

# Brain-to-Text Decoding: A Non-invasive Approach via Typing

Jarod Lévy<sup>1</sup>, Mingfang (Lucy) Zhang<sup>2,3</sup>, Svetlana Pinet<sup>4,5</sup>, Jérémy Rapin<sup>1</sup>, Hubert Banville<sup>1</sup>, Stéphane d'Ascoli<sup>1\*</sup>, Jean-Rémi King<sup>1\*</sup>

<sup>1</sup>Meta AI, <sup>2</sup>École Normale Supérieure, Université PSL, CNRS, <sup>3</sup>Hospital Foundation Adolphe de Rothschild, <sup>4</sup>Basque Center on Cognition, Brain and Language, San Sebastian, <sup>5</sup>Ikerbasque, Basque Foundation for Science, Bilbao

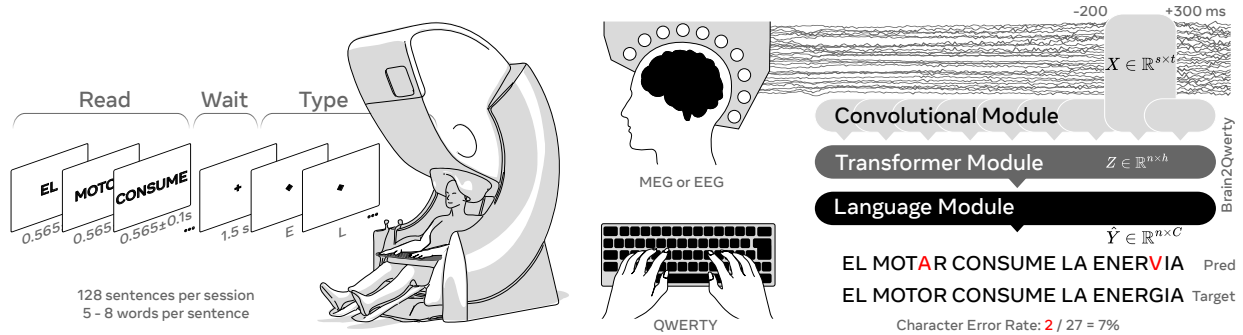
\*equal contribution

Modern neuroprostheses can now restore communication in patients who have lost the ability to speak or move. However, these invasive devices entail risks inherent to neurosurgery. Here, we introduce a non-invasive method to decode the production of sentences from brain activity and demonstrate its efficacy in a cohort of 35 healthy volunteers. For this, we present Brain2Qwerty, a new deep learning architecture trained to decode sentences from either electro- (EEG) or magneto-encephalography (MEG), while participants typed briefly memorized sentences on a QWERTY keyboard. With MEG, Brain2Qwerty reaches, on average, a character-error-rate (CER) of 32% and substantially outperforms EEG (CER: 67%). For the best participants, the model achieves a CER of 19%, and can perfectly decode a variety of sentences outside of the training set. While error analyses suggest that decoding depends on motor processes, the analysis of typographical errors suggests that it also involves higher-level cognitive factors. Overall, these results narrow the gap between invasive and non-invasive methods and thus open the path for developing safe brain-computer interfaces for non-communicating patients.

**Date:** February 5, 2025

**Correspondence:** {jarod,sdascoli, jeanremi}@meta.com

Meta



**Figure 1 Approach.** Recordings from 35 participants were obtained using electro-encephalography (EEG) and magneto-encephalography (MEG). Sentences were displayed word-by-word on a screen. Following the final word, a visual cue prompted them to begin typing this sentence, without visual feedback. Our Brain2Qwerty model includes three core stages to decode text from brain activity: (1) a convolutional module, input with 500 ms windows of M/EEG signals, (2) a transformer module trained at the sentence level, and (3) a pretrained language model to correct the outputs of the transformer. Performance is assessed using a Character Error Rate (CER) at the sentence level. An analysis of how the brain performs typing is described in a companion paper (Zhang et al., 2025).

## 1 Introduction

The past decade has been marked by rapid progress in brain-computer interfaces (BCIs) for individuals who, after a brain lesion, have lost their ability to speak or communicate. In particular, several patients suffering

from anarthria (Metzger et al., 2022; Moses et al., 2021), Amyotrophic Lateral Sclerosis (ALS) (Willett et al., 2023), or severe paralysis (Hochberg et al., 2012) have now been able to produce full sentences via a neuroprosthesis. Such an intracranial device, often implanted in the motor regions of the brain to record and decode its patterns of neural activity can assist with motor and communication tasks. Originally limited to decoding small sets of linguistic features (Herff et al., 2019; Angrick et al., 2019; Anumanchipalli et al., 2019; Moses et al., 2021; Card et al., 2024), words (Metzger et al., 2022), and gestures (Willett et al., 2021), the recent development of AI models has improved the precision and rapidity of brain-to-text decoding to a point of enabling natural language production at rates close to normal speech (Metzger et al., 2023; Wairagkar et al., 2024).

However, such invasive neuroprostheses require a neurosurgical procedure, and thus expose patients to non-negligible risks of brain hemorrhage and infection (Chung et al., 2019; Bullard et al., 2020; Leuthardt et al., 2021; Baranauskas, 2021). Additionally, maintaining functional cortical implants over extended time periods remains challenging (Fekete et al., 2023; Zhou et al., 2024; Yasar et al., 2024; Xie et al., 2020). As a result, in their current form, *invasive* BCIs are not easily scalable for diagnosing or restoring communication in the large groups of non- or poorly-responsive patients (Owen et al., 2006b; Claassen et al., 2019).

*Non-invasive* BCIs could potentially address this challenge. However, they are usually based on scalp electroencephalography (EEG), whose limited signal-to-noise ratio (Mak and Wolpaw, 2009) requires users to perform complex tasks. For example, EEG-based BCIs typically require individuals to maintain their attention on flickering stimuli (Abiri et al., 2019) or to imagine moving their hand or foot over long time periods (Bodien et al., 2024) – two tasks known to produce EEG patterns that can be relatively easily detected by a linear classifier. Even so, decoding performance remains moderate. For instance, a public BCI benchmark (Chevallier et al., 2024) using EEG achieves an accuracy of only 43.3% on a four-class classification task with a motor imagery dataset (Yi et al., 2014). In sum, current non-invasive methods fall short of providing a fast and reliable BCI.

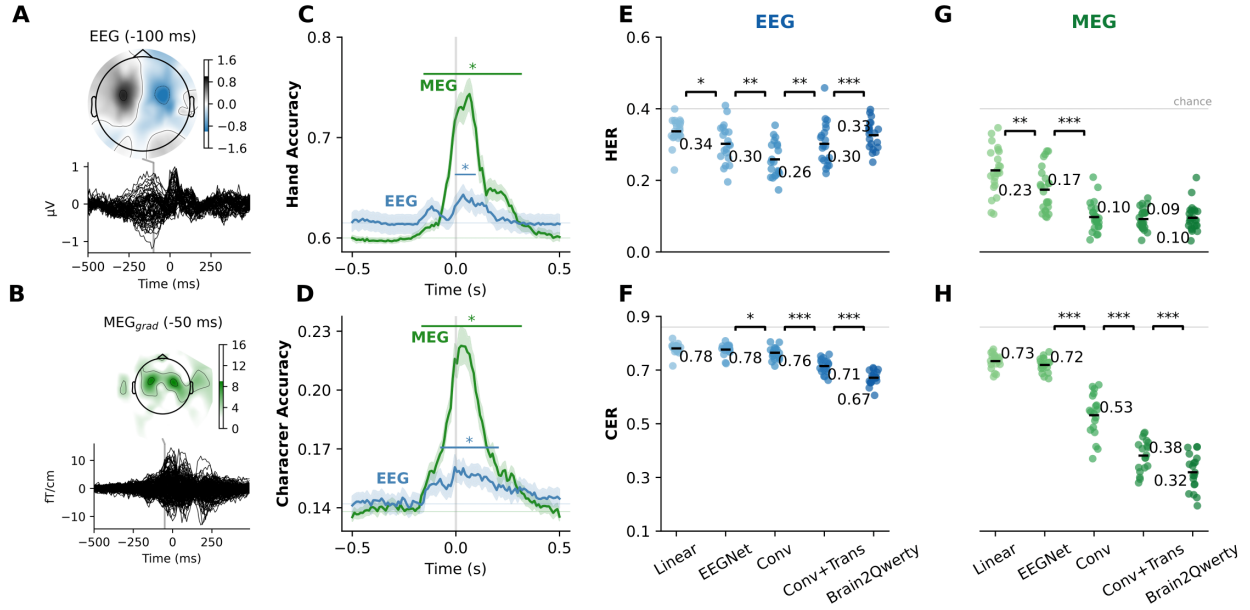
Two elements could address these challenges. First, magnetoencephalography (MEG), which measures the fluctuations of magnetic fields elicited in the cortex, has higher signal-to-noise ratio than EEG (Hämäläinen et al., 1993; Goldenholz et al., 2009; Baillet, 2017). Second, AI models trained to reconstruct natural language from MEG signals in language *comprehension* paradigms have recently demonstrated major improvements, especially in comparison to EEG (Défossez et al., 2023). Together, these elements thus indicate that with modern AI models, high-quality MEG signals and natural language tasks could be combined to decode the *production* of language from non-invasive recordings of the brain.

In this study, we introduce Brain2Qwerty, an AI model trained to decode text production from non-invasive recordings of brain activity (Fig. 1). For this, we tasked 35 participants to type briefly memorized sentences on a keyboard, while their brain activity was recorded with either EEG or MEG. We then train Brain2Qwerty, a three-stage deep neural network trained to decode text from these brain signals and evaluate it on both EEG (20 participants, 146K characters, 23K words and 4K sentences) and MEG recordings (20 participants, 193K characters, 30K words and 5K sentences). Note that the present study does *not* delve into *how* the brain produce language during typing. This neuroscientific issue is addressed in a companion paper (?).

## 2 Results

### 2.1 Linear Decoding

To verify that our typing protocol leads to the expected brain responses, we first focus on the differences in evoked responses elicited by left- and right-handed key presses (Fig. 2 A-B). The resulting topographies are typical of those associated with motor activity in the cortex (Donner et al., 2009). In addition, we trained a linear ridge classifier per subject to categorize left- vs right-handed responses at each time-sample relative to key presses. Decoding held-out peaks  $t=40$  ms after the key press (Fig. 2). MEG achieves a peak hand-classification accuracy of 74 ( $\pm 1.3$  standard-error-of-the-mean (SEM) across subjects) %, outperforming EEG by 16% (Mann Withney  $p: 10^{-7}$ ), which peaks at  $64 \pm 0.8\%$ . We then repeated this linear classification at the character-level. Character decoding peaks around the same time, and with a performance of  $22 \pm 0.8\%$  for MEG, and  $16 \pm 0.5\%$  for EEG. Overall, these findings confirm that the present protocol leads to the expected



**Figure 2 Decoding Performance across models.** **A.** Difference in EEG evoked responses between left vs right hand key presses. Each black line is the differential voltage of a sensor relative to key press. **B.** Same as A but for MEG. **C.** Linear classifiers are trained, at each time sample, to predict the left vs right hand relative to each key press. The gray line represents chance level and the error bar is the standard error of the mean across participants. Significant decoding scores ( $p < 0.05$ ) are marked with a star. **D.** Same as C but for character classification. **E-H.** Comparison of existing architectures (linear and EEGNet), and ablation of our three-step Brain2Qwerty model (Conv+Trans+Language Model), for both hand-error-rate (HER) and character-error-rate (CER). Each point represents the average score of a single participant. Statistical significance are denoted with  $p < 0.05$  (\*),  $p < 0.01$  (\*\*), and  $p < 0.001$  (\*\*\*).

brain responses to key presses (Pinet and Nozari, 2020).

## 2.2 Brain2Qwerty Performance

We next trained Brain2Qwerty, a new deep learning architecture, to decode individual characters from these M/EEG signals (cf Methods in section 4.2) and evaluated both the hands-error-rate (HER) and the character-error-rate (CER). The results show that Brain2Qwerty achieves a CER of  $32 \pm 0.6\%$  with MEG and  $67 \pm 1.5\%$  with EEG. This performance reflects a substantial difference across recording devices ( $p < 10^{-8}$ ). The worst and best EEG subjects reach a CER of  $61 \pm 2.0\%$  across sentences and  $71 \pm 2.3\%$ , respectively. Similarly, the best and worst MEG subjects reach a CER of  $19 \pm 1.1\%$  and  $45 \pm 1.2\%$ , respectively.

## 2.3 Comparing Brain2Qwerty to Baseline Models

How does Brain2Qwerty perform in comparison to classic baseline architectures? To address this issue we trained a linear model as well as EEGNet – a popular architecture used in BCIs (Lawhern et al., 2018) – with the same approach, and compared their decoding performance to Brain2Qwerty’s with a Wilcoxon test across subjects (Fig. 2C-F). EEGNet outperforms the linear model on both HER ( $p=0.008$ ) and CER ( $p < 10^{-4}$ ) for MEG – although only on HER for EEG ( $p=0.03$ ). However, EEGNet remains less effective than our model, which achieves, in comparison, a 1.14-fold improvement in CER with EEG ( $p < 10^{-5}$ ) and a 2.25-fold improvement for MEG ( $p < 10^{-6}$ ), respectively.

## 2.4 Brain2Qwerty Ablations

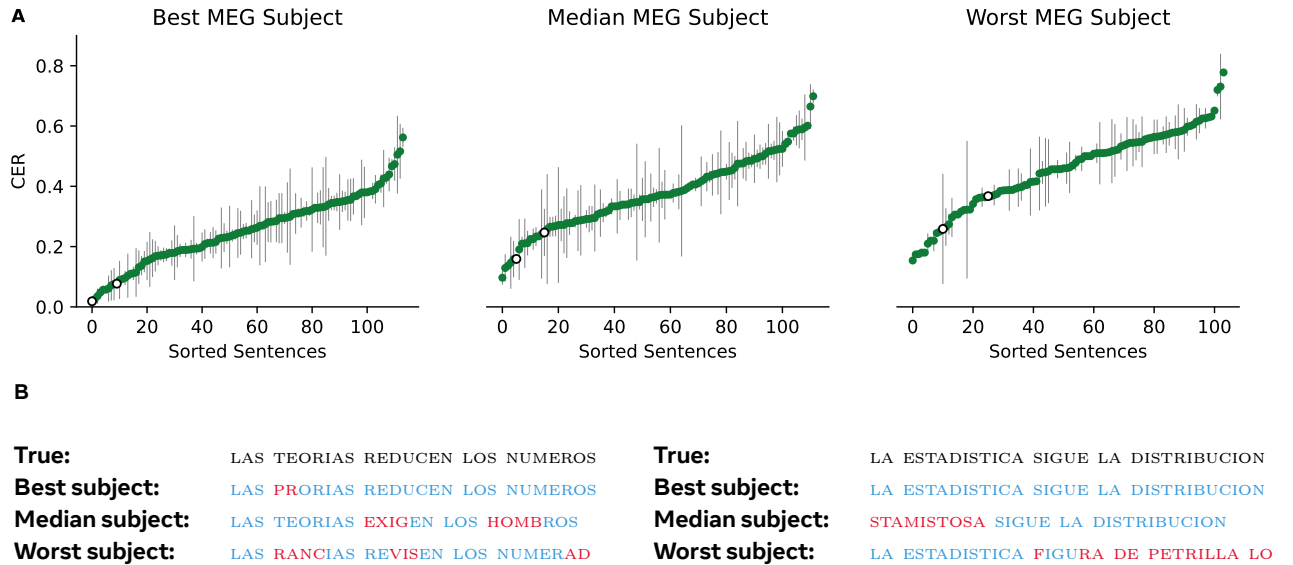
To validate our design choices, we then retrained different ablated versions of our model. Specifically, we re-trained and evaluated (i) the Convolutional Module (i.e. no transformer, no language model) and (ii) the Conv+Transformer (i.e. without Language Model) with the same hyperparameters. The Convolutional Module alone outperforms EEGNet both on the EEG’s HER ( $p=0.009$ ) and CER ( $p=0.03$ ) and on MEG

EEG		MEG	
<b>Read:</b>	LA CIENCIA DE LA IDEA ROMPE LA VISION	<b>Read:</b>	LA SILLA OCASIONA LAS LESIONES
<b>Typed:</b>	LA CIENCIA DE LA IDEA ROMPE LA <u>BISION</u>	<b>Type:</b>	LA SILLA OCASIOMA LAS LESIOMES
<b>Decode:</b>	LA CIENCIA DE LA IDEA LAS MAS DE ESOS	<b>Decode:</b>	LA SILLA OCASIONA LAS LESIONES
<b>Read:</b>	EL PROCESADOR EJECUTA LA INSTRUCCION	<b>Read:</b>	LAS TEORIAS REDUCEN LOS NUMEROS
<b>Typed:</b>	<u>ORDEN</u> ADOR EJECUTA LA INSTRUCCION	<b>Type:</b>	LAS TEORIAS REDUCEN LOS NUMEROS
<b>Decode:</b>	LAS CORRIDA PERITA LA INSTRUCCION	<b>Decode:</b>	LAS TEORIAS REDUCEN LOS NUMEROS
<b>Read:</b>	LA PRESENCIA DE LOS TIPOS IMPONE LOS RETOS	<b>Read:</b>	EL BENEFICIO SUPERA LOS RIESGOS
<b>Type:</b>	LA PRESENCIA DE LOS TIPOS IMPONE LOS RETOS	<b>Type:</b>	EK BENEFUCUI SYOERA KIS RUESGIS
<b>Decode:</b>	LA DECLARADA DE LOS CELOS ERAN A LOS ACTOS	<b>Decode:</b>	EL BENEFICIO SUPERA LOS RIESGOS

**Table 1** Examples of well-decoded sentences across subjects for EEG (left) and MEG (right) data.

Correct characters are highlighted in blue, mistakes in red, and typing errors underlined. Note that correct vs incorrect spaces are not visualized here.

data (HER:  $p < 10^{-5}$ , CER:  $p < 10^{-6}$ ). Adding the transformer only appears beneficial to CER with both EEG ( $p < 10^{-4}$ ) and MEG ( $p < 10^{-6}$ ). Finally, the use of a Language Model module leads to an additional 4% and 6% improvement of the CER of EEG ( $p < 10^{-5}$ ) and MEG ( $p < 10^{-6}$ ), respectively. Overall, these results show that the sentence-level contextualization provided by the transformer together with the leverage of natural language’s statistical regularities effectively improves the decoding of individual characters.



**Figure 3** Sentence-level performance for Best, Median and Worst MEG subjects.

**A.** Character-error-rate for three representative subjects. Each dot represents a unique sentence, with error bars indicating the standard error of the mean across repetitions. White dots corresponds to the sentences displayed below.

**B.** Decoding predictions for two sentences. Several splitting seeds were used to obtain the predictions across sentences.

## 2.5 Analyses of Decoded Sentences

The CER of all sentences for three representative participants recorded with MEG along with two example sentences from these subjects are displayed in Fig. 3. More decoding examples show that several sentences can be perfectly decoded for MEG (Tab. 1, right). Interestingly, some of these examples show that Brain2Qwerty’s language model can correct the typographical errors of the participant. For example, EL BENEFICIO SUPERA LOS RIESGOS was perfectly decoded, even though the participants typed: EK BENEFUCUI SYOERA KIS RUESGIS.

**Read:** LAS TEORIAS REDUCEN LOS NUMEROS  
**Typed:** LAS TEORIAS REDUCEN LOS NUMEROS  
**Conv:** LAS TANGPAS RE UDNDL SLINDIIS  
**Conv+Trans:** LAS GEOOLAS REDICEN LOS NUMEIOS  
**Brain2Qwerty:** LAS TEORIAS REDUCEN LOS NUMEROS

**Read:** EL BENEFICIO SUPERA LOS RIESGOS  
**Typed:** EL BEMEFICIO SUPERA LOS RIESFOS  
**Conv:** EL GEFEDISIO SUIERA NOA RIESTII  
**Conv+Trans:** EL GENEFICON CUPERA LOS RIESGOO  
**Brain2Qwerty:** EL BENEFICIO SUPERA LOS RIESGOS

**Table 2** Example of decoded sentences across ablations for MEG data. Color coding identical to Tab. 1.

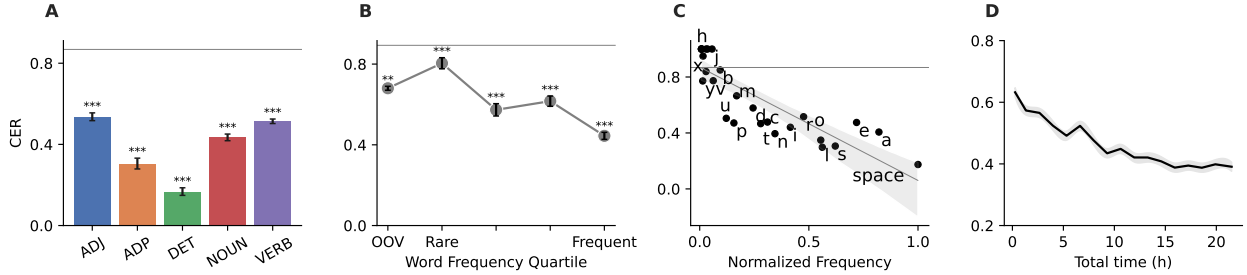
In comparison, the poor EEG decoding (Tab. 1, left) rarely leads to comprehensible text. Consistent with the statistical effects reported earlier, the examples in Tab. 2 highlight the impact of each module of our model, which together leads to perfect decoding after the language model.

## 2.6 Impact of Word Type and Frequency

To test whether Brain2Qwerty decodes words irrespectively of their grammatical type, we evaluated the CER for each part-of-speech (POS) categories separately (Fig. 4A). All POS categories are significantly decoded apart from determiners which exhibited a remarkably low CER ( $17 \pm 1.9\%$ ). This phenomenon may be due to two factors: their short length and their high frequency. To formally test this hypothesis, we first analyzed the impact of word frequency on CER (Fig. 4B). The results confirm that frequent words are better decoded than rare words ( $p=10^{-7}$ ). Interestingly, we verify that the words absent from the training set (out-of-vocabulary, OOV) can also be decoded, although with a relatively poor CER ( $68 \pm 2.1\%$ ). Note that this may be due to the fact that on a random partition of train/validation/test splits, OOV words tend to be rare words.

Second, we evaluated whether the frequency of each character also impacts decoding. The results show a significant correlation between character frequency and decoding accuracy:  $R=0.85$ ,  $p < 10^{-8}$  (Fig. 4C). Rare characters, such as "z," "k," and "w" in Spanish, are not decoded above chance level but only account for 0.08%, 0.08%, and 0.05% of the characters in our sentences. These results suggest that the number of repetitions (of words and characters) encountered during training directly affects performance.

To confirm this, we explore how decoding performance scales with the amount of data. For this, we re-trained our model on uniformly sampled subsets of the training set (Fig. 4D). Our results show that CER decreases as a function of the amount of training data:  $R=0.93$ ,  $p < 10^{-7}$ .

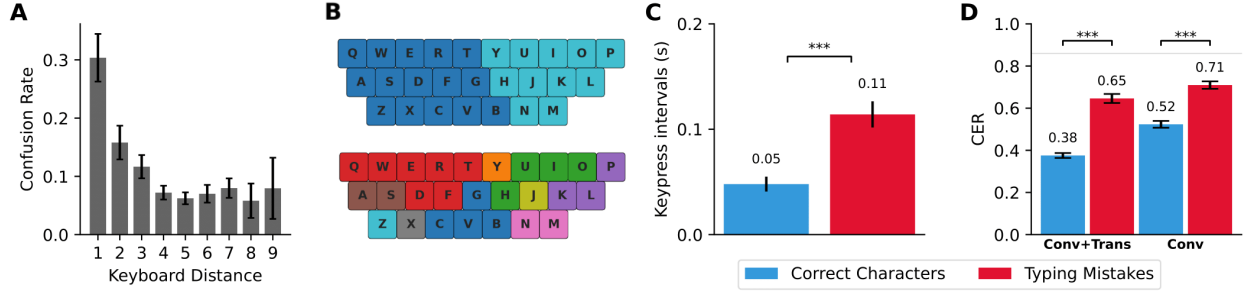


**Figure 4** Analysis of character- and word-level performance.

The results presented are specific to MEG data processed using the Conv+Trans model. **A.** Character-error-rate (CER) is evaluated across different part-of-speech categories to evaluate how performance varies across adjectives (ADJ), nouns, verbs, determiners (DET), and prepositions (ADP). **B.** CER as a function of word frequency. Out-of-vocabulary (OOV) decoding is used to test whether Brain2Qwerty can decode words absent from the training set. **C.** CER as a function of character frequency. **D.** CER as a function of recording time included in the training set.

## 2.7 Impact of Keyboard Layout

If Brain2Qwerty relies on brain activity from the motor cortex (as opposed to some amodal representations of language), then we expect its decoding errors to relate to the specific layout of the QWERTY keyboard. To



**Figure 5 Impact of keyboard layout and typing errors.**

The results presented are specific to MEG data processed using the Conv+Trans model. **A.** Keyboard Distance Effect. Confusion rate is analyzed against normalized keyboard distance. **B.** Clustering Analysis. K-means clustering of the model embeddings with 2 (top) and 10 (bottom) clusters respectively. **C.** Keypress Intervals Analysis. Comparison of keypress interval for correct keystrokes and typing errors, focusing on both preceding and subsequent characters. The sum of the two intervals is displayed. **D.** Typing Mistakes Performance Differences. Performance comparison for correct characters versus typing errors using the Conv+Trans model (left) and the Conv model (right).

test this hypothesis, we evaluated the confusion patterns of incorrectly predicted characters by analyzing the keyboard distance between decoded and actual key presses. The results show a strong Pearson correlation between physical distance and confusion rate:  $R=0.73$ ,  $p=0.02$  (Fig. 5A). To complement this analysis, we further perform a clustering analysis of the last-layer embeddings of the convolutional module using scikit-learn’s K-means clustering algorithm. When trained on two clusters, this unsupervised analysis fully separates the left-hand and right-hand keys. When using up to 10 clusters, the resulting partition remains consistent with the keyboard layout (Fig. 4B-E). This shows that the spatial layout of the keyboard is well encoded in the high-dimensional representations learned by our model. Overall, these results show that the decoder’s mistakes tend to be confused with the keys that are physically close to the target letter on the QWERTY keyboard, suggesting that the decoder primarily relies on motor representations.

## 2.8 Impact of Typing Errors

Our protocol does not allow participants to correct their mistakes. Typing errors account for 3.9% of the keystrokes and are present in 65% of the sentences. Furthermore, they are associated with a specific behavior (Fig. 5C): the time taken to type a character – as measured by the inter-key interval – doubles between correctly- ( $50 \pm 7$  ms) and incorrectly-typed characters ( $114 \pm 12$  ms,  $p=10^{-7}$ ). This well-known phenomenon (Logan and Crump, 2010) likely reflects hesitation or monitoring of mistakes. To evaluate the impact of typing error on decoding performance, we separately evaluated CER for correctly- and incorrectly-typed characters (Fig. 5D). The results show that with our Conv+Trans model, correctly-typed characters lead to a better CER (38%) than incorrectly-typed characters (65%,  $p=10^{-7}$ ). This result, however, may be partly driven by the contextualization enabled by the transformer. To minimize the impact of sentence context on this error analysis, we thus evaluate the performance of the Convolutional Module (Fig. 5D, right). Again, correctly-typed characters leads to a better CER (52%) than incorrectly-typed characters (71%,  $p=10^{-7}$ ). This result suggests that decoding performance diminishes when motor processes are inaccurately executed.

## 3 Discussion

The present study introduces a new method to decode the production of sentences from non-invasive brain recordings. With MEG, our Brain2Qwerty model achieves a character-error-rate (CER) of  $32 \pm 0.6\%$  on average across subjects, with the best-performing participants reaching a CER as low as 19%. Our analyses indicate that this decoding benefits from two main factors. First, the use of MEG signals instead of EEG signals resulted in a two-fold improvement. Second, our deep learning architecture, combined with a pretrained character-level language model, substantially outperforms standard models.

This work directly stems from the recent progress in decoding natural language from non-invasive recordings



of brain activity. In particular, Défossez et al. (2023) showed that the perception of natural speech segments could be decoded, from MEG signals, with up to 41% top-10 accuracy (chance level=0.1%). Similarly, Tang et al. (2023) showed that the meaning of perceived sentences could be decoded from functional Magnetic Resonance Imaging (fMRI). Our Brain2Qwerty model shares several elements with these approaches, notably through the use of both a subject layer (Défossez et al., 2023) and the use of a language model (Tang et al., 2023), although here restricted to a pretrained 9-gram character-level model. However, these two studies, which focus on decoding the *perception* of language rather than its *production*, remained limited in their downstream clinical applications.

Studies that directly decode text production from non-invasive recordings remain rare. For example, Crell et al. (2024) employed EEG to decode only 10 letters and achieved a character-error-rate of 75.8%, significantly higher than our 68% in the EEG setting with 29 characters. Similarly, EEG-based BCI benchmarks currently emphasize the constraints of poor signal quality and the variability across subjects (Chevallier et al., 2024). Our EEG findings are consistent with these observations. Beyond its mere metric performance, our approach is more efficient than traditional protocols used in non-invasive BCIs, such as the P300-speller (Marchetti and Priftis, 2014), SSVEP (Cheng et al., 2002), and fMRI-localizer (Owen et al., 2006a). These methods have historically relied on handcrafted signal processing and shallow classifiers. In contrast, our approach is based on a task that is comparatively easier to use.

While the decoding performance of Brain2Qwerty narrows the gap between non-invasive and invasive BCIs, this gap remains significant. In particular, for speech decoding, Metzger et al. (2023) achieved a rate of 79 words per minute and reported a CER of 15.2% on a dataset with 372 unique words, a vocabulary size comparable to ours. Willett et al. (2021) demonstrated typing speeds of 90 characters per minute with a CER below 6% and an offline CER under 1% when using a correction model. Both approaches rely on intracortical setups and require extensive recording sessions (11 hours per participant for Willett et al. (2021)). Consequently an important research avenue will be to scale and adapt those tasks for MEG experiments.

Several challenges remain to be solved before the present method could be adapted to clinical applications. First, our model does not operate in real time. In particular, the transformer and language model here operate at the sentence level and thus require the trial to conclude before an output can be produced. In addition, the input of Brain2Qwerty requires the MEG segments to be aligned to keystrokes. Overall, a real-time architecture, akin to what is done with electromyography (Sivakumar et al., 2024) and speech recognition (Défossez et al., 2023), remains necessary to make the present proof-of-concept applicable in real time.

Second, our study was conducted exclusively with healthy participants and with a strictly supervised model. Training indeed requires knowing both the timing and identity of each character. While this setup may be suitable for patients with neurodegenerative conditions who may still possess motor abilities, it is not applicable to locked-in individuals, who are completely unable to perform a typing task on a keyboard. Addressing this challenge may involve either adapting our typing task into an imagination task or designing AI systems capable of robust generalization across participants (Scotti et al., 2024).

Finally, while MEG outperforms EEG, current MEG systems, including the one used in the present study, are not wearable. This, however, may be resolved by the development of new MEG sensors based on optically pumped magnetometers (OPMs) (Shah and Wakai, 2013; Schofield et al., 2022; Brickwedde et al., 2024).

Overall, the present results serve as a stepping stone toward developing safer and more accessible non-invasive brain-computer interfaces, ultimately enabling solutions for individuals who have partially or completely lost the ability to communicate.

## 4 Methods

We aim to decode language production from non-invasive brain recordings. To achieve this, participants typed sentences on a keyboard while their brain activity was recorded using EEG or MEG. These two devices measure neural activity at a millisecond level, with EEG capturing electric fields and MEG detecting magnetic fields that are both generated by cortical neurons and recorded from sensors distributed across the scalp.

## 4.1 Experimental Protocol

*Cohort.* We recruited 35 healthy adult volunteers to participate in our study at the Basque Center on Cognition, Brain and Language (BCBL) in Spain. This group was composed of 23% of men and 77% of women with an average age of  $31.6 \pm 5.2$  years. All participants were right-handed and skilled at typing. Participants had to type words they were hearing on a keyboard covered by a cardboard box. They were selected if their typing accuracy was above or equal to 80%. They are all native Spanish speakers with no declared prior history of neurological or psychiatric disorders. Their brain activity was recorded with either EEG or MEG for  $0.88 \pm 0.02$  and  $0.93 \pm 0.01$  hours respectively, amounting to a total of 17.7 and 21.5 hours of typing. Five participants took part in both EEG and MEG sessions. One participant was excluded from the MEG study due to the presence of a metallic component during the recording. Participants gave their informed consent and were compensated 12 euros per hour for their participation. This study was approved by the local ethics committee. The same dataset is used to investigate the underlying neural mechanisms driving this task in a companion paper (?).

*Devices.* The MEG system is a Megin system with 306 channels (102 magneto-meters and 204 planar gradiometers) recording at a sampling rate of 1 kHz, with an online high-pass filter set at 0.1 Hz and a low-pass filter at 330 Hz. The EEG system is an actiCAP slim from BrainVision<sup>1</sup>, with 64 channels (61 EEG channels and 3 ocular channels), a BRAINAMP DC amplifier, and sampled at 1 kHz with an online high-pass filter set at 0.02 Hz.

Standard keyboards contain electronic and metallic parts that generate artifacts in the MEG. Consequently, we used a custom Magnetic-Resonance-compatible QWERTY keyboard from HybridMojo (LLC), and further modified it to replace the standard metallic springs with non-ferromagnetic silver-spring mechanisms.

*Task.* Participants were seated in front of a projected screen (100 cm for MEG and 70 cm for EEG away from their eyes), and with our custom keyboard placed on a stable platform. The distance between M/EEG sensors and the keyboard was 70 cm. This setup ensured participants could type in a natural position. Each trial consisted of three steps: read, wait, type. First, a sentence was presented on the screen, with a rapid serial visual presentation protocol (RSVP; i.e. one word at a time). Each word was presented in a black font, in all upper-case, on a 50% gray background for a random duration between 465 and 665 ms without intervals between words. Second, after the disappearance of the last word of each sentence, a black fixation cross was displayed on the screen for 1.5 seconds. Third, the disappearance of the fixation cross signaled the start of the typing phase. No letters were presented on the screen during typing. Nevertheless, we added minimal visual feedback: a small black square at the center of the screen rotated clockwise by 10 degrees on every keystroke. This feedback ensured that eye movements were not correlated with linguistic features, as it is usually the case in left-to-right reading. Each session consisted of two blocks of 64 sentences each. The first four sentences of each session were training sentences and were different from the 128 unique sentences in the protocol. During the first two training sentences, participants received visual feedback while typing. The other two sentences were used to train them on the task (typing with minimal visual feedback).

*Instruction and stimuli.* Participants were instructed to type the sentence that was presented in RSVP as accurately as possible without using backspaces to correct errors and while fixating on the center of the screen. To avoid using diacritic marks that occur in Spanish (e.g., é, á, í, ó, ú, ü, and ñ), all words were presented in upper case, and participants were instructed to think of writing in upper case, and without accents. Participants pressed the return key at the end of the trial. Each session consisted of 128 unique Spanish sentences. All sentences were declarative Spanish sentences that contained between 5 and 8 words. They consisted of determiners, nouns, adjectives, prepositions and verbs. This led to a total of 4K sentences and 146K characters across participants for EEG, and 5.1K sentences and 193K characters for MEG.

*MEG and EEG preprocessing.* In our pipeline, the EEG and MEG recordings were bandpass filtered between 0.1 and 20 Hz, and resampled to 50 Hz, using MNE-Python’s default parameters (Gramfort et al., 2014). These continuous recordings were then segmented into 0.5 s time windows around each key press from -0.2 s to +0.3 s. We applied baseline correction by subtracting the average channel-wise value in the (-0.2, 0) interval

---

<sup>1</sup><https://brainvision.com/products/acticap-slim-acticap-snap/>



from each window. To ensure that the data was on the same scale, a `RobustScaler` was applied across time from `scikit-learn` (Pedregosa et al., 2011) followed by a clamping operation. The scaler removes the median and scales the data according to the interquartile range.

*Text preprocessing.* We removed sentences that contained strictly more than 10 typographical errors (less than 5% for both EEG and MEG). To determine whether the participants performed typographical errors, we first used the `SequenceMatcher` from `difflib`<sup>2</sup> to align the typed sentence to the original sentence using a Gestalt pattern matching algorithm (Ratcliff and Metzner, 1988) that determines the minimal number of character-edits. This approach yields, for each keystroke event, two variables: the key pressed and the key that should have been pressed (target). Unless stated otherwise, all analyses are based on the target key. On average, participants typed  $152.0 \pm 3.2$  characters per minute, leading to an average sentence production time of  $5.7 \pm 0.2$  s.  $3.6 \pm 0.7\%$  of the keystrokes were typographical errors.

*Train/validation/test splits.* To ensure that our AI model does not memorize sentences, we divided the unique sentences into train (80%), validation (10%) and test splits (10%). As sentences may be similar (e.g. same sentence with an additional component), we adopted a maximally diverse splitting strategy by implementing a custom data splitter based on `sklearn`’s `AgglomerativeClustering` (Pedregosa et al., 2011). This method clusters unique sentences into groups of similar data points, where similarity is determined using Term Frequency-Inverse Document Frequency (TF-IDF) (Jones, 1972) cosine similarity with a threshold of 0.5 (i.e., sentences with a similarity score above 0.5 are grouped in the same split). These clusters are then assigned to the training, validation and testing splits with relative ratios of 80/10/10. This approach prevents the model from artificially improving test performance by memorizing similar sentences across participants. Throughout the study, we presented our results using a fixed splitting seed, except in Fig. 3 where we evaluated performance across sentences. For this figure, we aggregated results from multiple models with different splitting seeds to ensure broader sentence coverage, similar to a cross-validation approach.

## 4.2 Decoder

The goal for our decoding model is to predict each keystroke based on 0.5 s windows of M/EEG signals. Formally, the objective is to learn a mapping from brain recordings to class probabilities:  $f : \mathbb{R}^{s \times t} \rightarrow [0, 1]^C$ , where  $s$  represents the number of M/EEG sensors,  $t$  denotes the number of M/EEG time samples in the window, and  $C$  is the number of available keys. We consider  $C = 29$  distinct classes, which include all the letters of the Latin alphabet as well as three special classes: one for space, another for numbers, and the last for all other special characters.

### 4.2.1 Architecture

Our Brain2Qwerty model is composed of three successive modules (Fig. 1).

*Convolutional Module.* The first building block is a modified convolutional model originally introduced by Défossez et al. (2023). This model consists of four main parts. The first component employs a spatial attention mechanism to encode the relative positions of the sensors. The second introduces a subject-specific linear layer to account for differences between subjects. The third component is a convolutional neural network architecture, comprising 8 sequential blocks that employ a kernel size of 3 and a dilation period of 3, and incorporate skip connections, dropout regularization, and GELU activation functions. Finally, the temporal dimension is pooled with a single-head self-attention layer. Thus, for each window  $\mathbf{X} \in \mathbb{R}^{s \times t}$ , the CNN outputs  $\mathbf{z} \in \mathbb{R}^h$ , with  $h = 2,048$ . For clarity, this module will hereafter be referred to as Convolutional Module (Conv).

*Transformer Module.* The outputs of the Convolutional Module are then input to a transformer Module (Trans). The receptive field of the transformer is restricted to a unique sentence:  $\mathbf{Z} \in \mathbb{R}^{n \times h}$ . The transformer is used to refine the keystroke predictions by exploiting contextual information and consists of 4 layers with 2 attention heads per layer, maintaining consistent input and output dimensions. Finally, a linear layer projects the transformer’s outputs to obtain the logits of each character  $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times C}$ .

<sup>2</sup><https://docs.python.org/3/library/difflib.html>

*Language Model.* Finally, the output of the transformer  $\tilde{\mathbf{Y}}$  is input to a language model, so as to leverage the statistical regularities of natural language. For this, we used a 9-gram character-level model constructed using the KenLM library (Heafield, 2011) and pretrained on the Spanish Wikipedia Corpus (Wikimedia, 2005). This library optimizes both speed and memory efficiency by employing a prefix tree structure. At inference time, the language model is input with a sequence of predicted characters, and causally predicts the most likely next character given the preceding predicted ones.

Formally, let  $i \in [1, 2, \dots, n]$  denote the position of the character in a sequence of up to  $n$  characters and  $m = 9$  be the order of the  $m$ -gram model. The probability of the next character  $\hat{c}_i$  given its preceding predictions  $\hat{c}_{i-m}, \dots, \hat{c}_{i-1}$  is estimated as a weighted combination between the transformer’s logits and the probabilities of the language model:

$$\mathbf{P}(\hat{c}_i | \hat{c}_{i-m}, \dots, \hat{c}_{i-1}) = \mathbf{P}_{\text{trans}}(\hat{c}_i) + \alpha \cdot \mathbf{P}_{\text{lm}}(\hat{c}_i | \hat{c}_{i-m}, \dots, \hat{c}_{i-1}),$$

where  $\mathbf{P}_{\text{trans}}(\hat{c}_i) = \log \left( \text{softmax} \left[ (\tilde{\mathbf{Y}})_i \right] \right)$  represents the contribution from the core model,  $\mathbf{P}_{\text{lm}}(\hat{c}_i | \hat{c}_{i-m}, \dots, \hat{c}_{i-1})$  represents the probabilities from the language model, and  $\alpha$  is the language model weight. We use a beam search of size 30 and a language model weight of 5 found using a grid search, ensuring a tradeoff between inference time and decoding accuracy. This language modeling module outputs a sequence of  $n$  characters that aim to regularize the predictions of the transformer with the statistics of natural language. Brain2Qwerty’s final prediction is denoted as  $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times C}$ .

#### 4.2.2 Training

The Convolutional and Transformer modules are trained jointly with an unweighted cross-entropy loss in an end-to-end manner across all subjects, with the same hyperparameters for EEG and MEG recordings. This leads to a total of  $\sim 400\text{M}$  parameters (258M for Conv, 138M for Trans). The model is trained for 100 epochs with a batch size of 128 using the AdamW optimizer (Loshchilov and Hutter, 2019) with early stopping. We use the OneCycleLR scheduler (Smith and Topin, 2018) (weight decay= $10^{-4}$ ; pct\_start=0.1), warming up the learning rate to  $10^{-4}$  over the first 10 epochs then decaying linearly. Training was conducted on a single NVIDIA Tesla V100 Volta GPU with 32 GB of memory. The total runtime for training one model is  $\sim 12$  hours.

#### 4.2.3 Evaluation

*Hand Error Rate (HER).* For analysis purposes and comparison with the classic BCI literature (e.g. Yousefi et al. (2017); Hill et al. (2012)), we first consider HER. This metric estimates whether the target and the predicted characters correspond to the same left/right-hand split of the keyboard. Specifically, keys to the left of Y, H, and B are assigned to the left-hand category, while those to the right (including Y, H, and B) are assigned to the right-hand category. For this evaluation, special characters, numbers and space are excluded, as participants may use both hands to type them.

*Character-error-rate (CER).* CER is based on the Levenshtein distance which quantifies the minimum number of single-character edits required to transform the predicted sequence of keystrokes to the target sentence. A CER of 0 indicates perfect character-level accuracy. The formula for CER is given by  $\text{CER} = (s + d + a) \times \frac{1}{n}$ , where  $s$ ,  $d$  and  $a$  stand for substitutions, deletions, and additions in the  $n$ -character sentence, respectively. Unless stated otherwise, we compute the CER at the sentence level using the `Levenshtein`<sup>3</sup> python library, and report the average CER across sentences.

#### 4.2.4 Model comparison

*Statistics.* For statistical comparisons, we employed the non-parametric tests provided by the `scipy` package (Virtanen et al., 2020). For comparison across models within the same subjects, we used the Wilcoxon test. For comparison across subjects (e.g. for EEG versus MEG), we used the Mann Whitney U test. For

<sup>3</sup><https://pypi.org/project/Levenshtein/>

time-course decoding, we further applied a false-discovery-rate (FDR) correction for multiple comparisons across time samples.

*Baseline Models.* We consider two baseline models. The first one is a linear model implemented using the `RidgeClassifierCV` function from the scikit-learn library (Pedregosa et al., 2011), which was trained to predict characters from a single time sample of recording for each subject separately. The regularization parameter  $\alpha$  was selected through a nested cross-validation using a grid search logarithmic spanning from  $10^{-2}$  to  $10^8$ . This operation was repeated for each time sample between -0.5 and 0.5 seconds relative to the character onset.

As a second baseline, we need a model which uses the same setup as Brain2Qwerty, i.e. where all subjects are trained collectively and the temporal dimension is collapsed. We used EEGNet (Lawhern et al., 2018), a highly parameter-efficient model classically used in BCIs. We trained EEGNet with the same approach as our model. For our experiments, EEGNet is configured with a depth of 6 and a dropout rate of 0.3, which were selected based on a grid search over the validation set. EEGNet does not incorporate a subject-specific linear layer, which may compromise its performance in this particular setup where all subjects were trained collectively. To compute the chance level while accounting for the imbalance across characters, we evaluated the performance of a dummy model that always predicts the most frequent character.

*Companion paper.* We explore *how* the brain produce a hierarchy of language representations in a companion paper (Zhang et al., 2025). Note that Fig. 1 (left) and Fig. 2B are shared between these two studies.

## 5 Acknowledgments

The authors would like to thank Maite Kaltzakorta, Manex Lete, Jessi Jacobsen, Daniel Nieto, Jone Iraeta, Araitz Garnika, Jaione Bengoetxea, Natalia Louleli, Naroa Miralles, Eñaut Zeberio, Craig Richter, Amets Esnal, and Olatz Andonegui, as well as Abhishek Charnalia and Pierre-Louis Xech for their critical help. This research is supported by the Basque Government through the BERC 2022-2025 program and Funded by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010/AEI/10.13039/501100011033. Parts of this research were carried within the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 945304 - Cofund AI4theSciences hosted by PSL University.

## References

- R Abiri, S Borhani, E. W Sellers, Y Jiang, and X Zhao. A comprehensive review of eeg-based brain-computer interface paradigms. *Journal of Neural Engineering*, 16:011001, 2019.
- M Angrick, C Herff, E Mugler, M. C Tate, M. W Slutzky, D. J Krusienski, and T Schultz. Speech synthesis from ecog using densely connected 3d convolutional neural networks. *Journal of Neural Engineering*, 16(3):036019, 2019. doi: 10.1088/1741-2552/ab0b77.
- G. K Anumanchipalli, J Chartier, and E. F Chang. Speech synthesis from neural decoding of spoken sentences. *Nature*, 568(7753):493–498, 2019. doi: 10.1038/s41586-019-1119-1.
- Sylvain Baillet. Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, 20:327–339, 2017. doi: 10.1038/nn.4504. <https://doi.org/10.1038/nn.4504>.
- Gintaras Baranauskas. What limits the performance of current invasive bcis? *Frontiers in Systems Neuroscience*, 15: 617326, 2021. doi: 10.3389/fnsys.2021.617326. <https://doi.org/10.3389/fnsys.2021.617326>.
- Y. G Bodien, J. T Giacino, B. L Edlow, and N. D Schiff. Cognitive motor dissociation in disorders of consciousness. *New England Journal of Medicine*, 391(4):598–608, 2024. doi: 10.1056/NEJMoa2401376.
- Marion Brickwedde, Paul Anders, Andrea A Kühn, Roxanne Lofredi, Martin Holtkamp, Angela M Kaindl, Tineke Grent-’t Jong, Peter Krüger, Tilmann Sander, and Peter J Uhlhaas. Applications of opm-meg for translational neuroscience: a perspective. *Translational Psychiatry*, 14(1):341, 2024.
- Allison J. Bullard, Brent C. Hutchison, Jeffrey Lee, Cynthia A. Chestek, and Parag G. Patil. Estimating risk for future intracranial, fully implanted, modular neuroprosthetic systems: A systematic review of hardware complications in clinical deep brain stimulation and experimental human intracortical arrays. *Neuromodulation*, 23(4):411–426, 2020. doi: 10.1111/ner.13082. <https://doi.org/10.1111/ner.13082>.
- N. S Card, S. L Metzger, D. A Moses, F. R Willett, and E. F Chang. An accurate and rapidly calibrating speech neuroprosthesis. *New England Journal of Medicine*, 391(4):609–618, 2024. doi: 10.1056/NEJMoa2401435.
- Ming Cheng, Xiaorong Gao, Shangkai Gao, and Dingfeng Xu. Design and implementation of a brain-computer interface with high transfer rates. *IEEE Transactions on Biomedical Engineering*, 49(10):1181–1186, 2002.
- Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sebastien Velut, Salim Khazem, and Thomas Moreau. The largest eeg-based bci reproducibility study for open science: the moabb benchmark, 2024. <https://arxiv.org/abs/2404.15319>.
- Jeong E. Chung, Hyerim R. Joo, Jennifer L. Fan, Daniel F. Liu, Andrew H. Barnett, Samantha Chen, Connor Geaghan-Breiner, Mathias P. Karlsson, Matias Karlsson, Kimberly Y. Lee, Hui Liang, Jeremy F. Magland, Jeffrey A. Pebbles, Andrew C. Tooker, Leslie F. Greengard, Vanessa M. Tolosa, and Loren M. Frank. High-density, long-lasting, and multi-region electrophysiological recordings using polymer electrode arrays. *Neuron*, 101(1):21–31.e5, 2019. doi: 10.1016/j.neuron.2018.11.002. <https://doi.org/10.1016/j.neuron.2018.11.002>. Epub 2018 Nov 27.
- Jan Claassen, Kevin Doyle, Adu Matory, Caroline Couch, Kelly M. Burger, Angela Velazquez, Joshua U. Okonkwo, Jean-Rémi King, Soojin Park, Sachin Agarwal, David Roh, Murad Megjhani, Andrey Elisseyev, E. Sander Connolly, and Benjamin Rohaut. Detection of brain activation in unresponsive patients with acute brain injury. *New England Journal of Medicine*, 380(26):2497–2505, 2019. doi: 10.1056/NEJMoa1812757. <https://doi.org/10.1056/NEJMoa1812757>.
- M Crell et al. Non-invasive decoding of hand kinematics for letter classification using eeg. *Journal of Neural Engineering*, 2024.
- T. H. Donner, M. Siegel, P. Fries, and A. K. Engel. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, 19(18):1581–1585, Sep 2009. doi: 10.1016/j.cub.2009.07.066. Epub 2009 Sep 10.
- Alexandre Défossez, Charlie Caucheteux, Jeremy Rapin, et al. Decoding speech perception from non-invasive brain recordings. *Nature Machine Intelligence*, 5(10):1097–1107, 2023. doi: 10.1038/s42256-023-00714-5.
- Z. Fekete, A. Zátanyi, A. Kaszás, et al. Transparent neural interfaces: challenges and solutions of microengineered multimodal implants designed to measure intact neuronal populations using high-resolution electrophysiology and microscopy simultaneously. *Microsystems & Nanoengineering*, 9:66, 2023. doi: 10.1038/s41378-023-00519-x. <https://doi.org/10.1038/s41378-023-00519-x>.

- Daniel M. Goldenholz, Seppo P. Ahlfors, Matti S. Hämäläinen, David Sharon, Miki Ishitobi, Lucia M. Vaina, and Steven M. Stufflebeam. Mapping the signal-to-noise ratios of cortical sources in magnetoencephalography and electroencephalography. *Human Brain Mapping*, 30(4):1077–1086, 2009. doi: 10.1002/hbm.20571. <https://doi.org/10.1002/hbm.20571>.
- Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *neuroimage*, 86:446–460, 2014.
- Kenneth Heafield. Kenlm: faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, WMT ’11, pages 187–197, USA, 2011. Association for Computational Linguistics. ISBN 9781937284121.
- C Herff, M Angrick, G Johnson, J. J Shih, D. J Krusienski, and T Schultz. Generating natural, intelligible speech from brain activity in motor, premotor, and inferior frontal cortices. *Frontiers in Neuroscience*, 13:1267, 2019. doi: 10.3389/fnins.2019.01267.
- Jeremy Hill, Thomas N Lal, Michael Schröder, Tanya Hinterberger, Guido Widman, Christian E Elger, Stephen J Navin, and Niels Birbaumer. Decoding hand from human brain activity: A review of the bci literature. *NeuroImage*, 59(3):2614–2625, 2012.
- Leigh Hochberg, Daniel Bacher, Beata Jarosiewicz, et al. Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. *Nature*, 485:372–375, 2012. doi: 10.1038/nature11076. <https://doi.org/10.1038/nature11076>.
- Matti S. Hämäläinen, Riitta Hari, Risto J. Ilmoniemi, Jukka Knuutila, and Outi V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993. doi: 10.1103/RevModPhys.65.413. <https://doi.org/10.1103/RevModPhys.65.413>.
- Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28(1):11–21, 1972.
- Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.
- Eric C. Leuthardt, Daniel W. Moran, and Thomas R. Mullen. Defining surgical terminology and risk for brain-computer interface technologies. *Frontiers in Neuroscience*, 15:599549, 2021. doi: 10.3389/fnins.2021.599549. <https://doi.org/10.3389/fnins.2021.599549>.
- Gordon D. Logan and Matthew J. C. Crump. Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science*, 330(6004):683–686, 2010. doi: 10.1126/science.1190483.
- Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019. <https://arxiv.org/abs/1711.05101>.
- Joseph N. Mak and Jonathan R. Wolpaw. Clinical applications of brain-computer interfaces: current state and future prospects. *IEEE Reviews in Biomedical Engineering*, 2:187–199, 2009. doi: 10.1109/RBME.2009.2035356. <https://doi.org/10.1109/RBME.2009.2035356>.
- M Marchetti and K Priftis. Effectiveness of the p3-speller in brain–computer interfaces for amyotrophic lateral sclerosis patients: a systematic review and meta-analysis. *Frontiers in Neuroengineering*, 7:86422, 2014.
- S. L Metzger, F. R Willett, D. A Moses, and E. F Chang. A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 620(7970):1037–1046, 2023. doi: 10.1038/s41586-023-06199-4.
- Sarah L. Metzger, Jessie R. Liu, David A. Moses, et al. Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis. *Nature Communications*, 13:6510, 2022. doi: 10.1038/s41467-022-33611-3. <https://doi.org/10.1038/s41467-022-33611-3>.
- D. A Moses, S. L Metzger, J. R Liu, G. K Anumanchipalli, J. G Makin, and E. F Chang. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*, 385(3):217–227, 2021. doi: 10.1056/NEJMoa2026843.
- A. M Owen, M. R Coleman, M Boly, M. H Davis, S Laureys, and J. D Pickard. Detecting awareness in the vegetative state. *Science*, 313(5792):1402–1402, 2006a. doi: 10.1126/science.1130197.



- Adrian M. Owen, Martin R. Coleman, Mélanie Boly, Matthew H. Davis, Steven Laureys, and John D. Pickard. Detecting awareness in the vegetative state. *Science*, 313(5792):1402, 2006b. doi: 10.1126/science.1130197. <https://doi.org/10.1126/science.1130197>.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.
- Svetlana Pinet and Nazbanou Nozari. Electrophysiological correlates of monitoring in typing with and without visual feedback. *Journal of Cognitive Neuroscience*, 32(4):603–620, April 2020. doi: 10.1162/jocn\_a\_01500. [https://doi.org/10.1162/jocn\\_a\\_01500](https://doi.org/10.1162/jocn_a_01500). Epub 2019 Nov 8.
- John W. Ratcliff and David E. Metzener. Pattern matching: The gestalt approach. *Dr. Dobb’s Journal*, page 46, July 1988. <https://www.drdoobbs.com/database/pattern-matching-the-gestalt-approach/184407970?pgno=5>.
- Holly Schofield, Elena Boto, Vishal Shah, Ryan M Hill, James Osborne, Molly Rea, Cody Doyle, Niall Holmes, Richard Bowtell, David Woolger, et al. Quantum enabled functional neuroimaging: the why and how of magnetoencephalography using optically pumped magnetometers. *Contemporary physics*, 63(3):161–179, 2022.
- Paul S. Scotti, Mihir Tripathy, Cesar Kadir Torrico Villanueva, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A. Norman, and Tanishq Mathew Abraham. Mindeye2: Shared-subject models enable fmri-to-image with 1 hour of data, 2024. <https://arxiv.org/abs/2403.11207>.
- Vishal K Shah and Ronald T Wakai. A compact, high performance atomic magnetometer for biomedical applications. *Physics in Medicine & Biology*, 58(22):8153, 2013.
- Viswanath Sivakumar, Jeffrey Seely, Alan Du, Sean R Bittner, Adam Berenzweig, Anuoluwapo Bolarinwa, Alexandre Gramfort, and Michael I Mandel. emg2qwerty: A large dataset with baselines for touch typing using surface electromyography. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2024. <https://openreview.net/forum?id=LFCWIE5iS2>.
- Leslie N. Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates, 2018. <https://arxiv.org/abs/1708.07120>.
- J Tang, A LeBel, S Jain, and A. G Huth. Semantic reconstruction of continuous language from non-invasive brain recordings. *Nature Neuroscience*, 26(5):858–866, 2023.
- Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.
- M Wairagkar, S. L Metzger, D. A Moses, F. R Willett, and E. F Chang. An instantaneous voice synthesis neuroprosthesis. *bioRxiv*, 2024. doi: 10.1101/2024.08.25.712353.
- Wikimedia. Wikimedia downloads, 2005. <https://dumps.wikimedia.org>.
- F. R Willett, D. T Avansino, L. R Hochberg, J. M Henderson, and K. V Shenoy. High-performance brain-to-text communication via handwriting. *Nature*, 593(7858):249–254, 2021. doi: 10.1038/s41586-021-03506-2.
- F. R Willett, D. A Moses, S. L Metzger, J. R Liu, and E. F Chang. A high-performance speech neuroprosthesis. *Nature*, 620(7970):1031–1036, 2023. doi: 10.1038/s41586-023-06200-2.
- Yue Xie, Zhanfeng Zhang, and Xing Zhang. Chronic neural interface technology: new advances in flexible and durable materials. *Advanced Healthcare Materials*, 9(21):2000978, 2020. doi: 10.1002/adhm.202000978. <https://doi.org/10.1002/adhm.202000978>.
- T.B. Yasar, P. Gombkoto, A.L. Vyssotski, et al. Months-long tracking of neuronal ensembles spanning multiple brain areas with ultra-flexible tentacle electrodes. *Nature Communications*, 15:4822, 2024. doi: 10.1038/s41467-024-49226-9. <https://doi.org/10.1038/s41467-024-49226-9>.
- W. Yi, S. Qiu, K. Wang, H. Qi, L. Zhang, P. Zhou, F. He, and D. Ming. Evaluation of eeg oscillatory patterns and cognitive process during simple and compound limb motor imagery. *PloS One*, 9(12):e114853, 2014. doi: 10.1371/journal.pone.0114853.
- Siamak Yousefi, Ali Hashemi, and Sarah Ostadabbas. Hand classification from eeg signals using convolutional neural networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1736–1745, 2017.

- Mingfang (Lucy) Zhang, Jarod Lévy, Stéphane d’Ascoli, Jérémy Rapin, F.-Xavier Alario, Pierre Bourdillon, Svetlana Pinet, and Jean-Rémi King. From thought to action: How the hierarchy of neural dynamics supports language production. 2025.
- C. Zhou, Y. Tian, G. Li, et al. Through-polymer, via technology-enabled, flexible, lightweight, and integrated devices for implantable neural probes. *Microsystems & Nanoengineering*, 10:54, 2024. doi: 10.1038/s41378-024-00691-8. <https://doi.org/10.1038/s41378-024-00691-8>.