



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

☀ 상태	완료
≡ Journal	arXiv(Meta)
≡ Summary	전기뇌파검사(EEG) 또는 자기뇌파검사(MEG)에서 문장을 디코딩하도록 훈련된 새로운 심층 학습 구조인 Brain2Qwerty 소개
≡ Limitations	실시간 처리 가능성, 휴대용 MEG 장비 개발, 환자를 위한 적응형 학습 모델 등이 연구되어야 함
🔗 Link	https://scontent-ssn1-1.xx.fbcdn.net/v/t39.2365-6/475464888_600710912891423_9108680259802499048_n.pdf?_nc_cat=102&ccb=1-7&_nc_sid=3c67a6&_nc_ohc=VNE4iRNCZXYQ7kNvgGlllXU&_nc_oc=AdIWUBOD2RcfhdmsM_Jl2J8-7Msl4GQ6VNxZT6aMoNq4dNy3JQ6TOVZJydBpuAillo&_nc_zt=14&_nc_ht=scontent-ssn1-1.xx&_nc_gid=M9B-3mvVdDbPWVe7SQW8A&oh=00_AYHkuQzA5TWu1XOFa21-ZjwcD8W2yeYRbAYXrqWi51q4fQ&oe=67E21A16
≡ category	EEG Text Generation
≡ Year	2025.02

GitHub - facebookresearch/emg2qwerty: A surface electromyographic (sEMG) touch typing dataset with baselines. <https://arxiv.org/abs/2410.20081>.
A surface electromyographic (sEMG) touch typing dataset with baselines. <https://arxiv.org/abs/2410.20081>. - facebookresearch/emg2qwerty
🔗 <https://github.com/facebookresearch/emg2qwerty?tab=readme-ov-file>

Introduction

- 침습적 BCI : 신경외과적 절차를 요구하므로, 뇌 출혈 및 감염과 같은 위험 초래 → 비침습적 BCI 필요
- 비침습적 BCI
 - EEG의 경우, SNR로 인해 복잡한 작업을 수행해야 하고 해독 성능 보통임 (motor imagery 43.3%)
 - MEG의 경우, brain-to-text 태스크에서 EEG보다 높은 성능을 보임
- **Brain2Qwerty**라는 비침습적 뇌-텍스트 변환 시스템을 개발 → EEG 및 MEG 데이터를 사용해 문장을 예측하는 방법을 제안
 - 35명의 참여자에게 키보드에서 간단히 기억한 문장 입력 요청 → **EEG 및 MEG로 기록**
 - Brain2Qwerty 훈련 후, EEG(20명, 146K 문자, 23K 단어 및 4K 문장) 및 MEG(20명, 193K 문자, 30K 단어 및 5K 문장) 기록 모두에서 평가
- ⚠ 현재 연구는 타이핑 중 뇌가 언어를 생성하는 방식에 대해 다루지 않았는데, 이는 동반 논문인 [From Thought to Action(25.02)] 에서 다룸

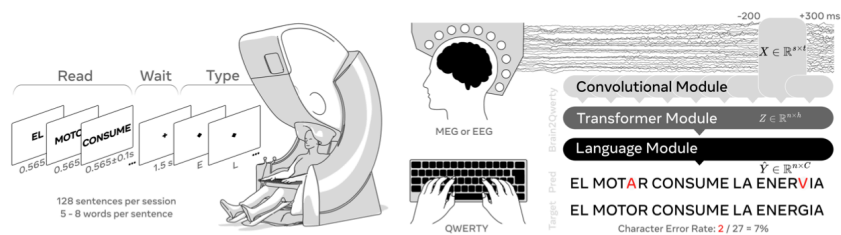


Figure 1 Approach. Recordings from 35 participants were obtained using electro-encephalography (EEG) and magneto-encephalography (MEG). Sentences were displayed word-by-word on a screen. Following the final word, a visual cue prompted them to begin typing this sentence, without visual feedback. Our Brain2Qwerty 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).

Methods

Experimental Protocol

- Cohort
 - 35명의 건강한 성인 자원자를 모집
 - EEG와 MEG로 0.88±0.02시간, 0.93±0.01시간 동안 기록 → 총 17.7시간과 21.5시간의 타이핑 시간

- Devices
 - EEG는 BrainVision의 actiCAP slim으로, 64채널임
 - MEG는 306채널,, 키보드도 맞춤형 QWERTY 키보드,,
- Task
 - read ; 한 번에 한 단어씩 화면에 문장 표시
 - wait ; 각 문장의 마지막 단어가 사라진 후, 1.5초 동안 검정색 화면
 - type ; 화면이 사라지면서 타이핑이 시작되고, 타이핑 동안 화면에 어떤 문자도 표시되지 않음
- Instruction and stimuli
 - 각 세션은 128개의 고유한 스페인어 문장으로 구성됨
 - EEG를 위한 146K 문자, 4K 문장 / MEG를 위한 193K 문자, 5K 문장
- MEG and EEG preprocessing
 - 0.1Hz ~ 20Hz 에서 밴드패스 필터링
 - MNE-Python의 기본 매개변수를 사용하여 50Hz로 재샘플링
 - 각 키 프레스 주변의 **0.5초 타임 윈도우**로 분할 (-0.2초부터 +0.3초까지) → 기저선 보정 진행
 - 데이터가 동일한 척도로 유지되도록 scikit-learn의 RobustScaler를 적용한 후 클램핑 작업 수행
- Text preprocessing
 - difflib의 SequenceMatcher를 사용하여 입력된 문장을 원본 문장과 정렬
 - 눌린 키와 눌려야 하는 키에 대해 매칭을 통해 오류 찾아냄 → 오류 10개 초과 문장 삭제
- Train/validation/test splits
 - train 80 / validation 10 / test 10
 - sklearn의 AgglomerativeClustering 기반 맞춤형 데이터 분할기 구현

Decoder



[목표] 뇌 기록에서 **클래스 확률에 대한 매핑** 학습 : $f : \mathbb{R}^{s \times t} \rightarrow [0, 1]^C$

[내용] M/EEG 신호의 0.5초 window 기반 각 키 입력 예측

[설명] C = 29개의 서로 다른 클래스 고려

- Architecture
 - **Convolutional Module (CNN 기반 특징 추출기)**
 - 센서의 상대적 위치를 인코딩하기 위해 spatial attention mechanism 블록 구성
 - 주체별 차이 고려를 위해 subject-specific linear layer 도입
 - 8개의 sequential 블록으로 구성된 convolutional neural network 아키텍처 구성
 - skip connections, dropout regularization, GELU activation functions ...
 - 시간 차원이 single-head self-attention layer로 풀링됨
 - 즉, 각 윈도우 $X \in \mathbb{R}^{s \times t}$ 에 대해 CNN은 $z \in \mathbb{R}^h$ 를 출력하며, h = 2048 임
 - **Transformer Module (문맥적 분석)**
 - Conv 모듈의 출력이 Trans 모듈의 입력으로 들어옴
 - **Transformer는 문맥 정보를 활용하여 키스트로크 예측을 세밀하게 조정되는데 사용**
 - 각 층에 2개의 attention heads를 가진 4개의 layers로 구성
 - linear layer가 transformer의 출력을 투영하여 각 문자에 대한 logits($\tilde{Y} \in \mathbb{R}^{n \times C}$)을 얻음
 - **Language Model (언어 모델 활용)**
 - Trans의 출력이 language model의 입력으로 들어옴
 - 자연어의 통계적 규칙성 활용하여 최종 예측 정교화
 - $P(\hat{c}_i | \hat{c}_{i-m}, \dots, \hat{c}_{i-1}) = P_{\text{trans}}(\hat{c}_i) + \alpha \cdot P_{\text{lm}}(\hat{c}_i | \hat{c}_{i-m}, \dots, \hat{c}_{i-1})$
 - 즉, **trans의 예측을 자연어의 통계로 정규화하려는 n개의 문자 시퀀스를 출력함**
 - 최종 예측은 다음과 같이 나타냄 : $\hat{Y} \in \mathbb{R}^{n \times C}$

- Training
 - 모델은 약 400M개의 파라미터를 가지고 있으며,
 - 32GB NVIDIA Tesla V100 GPU에서 약 12시간 동안 훈련
 - 학습률 스케줄링(OneCycleLR)과 AdamW 옵티마이저를 사용하여 최적화 수행
- Evaluation
 - Hand Error Rate(HER): 목표 문자와 예측된 문자가 키보드의 왼쪽/오른쪽 손 분할에 해당하는지 추정
 - Character Error Rate(CER): 예측된 키스트로크 시퀀스를 목표 문장으로 변환하는 데 필요한 수 측정
 - 0이면 완벽한 문자 수준 정확도
- Model comparison
 - Baseline 모델(선형 모델, EEGNet)과 성능 비교

Results

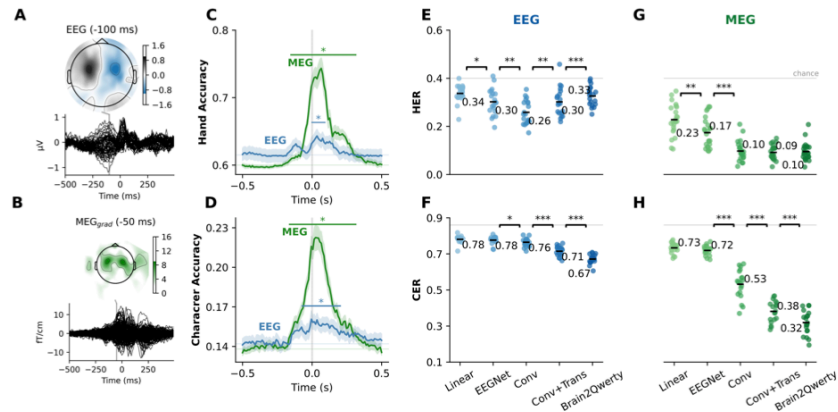


Figure 2 Decoding Performance across models. **A.** Difference in EEG evoked responses between left vs right hand key presses. Each black line is the differential voltage of a sensor relative to key press. **B.** Same as A but for MEG. **C.** Linear classifiers are trained, at each time sample, to predict 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 baselines (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 is denoted with $p < 0.05$ (*), $p < 0.01$ (**), and $p < 0.001$ (***).

Linear Decoding

- 실험 참가자들이 키보드를 사용할 때 EEG와 MEG 데이터를 기록하여 좌우 손 사용에 따른 뇌 신호 차이를 분석
- MEG의 분류 정확도가 EEG보다 높았으며, 특히 키 입력 후 40ms 시점에서 가장 높은 분류 성능을 보임

Brain2Qwerty Performance

- Brain2Qwerty 모델이 EEG와 MEG 데이터를 활용해 문장을 얼마나 정확하게 예측하는지 평가
- MEG 기반 예측의 문자 오류율(Character Error Rate, CER)은 32%였으며, EEG 기반 예측은 67%로 비교적 성능이 낮음
- 가장 성능이 좋은 참가자의 경우 MEG에서 CER 19%까지 낮출 수 있었음

Comparing Brain2Qwerty to Baseline Models

- Brain2Qwerty 모델을 기존 EEGNet 모델 및 선형 모델과 비교
- EEGNet이 선형 모델보다 성능이 좋았지만, Brain2Qwerty가 가장 우수한 성능을 보임
- MEG 데이터를 사용할 때 성능 차이가 더욱 보임

Brain2Qwerty Ablations

Read: LAS TEORIAS REDUCEN LOS NUMEROS
Typed: LAS TEORIAS REDUCEN LOS NUMEROS
Conv: LAS TANGPAS RE UDNNDL 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.

- Brain2Qwerty는 3단계 구조(CNN, Transformer, Language Model)로 구성됨
- 개별 요소를 제거하는 실험을 수행한 결과, Transformer와 Language Model을 추가할수록 성능이 향상됨을 확인

Analyses of Decoded Sentences

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

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

- MEG 기반 예측이 EEG보다 훨씬 정확하며, 일부 문장은 완벽하게 디코딩됨
- Brain2Qwerty의 언어 모델(Language Model)이 오타를 수정하는 능력을 가지고 있음이 확인됨

Impact of Word Type and Frequency

- 빈번하게 등장하는 단어는 예측 정확도가 높았고,
- 희귀한 단어는 정확도가 상대적으로 낮음
- 문장에서 짧은 단어(예: 관사, 전치사 등)가 더 정확하게 예측됨

Impact of Keyboard Layout

- 예측된 문자와 실제 입력된 문자가 키보드에서 물리적으로 가까운 경우, 오타가 발생할 확률이 높았음
- 즉, 뇌가 키보드의 물리적 배치를 반영하여 타이핑한다는 것을 확인할 수 있었음

Impact of Typing Errors

- 실험에서 오타가 발생하는 비율은 3.9%, 전체 문장의 65%에서 오타가 포함되었음
- 오타가 발생하면 키를 입력하는 시간이 증가하며(50ms → 114ms), 이는 뇌가 오류를 인지하고 있음을 시사
- 오타가 포함된 문장은 예측 정확도가 떨어지는 경향을 보임

Discussion

- Brain2Qwerty는 MEG를 사용할 경우 매우 높은 정확도를 보이지만, EEG의 성능은 상대적으로 낮음
- 현재 실시간 처리가 불가능하며, 문장을 다 입력한 후에야 결과가 생성됨
- 모델 훈련을 위해서는 입력 데이터(타이핑한 텍스트와 그 시점의 뇌파 데이터)가 필요하므로, 완전히 타이핑할 수 없는 환자에게는 사용이 어려움
- MEG 기기는 아직 휴대용이 아니며, 실생활에서 쉽게 사용하려면 하드웨어 발전이 필요함



- **Brain2Qwerty**는 EEG와 MEG를 사용하여 타이핑 중 뇌 활동을 분석하고 문장을 예측하는 모델
- MEG에서 특히 우수한 성능을 보였으며, 최고의 참가자는 **19%의 문자 오류율(CER)**을 기록
- 키보드의 물리적 배치, 단어 빈도수, 오타 여부가 예측 성능에 영향을 미침
- 실시간 디코딩이 어렵고, MEG 장비가 비싸고 휴대성이 떨어지는 것이 한계
- 향후 실시간 처리 가능성, 휴대용 MEG 장비 개발, 환자를 위한 적응형 학습 모델 등 연구