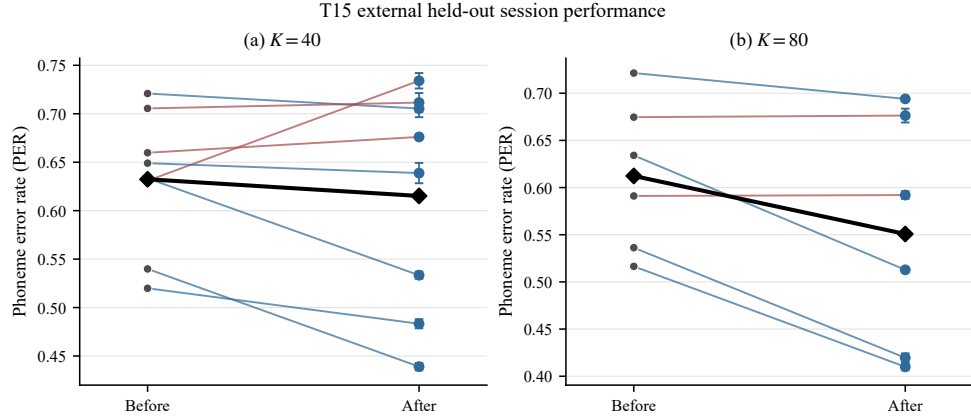


Fig. 5 Held-out T15 session performance. Panels show phoneme error rate before and after affine-only calibration for $K = 40$ and $K = 80$. All eight held-out sessions are included for $K = 40$, whereas six satisfy the predeclared evaluation-size criterion for $K = 80$. The same calibration hyperparameters selected during T12 development were used without held-out T15 retuning.

2.5 T15 error mechanisms reveal stronger budget dependence

The T15 error decomposition differed from the T12 pattern. At $K = 40$, the macro deletion rate decreased substantially from 0.171822 to 0.107795, and the mean decoded-to-reference sequence-length ratio increased from 0.850713 to 0.950242. However, the substitution rate increased from 0.438044 to 0.449359 and the insertion rate increased from 0.022535 to 0.058037. The large reduction in deletion errors was therefore partly offset by increases in the other edit-error components.

At $K = 80$, a more favorable pattern emerged. The substitution rate decreased from 0.423760 to 0.396516 and the deletion rate from 0.164680 to 0.106599. The insertion rate increased from 0.023893 to 0.047677, while the mean decoded-to-reference sequence-length ratio improved from 0.859213 to 0.941077.

These results indicate that the smaller T15 calibration budget primarily corrected under-decoding but did not consistently improve phoneme identity, whereas the larger budget produced reductions in both substitution and deletion errors (Fig. 6).

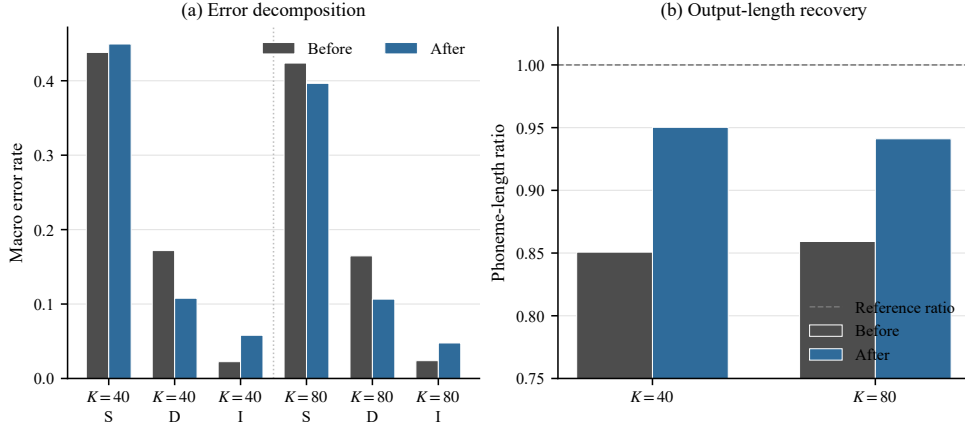


Fig. 6 T15 error decomposition after affine-only calibration. Panel (a) summarizes macro-averaged substitution, deletion, and insertion rates before and after calibration at $K = 40$ and $K = 80$. Panel (b) shows the corresponding macro-averaged decoded-to-reference phoneme-length ratios. The $K = 40$ condition primarily reduces deletion errors, whereas the larger $K = 80$ budget also reduces substitution errors.

2.6 Cross-dataset comparison and computational efficiency

The aggregate comparison in Fig. 7 demonstrates two related but distinct calibration behaviors. T12 exhibits a consistent future-session benefit at both calibration budgets, whereas T15 shows greater session dependence and a substantially stronger benefit at the larger budget. These observations do not imply that neural drift is itself affine. Rather, they indicate that the fraction of decoder-relevant session mismatch correctable by an affine input transformation differs across datasets, recording sessions, and calibration budgets.

Only the session-specific 256×256 affine matrix and its 256-dimensional bias were updated during final target-session calibration. The resulting 65,792 trainable parameters correspond to 0.049398% of the 133,187,113-parameter T12 source decoder and 0.048939% of the 134,436,649-parameter T15 decoder. More than 99.95% of either pretrained model therefore remained unchanged.

Mean T12 calibration time was 7.52 s for $K = 40$ and 7.56 s for $K = 80$, with peak allocated GPU memory of approximately 1.28 GB. T15 required 20.12 s for $K = 40$ and 23.46 s for $K = 80$, with peak allocated GPU memory of approximately 1.80 GB and 1.91 GB, respectively. These measurements quantify both the restricted optimization scope and the computational cost of future-session calibration.

Table 1 summarizes the principal held-out results across both datasets and calibration budgets.

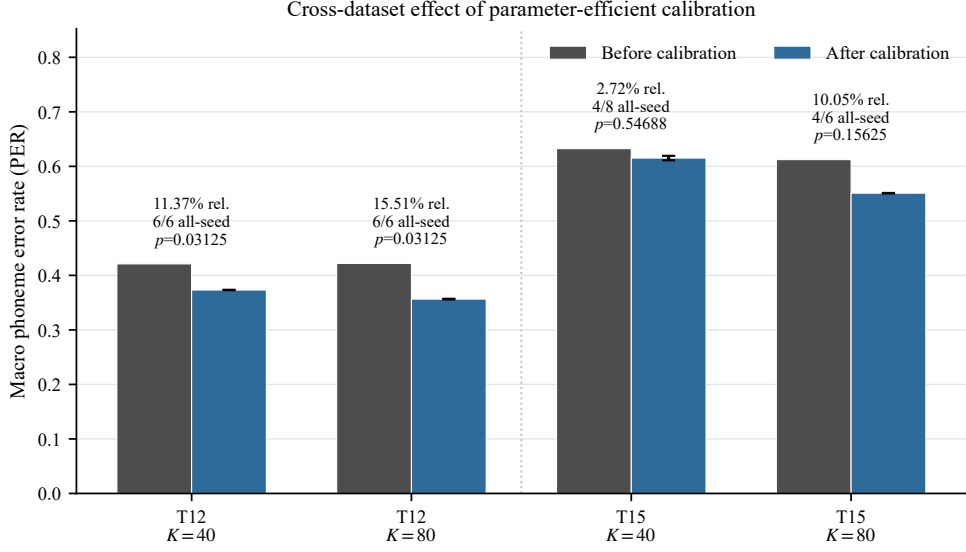


Fig. 7 Cross-dataset effect of parameter-efficient calibration. Macro phoneme error rate is shown before and after calibration for T12 and T15 at $K = 40$ and $K = 80$. Error bars represent the across-seed standard deviation of the post-calibration macro PER. Annotations summarize relative improvement, all-seed session consistency, and nominal exact paired Wilcoxon values.

Table 1 Final held-out cross-dataset results. Post-calibration macro PER is the mean across eligible sessions and is reported as mean $\pm$ standard deviation across three calibration seeds. The p values are nominal exact two-sided Wilcoxon signed-rank values computed from paired session-level observations.

Dataset	K	n	Macro before	Macro after	Rel. imp.	Pooled before	Pooled after	All-seed imp.
T12	40	6	0.4211	0.3732 ± 0.0004	11.37%	0.4279	0.3793	6/6
T12	80	6	0.4219	0.3565 ± 0.0006	15.51%	0.4293	0.3629	6/6
T15	40	8	0.6324	0.6152 ± 0.0041	2.72%	0.6136	0.5835	4/8
T15	80	6	0.6123	0.5508 ± 0.0004	10.05%	0.6006	0.5260	4/6

Nominal p values: T12 $K = 40$, 0.03125; T12 $K = 80$, 0.03125; T15 $K = 40$, 0.546875; T15 $K = 80$, 0.15625.

3 Discussion

This study examined whether future-session degradation of an intracortical speech decoder can be reduced without retraining its high-capacity recurrent sequence model. Under a strictly chronological evaluation, adaptation confined to a 65,792-parameter affine input transformation produced substantial and highly consistent recovery on T12. Relative macro PER reductions were 11.37% with 40 labeled trials and 15.51% with 80 trials, with every held-out T12 session improving under all three repeated calibration seeds. The same frozen adaptation principle also improved aggregate decoding

on the independent T15 dataset, particularly at the larger calibration budget, although session-level responses were more heterogeneous.

The T12 results suggest that an important component of the mismatch between the source decoder and these later recording sessions was expressed sufficiently close to the neural input to be corrected without relearning the recurrent temporal representation or phoneme classifier. This interpretation is consistent with previous work showing that alignment of session-dependent neural representations can stabilize longitudinal intracortical decoding [15]. It should not, however, be interpreted as evidence that the biological processes underlying neural drift are inherently affine. The transformation is better understood as a decoder-level correction that can compensate for a useful fraction of the distribution shift encountered by the fixed sequence model.

The independent T15 evaluation provides an important boundary to this interpretation. The $K = 40$ condition produced only a 2.72% reduction in macro PER, and only half of the held-out sessions improved under all three calibration seeds. With $K = 80$, the relative macro reduction increased to 10.05%, but the session-level response remained less uniform than on T12. Neither T15 condition fell below the nominal 0.05 reference threshold. The T15 result therefore argues against a universal claim that a single low-capacity transformation can remove cross-session mismatch. Instead, it indicates that the correctable fraction depends on both the target recording session and the amount of calibration information available.

The error decomposition helps explain the budget dependence. On T12, both substitution and deletion rates decreased at the two budgets, and decoded sequence length moved closer to the reference length. The $K = 40$ T15 condition behaved differently: a large decrease in deletion errors and substantial recovery of sequence length were accompanied by increases in substitutions and insertions. With 80 calibration trials, both substitution and deletion errors decreased. A larger labeled sample may therefore provide a more reliable estimate of the target-session input transformation when the target distribution differs more strongly from the source distribution.

Validation ablations also support restricting the final update to the input transformation. Affine-only and affine-plus-adapter calibration produced almost identical validation PER values (0.295796 and 0.295989, respectively). The combined configuration is therefore not shown to be inferior in a practically meaningful statistical sense. The simpler affine-only protocol was selected because the additional residual adapter did not improve the evaluated validation accuracy and introduced more trainable parameters. This distinction is important because the study is not intended to establish affine adaptation as universally superior to more expressive parameter-efficient methods. Rather, it tests how far a deliberately restricted input-space correction can recover a pretrained decoder.

The restricted update also has practical computational implications. More than 99.95% of each source decoder remained frozen, and target-session optimization required approximately 8 s on T12 and 20–23 s on T15 on the evaluated GPU. Parameter-efficient calibration does not remove the need to collect labeled attempted-speech examples, but it separates that calibration burden from the much larger

507 computational cost of source-model training and full-model optimization. Such a separation
508 could be useful in future systems in which a stable sequence decoder is retained
509 over time while compact session-specific components are updated.

510 Several limitations constrain the conclusions. The study evaluates two public
511 intracortical speech datasets, each representing a single BrainGate2 participant, and
512 all analyses are offline. Generalization to additional participants, implant configurations,
513 longer recording periods, and closed-loop communication therefore remains
514 unknown. The calibration procedure is supervised and requires 40 or 80 labeled target-
515 session trials. The validation ablation compares restricted affine and adapter scopes
516 but does not exhaust the range of modern domain-adaptation or parameter-efficient
517 approaches. T12 and T15 also differ in native feature dimensionality, recording chronology,
518 and source-model construction, so the T15 experiment should be interpreted as
519 independent cross-dataset evaluation rather than a controlled single-factor replication.

520 Within these limitations, the results show that a meaningful component of decoder-
521 relevant future-session shift can be corrected with a very small fraction of the full model
522 parameters. The consistency observed on T12 and the more variable behavior observed
523 on T15 are both informative: restricted input-space adaptation can provide rapid
524 recovery when the target mismatch is compatible with the transformation, but greater
525 calibration data or more expressive adaptation may be required for more difficult sessions.
526 Future work should therefore evaluate label-efficient and label-free alternatives,
527 additional participants, and closed-loop longitudinal operation while preserving the
528 strict separation between adaptation development and genuinely future-session testing.

529

530 4 Methods

531

532 4.1 Study design

533

534 The study was designed to evaluate rapid adaptation to a later recording session while
535 preventing information from future sessions from influencing model or protocol development.
536 Complete recording sessions were therefore partitioned chronologically rather
537 than by randomly splitting individual trials. Earlier sessions were used for source-
538 model training, subsequent sessions were used for validation and protocol selection,
539 and the latest sessions were held out for final evaluation.

540 T12 served as the development dataset. Source-checkpoint selection, initialization
541 auditing, adaptation-scope comparison, and calibration duration were completed
542 using its training and validation partitions before final T12 testing. T15 used a separately
543 trained source decoder. The final target-session calibration scope, learning
544 rate, number of optimization steps, calibration budgets, and seed set selected during
545 T12 development were transferred to the T15 held-out evaluation without retuning in
546 response to T15 test performance.

547

548 4.2 Public intracortical speech datasets and ethics

549

550 The analyses used two publicly released intracortical attempted-speech datasets. T12
551 was released with the high-performance speech neuroprosthesis study of Willett et
552 al. [5, 23], and T15 was released with the rapidly calibrating speech neuroprosthesis

study of Card et al. [7, 24]. No new neural recordings or participant recruitment were performed for the present study.

Both datasets originated from participants enrolled in the BrainGate2 Neural Interface System pilot clinical trial (ClinicalTrials.gov NCT00912041). The original T12 data collection was conducted under the US Food and Drug Administration Investigational Device Exemption G090003 and Stanford University Institutional Review Board protocol 20804, and the participant provided informed consent. The original T15 study was conducted under the same BrainGate2 Investigational Device Exemption, with Institutional Review Board approval from Mass General Brigham (2009P000505) and the University of California, Davis (1843264), and informed consent was obtained from the participant. The original studies report that research procedures were performed in accordance with relevant institutional, regulatory, and ethical guidelines. The present work consisted solely of secondary computational analysis of the publicly released, de-identified datasets and included no new intervention or data collection.

4.3 Chronological data partitions

For T12, 24 recording sessions containing 9,680 labeled attempted-speech trials were retained. The first 14 sessions (5,360 trials) formed the source-training partition, the subsequent four sessions (1,560 trials) formed the validation partition, and the final six sessions (2,760 trials) were held out for final evaluation. The training period extended from April 28 to July 5, 2022. Validation sessions were recorded on July 14, July 21, July 27, and July 29, and the held-out sessions were recorded on August 2, August 11, August 13, August 18, August 23, and August 25, 2022.

For T15, 45 sessions containing 9,498 labeled trials were retained. The first 31 sessions (6,972 trials) formed the training partition, the next six sessions (1,275 trials) were used for validation, and the final eight sessions (1,251 trials) were reserved for held-out evaluation. The held-out sessions were recorded on July 21 and July 28, 2024, and January 10, January 12, March 14, March 16, March 30, and April 13, 2025.

Within a held-out session, the first K chronological trials were used for supervised calibration and all remaining eligible trials were reserved for evaluation. For T15, chronological trial order was reconstructed using the released recording-block identifier and within-block trial index because dataframe storage order does not consistently reproduce recording chronology.

No evaluation trial was used for gradient computation, source-checkpoint selection, adaptation-scope selection, or calibration-hyperparameter tuning.

4.4 Neural features and target sequences

Each trial consisted of a variable-length neural feature sequence X_i and an ordered phoneme target sequence y_i . T12 contains 256 neural features per time bin and T15 contains 512 features per time bin. The released feature representations were retained without per-trial z -score normalization.

Temporal Gaussian smoothing with standard deviation $\sigma = 2$ was applied in the native feature space of each dataset. Let $x_t^{(d)} \in \mathbb{R}^{D_d}$ denote the native neural feature

599 vector for dataset d at time t . The smoothed representation can be written as

$$600 \quad \tilde{x}_t^{(d)} = \sum_{\tau} g_{\sigma}(\tau) x_{t-\tau}^{(d)}, \quad (1)$$

603 where g_{σ} denotes the Gaussian smoothing kernel.

604 For T12, $D_{T12} = 256$ and the smoothed representation enters the session-specific
606 affine transformation directly. For T15, $D_{T15} = 512$ and the smoothed representation
607 is first mapped to a 256-dimensional calibration space using a learned bias-free bridge,

$$608 \quad u_t^{(T15)} = W_{\text{bridge}} \tilde{x}_t^{(T15)}, \quad W_{\text{bridge}} \in \mathbb{R}^{256 \times 512}. \quad (2)$$

611 The bridge is learned during T15 source-model training and remains frozen during
612 target-session calibration.

613 Both datasets use the same structural target space of 40 phoneme classes. Con-
614 nectionist temporal classification (CTC) blank is assigned index zero, producing 41
615 output classes in total.

617 4.5 Source decoder architecture

618 After native-space smoothing and the optional T15 input bridge, a session-specific
619 affine transformation maps the representation into the decoder input space,

$$620 \quad h_t = W_s u_t + b_s, \quad W_s \in \mathbb{R}^{256 \times 256}, \quad b_s \in \mathbb{R}^{256}. \quad (3)$$

623 For T12, $u_t = \tilde{x}_t^{(T12)}$.

625 A Softsign nonlinearity is applied to the affine output. Temporal context is then
626 constructed by unfolding 32 consecutive 256-dimensional feature vectors with stride
627 four, producing an 8,192-dimensional context vector for each recurrent step. The
628 unfolded sequence is processed by a five-layer bidirectional gated recurrent unit
629 (BiGRU) encoder with 1,024 hidden units in each direction per layer. Gated recur-
630 rent units provide recurrent state updates suitable for variable-length sequential
631 representations [25].

632 The final recurrent representation is projected to 41 logits,

$$633 \quad o_n = W_{\text{CTC}} z_n + b_{\text{CTC}}, \quad (4)$$

636 where z_n denotes the final BiGRU representation. CTC is used because the frame-level
637 alignment between the neural sequence and target phoneme sequence is not observed
638 explicitly [26]. For target sequence y ,

$$639 \quad P(y|X) = \sum_{\pi \in B^{-1}(y)} P(\pi|X), \quad (5)$$

where π denotes an alignment path and $\mathcal{B}$ is the standard CTC collapse operation. Source-model training and supervised target-session calibration minimize

$$\mathcal{L}_{\text{CTC}} = -\log P(y|X). \quad (6)$$

Separate source decoders were trained for T12 and T15. T15 was not obtained by fine-tuning the T12 model.

4.6 Source-model training

Both source models were optimized using Adam [27] with learning rate 0.02, $\epsilon = 0.1$, and weight decay 10^{-5} . A physical mini-batch contained four trials, and gradients were accumulated across 16 mini-batches to produce an effective batch size of 64. bfloat16 (BF16) precision was used during source-model training.

Training examples were sampled across available source sessions to reduce dominance by recording days containing more trials. Session-specific affine transformations were optimized jointly with the shared recurrent decoder. For T15, the bias-free 512-to-256 bridge was also learned during source training.

Neural augmentation was applied only during source-model training. For native neural vector x_t , the augmented representation was

$$x_t^{\text{aug}} = x_t + \epsilon_t + c, \quad (7)$$

where ϵ_t is time-varying white Gaussian perturbation with standard deviation 0.8 and c is a trial-level constant feature offset sampled with standard deviation 0.2. Augmentation preceded temporal smoothing and was disabled during validation, target-session calibration, and held-out evaluation.

Checkpoint selection used only validation-session PER. The selected T12 checkpoint was optimizer step 7,750 (validation PER 0.314116), and the selected T15 checkpoint was step 11,750 (validation PER 0.395386).

4.7 Validation-only selection of adaptation scope

Three restricted target-session parameter scopes were evaluated on T12 validation sessions:

$$\Theta_{\text{adapt}} \in \{\Theta_{\text{affine}}, \Theta_{\text{adapter}}, \Theta_{\text{affine}} \cup \Theta_{\text{adapter}}\}. \quad (8)$$

The affine-only condition updates the target-session affine weight and bias. A compact residual adapter was considered as a validation ablation. If h_t denotes the representation entering this component, the adapted representation is

$$h_t^{\text{adapt}} = h_t + A_\phi(h_t), \quad (9)$$

where A_ϕ is the trainable residual transformation.

Validation PER was 0.307066 for adapter-only calibration, 0.295796 for affine-only calibration, and 0.295989 for joint affine-plus-adapter calibration. Affine-only calibration was selected because it achieved the lowest evaluated validation PER while

691 requiring fewer parameters than the combined scope. The residual adapter is therefore
 692 an ablation component and is not present in final T12 or T15 held-out calibration.
 693 Optimization duration was subsequently selected using the validation data, with
 694 the lowest evaluated value obtained at 200 steps. Both adaptation scope and duration
 695 were fixed before final held-out testing.

697 4.8 Initialization of an unseen target session

698 A held-out future session does not have a previously learned affine transformation.
 699 Its initial affine parameters were therefore formed from the elementwise mean of the
 700 affine weights and biases associated with the last three source-training sessions:
 701

$$702 \quad W_{\text{init}} = \frac{1}{3} \sum_{j=1}^3 W_{s_j}, \quad b_{\text{init}} = \frac{1}{3} \sum_{j=1}^3 b_{s_j}. \quad (10)$$

703 The same initialization was used for the unadapted baseline and as the starting
 704 point for target-session calibration. A fresh copy was created for every held-out ses-
 705 sion, calibration budget, and random seed; parameters were never carried from one
 706 condition to another.

711 4.9 Future-session affine calibration

712 For each held-out session s , the first K chronologically ordered labeled trials formed
 713 the calibration set,
 714

$$715 \quad \mathcal{C}_{s,K} = \{(X_i, y_i)\}_{i=1}^K, \quad K \in \{40, 80\}. \quad (11)$$

716 The remaining eligible trials formed the evaluation set $\mathcal{E}_{s,K}$.

717 Only the target-session affine parameters $\{W_{\text{cal}}, b_{\text{cal}}\}$ were optimized:

$$718 \quad \theta_{s,K}^* = \arg \min_{\{W_{\text{cal}}, b_{\text{cal}}\}} \sum_{(X_i, y_i) \in \mathcal{E}_{s,K}} \mathcal{L}_{\text{CTC}}(X_i, y_i). \quad (12)$$

719 The recurrent encoder and CTC classifier remained frozen. The T15 input bridge
 720 also remained frozen. The number of trainable parameters was therefore

$$721 \quad 256 \times 256 + 256 = 65,792 \quad (13)$$

722 for both datasets.

723 Calibration used Adam with learning rate 0.001, $\epsilon = 0.1$, zero weight decay, 200
 724 optimizer steps, batch size four, and gradient-norm clipping at five. Neural augmen-
 725 tation was disabled. Calibration examples were sampled with replacement so that
 726 optimization duration remained fixed for both budgets.

727 Three calibration seeds (0, 1, and 2) were evaluated. The source checkpoint, initial-
 728 ization, calibration trials, hyperparameters, and evaluation trials were identical across
 729 seeds; only the stochastic sampling order of calibration examples differed.

For T15, a predeclared minimum of 40 trials was required to remain for evaluation. All eight sessions were therefore eligible at $K = 40$. At $K = 80$, six sessions remained eligible. The March 14, 2025 session would have left only three evaluation trials, and the April 13, 2025 session only 14, so these sessions were excluded from the $K = 80$ analysis independently of their decoding performance.

4.10 Phoneme decoding and evaluation metrics

The primary analysis used greedy phoneme-level CTC decoding, thereby excluding downstream language-model correction from the calibration comparison. At each recurrent step,

$$\hat{\pi}_n = \arg \max_c P(c|X, n), \quad (14)$$

and the final phoneme sequence was obtained by applying the standard CTC collapse operator $\mathcal{B}$.

For a reference phoneme sequence and prediction, PER was defined as

$$\text{PER} = \frac{S + D + I}{N_{\text{ph}}}, \quad (15)$$

where S , D , and I denote substitution, deletion, and insertion counts and N_{ph} is the number of reference phonemes.

The primary dataset-level summary was macro PER,

$$\text{PER}_{\text{macro}} = \frac{1}{M} \sum_{s=1}^M \text{PER}_s, \quad (16)$$

which assigns equal weight to each eligible held-out session. Complementary pooled PER was computed from total edit errors and total reference phonemes across the eligible evaluation set.

Relative improvement was calculated as

$$\Delta_{\text{rel}} = 100 \frac{\text{PER}_{\text{before}} - \text{PER}_{\text{after}}}{\text{PER}_{\text{before}}}. \quad (17)$$

Sequence-length behavior was measured using the decoded-to-reference phoneme-length ratio,

$$r_{\text{len},s} = \frac{\sum_{i \in \mathcal{E}_{s,K}} |\hat{y}_i|}{\sum_{i \in \mathcal{E}_{s,K}} |y_i|}. \quad (18)$$

Values below one indicate that decoded sequences are shorter than the references. This diagnostic was interpreted together with explicit deletion rate rather than as an independent accuracy measure.

For T12, substitution, deletion, insertion, and length-ratio diagnostics were pooled across eligible sessions. For T15, these quantities were computed separately by session, averaged across calibration seeds, and then macro-averaged across eligible sessions.

783 4.11 Statistical analysis

784 The unit of paired statistical analysis was the held-out recording session, rather than
 785 individual trials or calibration seeds. For each session and budget, the post-calibration
 786 PER was first averaged over the three calibration seeds and paired with its corre-
 787 sponding before-calibration PER. This prevents repeated stochastic calibrations of the
 788 same recording session from being treated as statistically independent observations.

789 Because the numbers of held-out sessions were small and no normality assumption
 790 for session-level PER differences was justified, an exact two-sided Wilcoxon signed-
 791 rank test was applied to

$$792 \delta_{s,K} = \text{PER}_{s,K}^{\text{after}} - \text{PER}_{s,K}^{\text{before}}. \quad (19)$$

793 The independent session counts were $n = 6$ for both T12 budgets, $n = 8$ for T15
 794 at $K = 40$, and $n = 6$ for T15 at $K = 80$. Exact two-sided p values are reported.
 795 Four prespecified dataset–budget tests were evaluated. No multiplicity-adjusted con-
 796 firmatory inference is claimed; the p values are therefore treated as nominal inferential
 797 summaries and interpreted together with effect size, direction of session-level changes,
 798 all-seed consistency, and macro and pooled PER. A nominal $\alpha = 0.05$ threshold is
 799 used only as a descriptive reference.

800 Across-seed variability of the post-calibration macro PER is reported as the sample
 801 standard deviation over the three calibration seeds.

802 4.12 Computational analysis and reproducibility

803 Parameter efficiency for dataset d was calculated as

$$804 \rho_d = 100 \frac{N_{\text{adapt}}}{N_{\text{model},d}}, \quad (20)$$

805 where $N_{\text{adapt}} = 65,792$. Calibration duration was recorded for every eligible held-
 806 out session and seed, and peak allocated GPU memory was measured during each
 807 calibration run.

808 Experiments were implemented in Python using PyTorch [28] and CUDA acceler-
 809 ation. Final computational measurements were obtained using an NVIDIA GeForce
 810 RTX 4060 GPU. Source data, chronological partitions, final analysis tables, plotting
 811 scripts, and reproducibility documentation are described in the public repository cited
 812 below. The immutable release associated with this manuscript fixes the submitted
 813 implementation and final numerical artifacts.

814 4.13 Use of artificial intelligence-assisted tools

815 ChatGPT and Codex (OpenAI) were used as software-development and writing-
 816 support tools for selected code drafting, debugging, reproducibility documentation,
 817 manuscript organization, and language refinement. All experimental protocols, chrono-
 818 logical partitions, model configurations, numerical analyses, statistical procedures,
 819 figures, and reported results were executed, reviewed, and verified by the authors

against the saved experimental artifacts. Scientific conclusions and final interpretive judgments were made by the authors, who take full responsibility for the content of the manuscript.

Data availability The T12 intracortical speech dataset analysed in this study is publicly available through the Dryad Digital Repository under <https://doi.org/10.5061/dryad.x69p8czpq>. The T15 dataset is publicly available through Dryad under <https://doi.org/10.5061/dryad.dncjsxm85>. No new human-participant data were collected for the present study.

Code availability Source code and reproducibility resources for neural data preprocessing, source-model training, validation-only protocol selection, target-session calibration, held-out evaluation, statistical analysis, and figure generation are publicly available at <https://github.com/roy056/rapid-cross-session-ibci-calibration>. The version corresponding to this manuscript is preserved as the immutable release v1.0.0-paper-submission, targeting commit d867b99f3cffc3b9462097201038fe804bfeac7d.

References

- [1] Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498; <https://doi.org/10.1038/s41586-019-1119-1> (2019).
- [2] Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227; <https://doi.org/10.1056/NEJMoa2027540> (2021).
- [3] Makin, J. G., Moses, D. A. & Chang, E. F. Machine translation of cortical activity to text with an encoder–decoder framework. *Nat. Neurosci.* **23**, 575–582; <https://doi.org/10.1038/s41593-020-0608-8> (2020).
- [4] Willett, F. R. et al. High-performance brain-to-text communication via handwriting. *Nature* **593**, 249–254; <https://doi.org/10.1038/s41586-021-03506-2> (2021).
- [5] Willett, F. R. et al. A high-performance speech neuroprosthesis. *Nature* **620**, 1031–1036; <https://doi.org/10.1038/s41586-023-06377-x> (2023).
- [6] Metzger, S. L. et al. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature* **620**, 1037–1046; <https://doi.org/10.1038/s41586-023-06443-4> (2023).
- [7] Card, N. S. et al. An accurate and rapidly calibrating speech neuroprosthesis. *N. Engl. J. Med.* **391**, 609–618; <https://doi.org/10.1056/NEJMoa2314132> (2024).
- [8] Angrick, M. et al. Online speech synthesis using a chronically implanted brain–computer interface in an individual with ALS. *Sci. Rep.* **14**, 9617; <https://doi.org/10.1038/s41598-024-60277-2> (2024).

- 875 [9] Komeiji, S., Mitsuhashi, T., Iimura, Y. et al. Feasibility of decoding covert speech
876 in ECoG with a Transformer trained on overt speech. *Sci. Rep.* **14**, 11491; <https://doi.org/10.1038/s41598-024-62230-9> (2024).
877
878
- 879 [10] Chestek, C. A. et al. Long-term stability of neural prosthetic control signals from
880 silicon cortical arrays in rhesus macaque motor cortex. *J. Neural Eng.* **8**, 045005;
881 <https://doi.org/10.1088/1741-2560/8/4/045005> (2011).
882
- 883 [11] Perge, J. A., Homer, M. L., Malik, W. Q. et al. Intra-day signal instabilities affect
884 decoding performance in an intracortical neural interface system. *J. Neural Eng.*
885 **10**, 036004; <https://doi.org/10.1088/1741-2560/10/3/036004> (2013).
886
- 887 [12] Jarosiewicz, B. et al. Virtual typing by people with tetraplegia using a self-
888 calibrating intracortical brain–computer interface. *Sci. Transl. Med.* **7**, 313ra179;
889 <https://doi.org/10.1126/scitranslmed.aac7328> (2015).
890
- 891 [13] Sussillo, D. et al. Making brain–machine interfaces robust to future neu-
892 ral variability. *Nat. Commun.* **7**, 13749; <https://doi.org/10.1038/ncomms13749>
893 (2016).
894
- 895 [14] Brandman, D. M. et al. Rapid calibration of an intracortical brain–computer
896 interface for people with tetraplegia. *J. Neural Eng.* **15**, 026007; <https://doi.org/10.1088/1741-2552/aa9ee7> (2018).
897
- 898 [15] Degenhart, A. D. et al. Stabilization of a brain–computer interface via the align-
899 ment of low-dimensional spaces of neural activity. *Nat. Biomed. Eng.* **4**, 672–685;
900 <https://doi.org/10.1038/s41551-020-0542-9> (2020).
901
- 902 [16] Rouanne, V., Costecalde, T., Benabid, A. L. et al. Unsupervised adaptation
903 of an ECoG based brain–computer interface using neural correlates of task
904 performance. *Sci. Rep.* **12**, 21316; <https://doi.org/10.1038/s41598-022-25049-w>
905 (2022).
906
- 907 [17] Farshchian, A. et al. Adversarial domain adaptation for stable brain–machine
908 interfaces. In *International Conference on Learning Representations* (ICLR,
909 2019).
910
- 911 [18] Ganin, Y. et al. Domain-adversarial training of neural networks. *J. Mach. Learn.*
912 *Res.* **17**, 1–35 (2016).
913
- 914 [19] Ying, J., Wei, Q. & Zhou, X. Riemannian geometry-based transfer learning for
915 reducing training time in c-VEP BCIs. *Sci. Rep.* **12**, 9818; <https://doi.org/10.1038/s41598-022-14026-y> (2022).
916
- 917 [20] Khalil, K., Asgher, U. & Ayaz, Y. Novel fNIRS study on homogeneous symmetric
918 feature-based transfer learning for brain–computer interface. *Sci. Rep.* **12**, 3198;
919 <https://doi.org/10.1038/s41598-022-06805-4> (2022).
920

- [21] Housby, N. et al. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, **97**, 2790–2799 (2019).
[22] Hu, E. J. et al. LoRA: low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR, 2022)*.
[23] Willett, F. R. et al. Data for: A high-performance speech neuroprosthesis. *Dryad*; <https://doi.org/10.5061/dryad.x69p8czpq> (2023).
[24] Card, N. S. et al. Data for: An accurate and rapidly calibrating speech neuroprosthesis. *Dryad*; <https://doi.org/10.5061/dryad.dncjsxm85> (2025).
[25] Cho, K. et al. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 1724–1734; <https://doi.org/10.3115/v1/D14-1179> (2014).
[26] Graves, A., Fernández, S., Gomez, F. & Schmidhuber, J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine Learning*, 369–376; <https://doi.org/10.1145/1143844.1143891> (2006).
[27] Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *International Conference on Learning Representations (ICLR, 2015)*.
[28] Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **32** (2019).

Acknowledgements The authors thank the participants and research teams whose efforts made the public T12 and T15 intracortical speech datasets available to the research community. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions D.R. contributed to conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing of the original draft, and review and editing of the manuscript. F.T. contributed to supervision, conceptualization, methodology, validation, resources, and review and editing of the manuscript. R.M. contributed to methodology, validation, supervision, and review and editing of the manuscript.

Competing interests The authors declare no competing interests.