

Vers un décodage sémantique de l'EEG en dénomination d'images à partir d'embeddings neuronaux et textuels

Wei Hu¹, Binbin Xu²

¹ Université de Bordeaux

² EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Alès, France

Résumé

Nous présentons une preuve de concept pour le décodage sémantique multiclasse à partir d'EEG haute densité enregistré lors d'une tâche de dénomination d'images. Les noms des images sont regroupés en neuf catégories sémantiques au moyen d'embeddings textuels, tandis que les signaux EEG sont transformés canal par canal à l'aide de SingLEM, un encodeur EEG pré-entraîné. Un classificateur KNN appliqué aux représentations EARLY, NAMING et EARLY+NAMING atteint des exactitudes équilibrées de 0,562, 0,610 et 0,781, ainsi qu'un score Macro-F1 maximal de 0,784 pour la représentation combinée. Ces performances démontrent que les embeddings EEG pré-entraînés permettent d'extraire une structure sémantique robuste à partir de signaux EEG traités canal par canal.

Mots-clés

décodage EEG, décodage sémantique, dénomination d'images, embeddings neuronaux, embeddings textuels

Motivation

L'inférence du contenu sémantique à partir d'EEG non invasifs demeure un défi en raison du faible rapport signal/bruit, de la variabilité interindividuelle marquée et de l'influence des composantes perceptives et motrices. La dénomination d'images constitue un paradigme contrôlé qui mobilise la reconnaissance visuelle, l'accès lexicosémantique et la planification articulatoire selon une séquence temporelle bien documentée [1]. Cependant, les réponses orales introduisent des artefacts articulatoires susceptibles de masquer les étapes finales de cette planification. Ce projet en cours vise à développer un décodage sémantique à partir d'EEG haute densité dans le cadre d'une tâche de dénomination d'images. Notre contribution actuelle repose sur un pipeline de décodage reproductible, fondé sur une référence de contrôle rigoureuse et sur un regroupement sémantique guidé par les données au moyen d'embeddings textuels. Elle combine ainsi une référence de contrôle, un regroupement sémantique basé sur des embeddings textuels et un décodage multiclasse à partir de représentations EEG pré-entraînées.

Données et tâche

Vingt participants sains, francophones natifs, ont été recrutés pour cette étude. Quatre d'entre eux ont été exclus en

raison de problèmes d'acquisition, laissant un échantillon final de $n = 16$ sujets pour l'analyse. Les participants devaient dénommer des dessins au trait en noir et blanc issus d'un corpus d'images standardisé [2]. Chaque essai comprenait une période de référence pré-stimulus (-1 à 0 s), la présentation de l'image (0 à 1,5 s) et une période de dénomination orale (1,5 à 3 s). Chaque participant a réalisé 270 essais, dont 70 essais de contrôle sur une image de chien répétée et 200 essais sur des images distinctes, chacune associée à un mot différent. L'EEG a été enregistré sur 99 canaux, dont 96 dédiés à l'EEG et 3 à l'EOG et aux signaux faciaux [3]. Les analyses sémantiques portent sur les 200 images distinctes ; les 70 essais de contrôle sont conservés dans le pipeline mais exclus de la classification sémantique.

Premier test de séparabilité sémantique

Avant le décodage multiclasse, nous avons évalué un contraste contrôle-tâche opposant l'image contrôle répétée aux 200 images de dénomination. Ce contraste ne se réduit pas à une simple séparation perceptive : la catégorie sémantique du contrôle est partagée avec certains essais de la tâche, tandis que les autres images couvrent un ensemble sémantique hétérogène. À partir de caractéristiques temps-fréquence extraites des signaux EEG, un SVM atteint une AUC de 0,793 en validation croisée à 5 plis (Fig. 1).

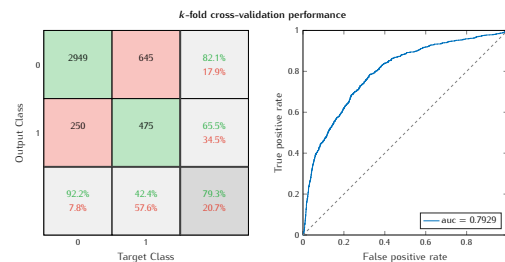


FIGURE 1. Performance du SVM pour le contraste contrôle-tâche, évaluée par validation croisée à 5 plis.

Cette première analyse établit une séparabilité EEG de base ; nous passons ensuite à des représentations SingLEM canal par canal pour tester un décodage sémantique multiclasse plus structuré.

Définition des catégories sémantiques

Le contraste contrôle-tâche offre un premier test de séparabilité, mais ne permet pas d'évaluer l'organisation sé-

mantique des représentations EEG. Nous avons donc défini une cible multiclass à partir des noms associés aux 200 images de dénomination. Chaque nom a été encodé à l’aide de Qwen3-Embedding-0.6B [4], choisi pour ses représentations sémantiques contextuelles et multilingues, puis les embeddings textuels ont été regroupés par clustering hiérarchique de Ward afin d’estimer la structure de similarité entre les items.

Les regroupements fins ont ensuite été fusionnés en neuf catégories interprétables, avec des contraintes sur le nombre de groupes et leur taille minimale, à l’aide de Qwen3-4B-Instruct-2507 : *DevicesVehicles*, *FoodNature*, *HouseholdObjects*, *HumanRelated*, *LandAnimals*, *OtherAnimals*, *PlacesSymbolsReading*, *ToolsInstrumentsMedical* et *ToysSports* (Fig. 2). Cette partition définit la cible du décodage EEG multiclass.

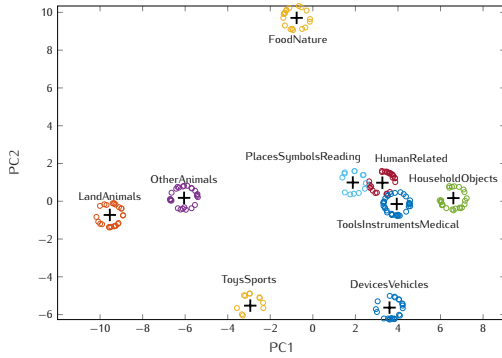


FIGURE 2. Structure sémantique des labels d’images dérivée des embeddings. Les croix noires indiquent les centroïdes des clusters projetés dans l’espace PCA, et les points colorés les labels individuels.

Décodage sémantique à partir des embeddings EEG

Les signaux EEG ont été transformés en représentations compactes au moyen de SingLEM, un modèle EEG pré-entraîné à grande échelle [5]. L’encodeur a été appliqué séparément à chaque canal, préservant une représentation par canal et par essai : chaque essai fournit ainsi 96 échantillons, correspondant aux 96 canaux EEG. Deux fenêtres temporelles ont été extraites : *EARLY*, associée à la phase précoce suivant l’apparition de l’image, et *NAMING*, associée à la phase tardive de dénomination. Chaque fenêtre fournit 16 caractéristiques.

Nous avons évalué un classifieur KNN standardisé pour le décodage des neuf classes sémantiques définies précédemment, en excluant les essais de contrôle car ils ne constituent qu’une seule classe. La validation croisée stratifiée à 5 plis est réalisée au niveau des échantillons canal-essai ; elle estime la séparabilité sémantique dans l’ensemble des événements observés, plutôt que la généralisation à des images entièrement nouvelles.

Comme le montre le Tab. 1, l’exactitude équilibrée atteint 0.562 avec *EARLY*, 0.610 avec *NAMING*, et 0.781 avec *EARLY+NAMING*. Le score Macro-F1 suit le même pro-

Caractéristiques	Dim.	Acc. équilibrée	Macro-F1
EARLY	16	0.562 ± 0.002	0.566 ± 0.002
NAMING	16	0.610 ± 0.003	0.613 ± 0.003
EARLY+NAMING	32	0.781 ± 0.002	0.784 ± 0.002

TABLE 1. Décodage sémantique en neuf classes. Moyenne ± écart-type sur cinq plis stratifiés au niveau canal-essai.

fil, passant de 0.566 à 0.613, puis à 0.784 pour la représentation combinée. L’amélioration de *NAMING* par rapport à *EARLY* indique une augmentation de l’information sémantique pendant la phase de dénomination. Le gain plus marqué obtenu avec *EARLY+NAMING* suggère que les deux fenêtres temporelles apportent des informations complémentaires pour la discrimination des catégories sémantiques.

Conclusion

Ces résultats montrent qu’un décodage sémantique multiclass est possible à partir de représentations EEG compactes extraites au niveau des canaux. Le passage de *EARLY* à *NAMING* améliore les performances, ce qui est cohérent avec une augmentation de l’information sémantique pendant la phase de dénomination. La performance optimale est obtenue avec la représentation *EARLY+NAMING*, indiquant que les deux fenêtres temporelles fournissant des informations complémentaires plutôt que redondantes.

Cette étude établit ainsi une première référence pour relier des embeddings EEG pré-entraînés à des catégories sémantiques dérivées d’embeddings textuels. Les analyses futures porteront sur les métriques par classe, les matrices de confusion et la distribution spatiale des performances afin d’identifier les catégories les mieux décodées et les canaux les plus informatifs.

Références

- [1] P. Indefrey and W. Levelt, “The spatial and temporal signatures of word production components,” *Cognition*, vol. 92, no. 1, pp. 101–144, 2004. Towards a New Functional Anatomy of Language.
- [2] J. G. Snodgrass and M. Vanderwart, “A standardized set of 260 pictures : Norms for name agreement, image agreement, familiarity, and visual complexity.,” *Journal of Experimental Psychology : Human Learning and Memory*, vol. 6, no. 2, p. 174–215, 1980.
- [3] V. Volpert, B. Xu, A. Tchekmedjiev, S. Harispe, A. Aksenov, Q. Mesnildrey, and A. Beuter, “Characterization of spatiotemporal dynamics in EEG data during picture naming with optical flow patterns,” *Mathematical Biosciences and Engineering*, vol. 20, no. 6, pp. 11429–11463, 2023.
- [4] Y. Zhang, M. Li, D. Long, X. Zhang, H. Lin, B. Yang, et al., “Qwen3 embedding : Advancing text embedding and reranking through foundation models,” *arXiv preprint arXiv :2506.05176*, 2025.
- [5] J. Sukhbaatar, S. Imamura, I. Inoue, S. Murakami, K. M. Hassan, S. Han, I. Chanpornpakdi, and T. Tanaka, “Singlem : Single-channel large eeg model,” *arXiv preprint arXiv :2509.17920*, Sept. 2025.