BASES CÉRÉBRALES DU TRAITEMENT DE LA PAROLE CHEZ LE NOURRISSON : EEG ET APPRENTISSAGE AUTOMATIQUE

Haroun Hassan BRAHIM

Aix Marseille Université – Faculté des sciences (Luminy)
M1 Informatique : Parcours IAAA

RESUMÉ

La perception de la parole chez le nourrisson constitue une étape fondamentale dans l'acquisition du langage. Ce travail vise à caractériser les réponses cérébrales aux phonèmes élémentaires de la parole (b et d) chez des nourrissons de trois mois à l'aide de l'électroencéphalographie (EEG) et de techniques d'apprentissage automatique. L'étude s'appuie sur l'analyse d'enregistrements EEG recueillis auprès de 24 nourrissons exposés à un continuum acoustique entre /ba/ et /da/. Les signaux EEG sont prétraités, segmentés, filtrés et soumis à une procédure de sélection des essais. Un pipeline d'apprentissage automatique basé sur la régression logistique est utilisé pour le décodage temporel des réponses cérébrales, permettant d'identifier les moments où la distinction entre les deux phonèmes émerge. Les résultats montrent la faisabilité du décodage des catégories phonémiques à partir des signaux EEG, ouvrant la voie à une meilleure compréhension des mécanismes neuronaux précoces impliqués dans la perception de la parole chez les nourrissons.

Mots clés— Apprentissage automatique, EEG, décodage, perception catégorielle, phonèmes, nourrisson.

1. INTRODUCTION

La perception de la parole est un processus cognitif évolutif qui se développe dès les premiers mois de la vie. Chez le nourrisson, cette capacité est cruciale pour l'acquisition du langage et la communication. Comprendre les bases cérébrales du traitement de la parole chez le nourrisson est un enjeu majeur en neurosciences. L'électroencéphalographie (EEG) offre une résolution temporelle élevée pour explorer les mécanismes neuronaux impliqués dans la perception de la parole [1, 2].

Dans ce projet, nous nous intéressons à la caractérisation des réponses cérébrales aux sons élémentaires de la parole (phonèmes) chez les nourrissons, en particulier au phénomène de la perception catégorielle : la capacité du système perceptif à transformer un continuum acoustique en catégories distinctes, avec un seuil net de bascule perceptive[3]. Nous nous concentrons ici sur la catégorisation des phonèmes ***/ba/*** et ***/da/***, à partir des signaux EEG recueillis auprès de nourrissons, à l'aide de méthodes d'apprentissage automatique, notamment la régression logistique, pour déterminer s'il est possible de distinguer les réponses cérébrales à ces deux phonèmes.

TER M1 encadré par Claire KABDEBON et Thomas SCHATZ.

2. MATÉRIEL ET MÉTHODOLOGIE

2.1. Participants

Dans ce projet, nous analysons les enregistrements EEG de 24 nourrissons âgés d'environ 3 mois, recrutés en région parisienne (10 filles ; âge moyen : 16 ± 1 semaine ; tranche d'âge : 13 à 18 semaines). Tous les nourrissons inclus dans cette étude sont nés en bonne santé après 37 semaines de gestation, avec un poids de naissance supérieur à 2,5 kg.

2.2. Procédure expérimentale

Les données EEG analysées dans ce projet ont été acquises grâce au matériel EGI (Eugene, Oregon, USA), permettant l'enregistrement simultané de l'activité électrique au niveau de 128 capteurs répartis sur toute la surface du crâne grâce à un bonnet spécifiquement élaboré pour l'étude du bébé. Lors de l'enregistrement, le bébé était confortablement installé sur les genoux de l'un de ses parents dans une cabine insonorisée pendant que des syllabes étaient diffusées via des haut-parleurs dans la cabine. Huit syllabes différentes (Figure 1), synthétisées à intervalles réguliers le long d'un continuum acoustique entre /ba/ et /da/ était présentées aux participants. Ces huit syllabes étaient synthétisées à partir de l'enregistrement d'une syllabe /ba/ et d'une syllabe /da/ prononcées par une même locutrice de langue maternelle française. Un algorithme de morphing entre ces deux syllabes était ensuite utilisé pour synthétiser les 8 stimuli à intervalles réguliers le long du continuum acoustique, de sorte que les 4 premiers stimuli du continuum étaient clairement perçus comme un /da/ et les 4 derniers comme un /ba/ par des auditeurs adultes de langue maternelle française. Ces 8 stimuli étaient présentés continuellement, dans un ordre pseudo-aléatoire de façon à contrôler les effets

de répétition et de contexte : au total, chaque stimulus du continuum était précédé d'une même proportion de chacun des 8 stimuli restants, et n'était jamais présenté 2 fois à la suite. Les stimuli était séparés d'un intervalle inter-stimulus aléatoirement compris entre 1000 ms et 1300 ms. Pour minimiser les mouvements et maximiser la durée d'enregistrement, les bébés étaient encouragés à s'endormir. L'enregistrement se terminait dès que les nourrissons devenaient agités. Pendant l'enregistrement, l'expérimentateur pouvait moduler le volume sonore des syllabes pour faciliter l'endormissement du participant. Pour chaque *essai*, c'est à dire chaque présentation d'un stimulus, un indicateur de ce volume sonore était enregistré. Par ailleurs, pendant l'enregistrement, l'expérimentateur codait manuellement l'état (agité ou non, endormi ou non) du bébé.

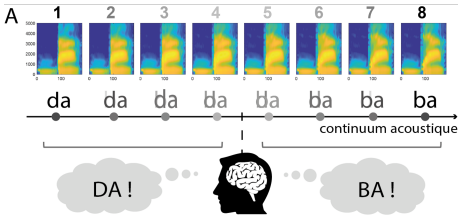


Fig. 1 – Représentations des 8 stimuli de parole utilisés, répartis à intervalles réguliers le long d'un continuum acoustique ba - da.

2.3. Traitement des données EEG

Pour chaque participant, l'activité électrique des 128 électrodes ($n_{channels}$), numérisée en continu à 250 Hz est initialement enregistrée sous forme de matrice de taille $128 \times n_{times}$ où n_{times} représente le nombre de points de temps enregistrés. Dans un premier temps, cet enregistrement EEG a été filtré (filtre passe bande entre 0.1 Hz et 40 Hz). Dans un second temps, les données filtrées ont été segmentées en *époques* : on extrait l'activité électrique autour de la présentation de la syllabe, c'est à dire entre $-200\ ms$ et $+1000\ ms$ par rapport à l'onset de la syllabe, soit 300 points de temps. Les données sont alors enregistrées sous forme d'une matrice de taille $n_{epochs} \times n_{channels} \times 300$ où n_{epochs} représente le nombre d'*époques*, c'est à dire le nombre de syllabes présentées au participant. Dans un troisième et dernier temps, un algorithme de réjection d'artefacts était appliqué aux données de marnière à rejeter les *époques* contenant des signaux parasites évidents.

2.4. Sélection des données

Une des étapes centrales du prétraitement que nous avons effectuée est l'élimination des segments de données les moins pertinents. Comme mentionné précédemment (2.2), certaines conditions telles que le volume sonore et l'état d'agitation du bébé, prises en compte lors de l'enregistrement, jouent un rôle prépondérant sur la qualité des données.

Par souci de cohérence, nous avons restreint les données en appliquant un seuil minimal de volume sonore, fixé à *0,2* [1], et en excluant les enregistrements marqués comme *agités*. Ce processus entraîne une réduction de la taille des données, ce qui peut être conséquent si l'agitation est fréquente, et peut également conduire à un déséquilibre dans les classes de données.

C'est notamment le cas pour un sujet dont les données ont été rejetées, sachant que deux tiers $(2/3)$ de sa taille initiale étaient affectés. Ainsi, nous sommes passés d'un groupe initial de 25 sujets à un groupe final de 24 sujets.

2.5. Apprentissage automatique

2.5.1. *Choix du modèle*

Nous rappelons que notre but est d'évaluer la capacité du cerveau à discriminer les phonèmes *b* et *d*. C'est-à-dire prédire ou reconstituer, à l'aide de méthodes mathématiques ou d'apprentissage automatique, les réponses cérébrales aux différents stimuli à partir des signaux EEG dont nous disposons. Du point de vue de l'apprentissage automatique, cette tâche peut être considérée comme un problème de **classification binaire**. Pour cela, nous avons fait usage de l'algorithme de la **régression logistique**, utilisé spécifiquement dans des tâches de classification binaire.

Cependant, l'utilisation de ce modèle linéaire se justifie principalement par le fait que le signal EEG mesuré par chaque capteur résulte d'une superposition linéaire des activités issues de multiples sources neuronales. Ainsi, la relation intrinsèque entre l'activité cérébrale et le signal enregistré est naturellement linéaire. Ceci rend les modèles linéaires particulièrement adaptés, à la fois pour refléter fidèlement la nature physique des signaux EEG, pour faciliter l'interprétation des résultats et garantir une extraction efficace et robuste de l'information pertinente malgré le caractère bruyant des signaux[4].

2.5.2. *Rapport signal sur bruit*

Les données de base peuvent être affectées par des bruits de diverses natures (physiologiques, techniques ou environnementaux) [2, 5]. Bien que les techniques de filtrage appliquées durant la phase de prétraitement soient destinées à réduire ces bruits, elles ne permettent pas de les éliminer complètement. Nous avons donc implémenté une méthode de réduction de bruits basée sur le moyennage des essais (*micro-moyennage)*, pour améliorer les performances de classification[6].

Dans cette approche, on moyenne N réponses cérébrales associées à un même stimulus, sélectionnées aléatoirement, avec $N \in [1, 10]$.

– Tirage sans remise : Dans cette première approche chaque essai ne sera utilisé qu'une seule fois. Par conséquent,

1. Valeur arbitraire sans unité, utilisée pour régler le volume sonore

la taille des données après micro-moyennage sera d'autant plus faible que N est grand.

– Tirage avec remise : Dans cette seconde approche chaque essai peut être utilisé une ou plusieurs fois. Avec cette approche, nous garantissons une meilleure qualité des données vis-à-vis du rapport signal sur bruit, tout en gardant inchangée la taille des données.

Cette technique nous permet aussi d'introduire un paramètre *facteur*, en fonction duquel on peut augmenter artificiellement la taille de données sur la base des moyennes effectuées, en ajustant sa valeur en terme de pourcentage souhaité. L'augmentation est très coûteuse en termes de temps d'exécution et d'utilisation de la mémoire RAM, si le pourcentage voulu est élevé.

Notons que cette méthode est utilisée avec précaution, c'est-à-dire après avoir séparé les données lors de la validation croisée en ensembles indépendants d'entraînement et de validation, sur lesquels elle sera appliquée. Cela permet d'éviter de se trouver dans une situation où les essais ayant participé à la moyenne se distribuent de part et d'autre des ensembles d'entraînement et de validation, pouvant biaiser positivement les performances de décodage.

2.5.3. Décodage

L'application de modèles de classification aux données neuronales, connue sous le nom de *décodage* est devenue une technique standard dans les recherches en neurosciences basées sur l'EEG. Le décodage permet de quantifier la manière dont les états cognitifs, les caractéristiques perceptives ou les conditions expérimentales se reflètent dans les schémas d'activité cérébrale[7].

Dans cette étude, nous avons utilisé un **décodage à résolution temporelle**, où un modèle prédictif multivarié est ajusté à chaque instant temporel pour évaluer ses performances au même instant sur de nouvelles données[8]. Ainsi, pour chaque point temporel, on entraîne et valide n_{times} modèles, chacun sur l'ensemble des n_{epochs} essais, et on note le score obtenu à l'instant $t \in [1, n_{times}]$. Cela nous permettra d'analyser, de comprendre et de déterminer quand, nos stimuli d'intérêt (phonèmes **d** et **b**) deviennent distinguables dans le temps[9] : c'est le but de notre étude.

Nous avons configuré un pipeline de traitement qui comprend trois étapes principales :

Normalisation + classification + validation croisée

– Normalisation : Les données ont été normalisées à l'aide de la méthode *z-score* (standardisation), qui transforme les variables en une distribution normale centrée et réduite avec une moyenne de 0 et un écart type de 1. Cette étape est cruciale pour améliorer les performances des algorithmes sensibles à l'échelle[10].

– Classification : Nous avons utilisé la régression logistique avec le solveur intégré *liblinear*. Ce modèle linéaire prédit la probabilité qu'un événement appartienne à une classe donnée (0 ou 1) à partir des observations, grâce à des représentations mathématiques sous-jacentes. Autrement dit, elle vise à expliquer une variable binaire Y à l'aide de p variables explicatives $X_1, \ldots, X_p$, à partir de n observations $(x_1, y_1), \ldots, (x_n, y_n)$, où $x_i \in \mathbb{R}^p$ et $y_i \in \{0, 1\}$. Dans notre cas, il s'agit de distinguer /da/ (classe 0) de /ba/ (classe 1).

– Validation croisée : Pour évaluer la performance du modèle, nous nous sommes basés sur la technique de la validation croisée (*K-Folds cross-validation*). Nous en avons utilisé la variante dite *Stratifiée*, où les plis sont obtenus en préservant le pourcentage d'échantillons pour chaque classe de y[11]. Pour le besoin de nos expérimentations, nous avons fixé le nombre de plis à 5, afin d'assurer la robustesse des résultats. L'ensemble de ces scores moyenné sur toutes les itérations des plis nous permet d'obtenir une mesure de la précision spécifique à chaque instant, reflétant la caractérisation des réponses cérébrales dans le temps.

Nous nous sommes rendus compte que les performances varient fortement d'un découpage à un autre. Ainsi, pour obtenir une estimation plus stable et fiable de la performance, nous répétons n_{iter} fois ce processus d'entraînement. La moyenne des résultats de toutes les itérations donne une meilleure idée sur la capacité réelle du modèle à généraliser. On note que ces n_{iter} répétitions bien que avantageuse, sont très coûteuse en temps de calcul pour n plus grand, car on aura $n_{iter} \times n_{times} \times 5$ modèles à entraîner.

2.6. Infrastructure et configuration logicielle

Dans ce projet, le développement s'est appuyé sur l'environnement *Python* avec *Visual Studio Code* [2] et le système de gestion de versions *Git*, hébergé sur *GitLab*. L'entraînement des modèles a été réalisé sur un cluster HPC [3] du Centre de Recherche en Psychologie et Neurosciences (CRPN) dans lequel le stage a été réalisé, disposant de 5 nœuds CPU Intel Xeon dont 4 hautement disponibles (ayant jusqu'à 32 cœurs et 256 Go de RAM par nœud). Chaque tâche a été exécutée sur 4 cœurs avec 24 Go de RAM alloués, permettant de paralléliser les validations croisées et les n_{iter} répétitions. La gestion des tâches a été assurée par *Slurm* [4], ce qui a permis de saturer efficacement les nœuds disponibles et d'obtenir un gain significatif en temps d'exécution.

3. RÉSULTATS ET ANALYSE

Dans cette section, nous présentons les résultats des expériences obtenus grâce à la méthodologie décrite précédemment.

La figure 2 illustre les performances de décodage pour un sujet spécifique sur 50 itérations. Une forte variabilité des

2. VsCode est un éditeur de code extensible développé par Microsoft
3. HPC : High Performance Computing (calcul haute performance)
4. Système open source de gestion de cluster et de planification de tâches.

courbes est observée d'une itération à l'autre, ce qui justifie la réalisation de plusieurs itérations de l'algorithme. Ainsi, les scores obtenus à chaque itération sont moyennés afin d'obtenir une estimation plus robuste de la performance globale du modèle. Cette variabilité reflète le faible rapport signal-sur-bruit caractéristique des données EEG.

Les figures 3 (a) et (b) présentent les performances de décodage d'un sujet spécifique sur 50 itérations de la validation croisée, en utilisant la méthode de micro-moyennage. On observe dans les deux cas une forte fluctuation des performances au fil du temps, avec des pics de score après la présentation des stimuli, ce qui pourrait correspondre à une phase critique de traitement neuronal. En comparant les résultats de ces deux approches à ceux de la figure 2 (sans micro-moyennage), il apparaît que l'utilisation du micro-moyennage améliore légèrement les performances. C'est pourquoi nous insistons sur son utilisation.

Les figures 4 (a) et (b) montrent les performances de décodage inter-sujets, en fonction des valeurs de N, avec et sans remise lors du micro-moyennage. On observe un effet de généralisation : les performances tendent à s'améliorer progressivement en fonction de la valeur de N.

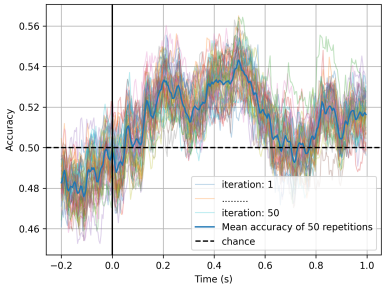
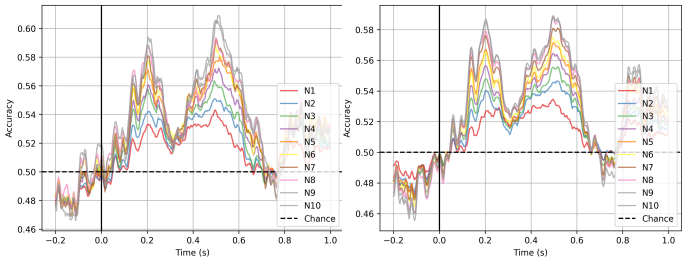
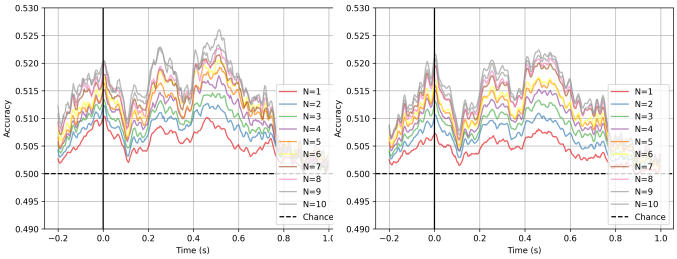


Fig. 2 – Performances de décodage pour un sujet spécifique sur 50 répétitions, sans micro-moyennage. Chaque courbe colorée représente une moyenne de la validation croisée à 5 plis, tandis que la courbe bleue correspond à leur moyenne globale. La ligne pointillée noire indique le niveau de hasard (50%).



(a) Sans remise (b) Avec remise

Fig. 3 – Performances de décodage pour un sujet sur 50 itérations, basé sur la technique de micro-moyennage, respectivement sans (a) et avec remise (b). Chaque courbe colorée représente les performances individuelles au groupement selon la valeur de N allant de 1 à 10. La ligne pointillée noire indique le niveau de hasard (50%).



(a) Sans remise (b) Avec remise

Fig. 4 – Performances de décodage au niveau du groupe basé sur la technique de micro-moyennage, respectivement sans (a) et avec remise (b). Les courbes colorées représentent les performances de 50 itérations moyennées sur tous les sujets $\forall N \in [1-10]$. La ligne pointillée noire indique le niveau de hasard (50%).

4. DISCUSSION

Les résultats de cette étude montrent que les techniques d'apprentissage automatique peuvent effectivement être utilisées pour distinguer les réponses cérébrales associées aux phonèmes chez les nourrissons.

Cependant, globalement, les performances demeurent proches du niveau aléatoire. Plusieurs facteurs peuvent expliquer ces limites. Par exemple, les techniques de prétraitement utilisées pourraient ne pas être optimisées pour extraire les caractéristiques pertinentes. De plus, une forte variabilité intra-sujet et inter-sujet a été observée, affectant fortement les scores de classification, aussi bien au niveau individuel que groupal. Enfin, les réponses cérébrales sont plus faibles pendant le sommeil et donc plus difficiles à capturer.

5. CONCLUSION ET PERSPECTIVES

Ce projet nous a permis d'explorer les bases cérébrales du traitement de la parole chez les nourrissons à l'aide de techniques d'imagerie EEG couplées à l'apprentissage automatique. Bien que les performances de décodage restent proches du seuil aléatoire, certaines tendances intéressantes ont été observées, notamment une phase critique de traitement neuronal suivant la présentation des stimuli. Ces résultats soulignent la complexité du signal EEG et la nécessité de développer des approches plus avancées permettant de capturer avec une meilleure précision les informations contenues dans les données EEG.

Finalement, cette étude représente une première étape prometteuse vers une meilleure compréhension des mécanismes neuronaux sous-jacents à la perception catégorielle des phonèmes chez les nourrissons. Des recherches futures devraient porter sur l'optimisation des techniques de prétraitement, l'utilisation de modèles d'apprentissage automatique plus performants, ainsi que sur la mise en place de tests statistiques visant à valider la significativité des effets observés

6. REFERENCES

[1] Katarzyna Blinowska and Piotr Durka, "Electroencephalography (eeg)," *Wiley encyclopedia of biomedical engineering*, 2006.

[2] Gernot R Müller-Putz, "Electroencephalography," *Handbook of Clinical Neurology*, vol. 168, pp. 249–262, 2020.

[3] Edward F Chang, J William Rieger, Keith Johnson, Mitchel S Berger, Nicholas M Barbaro, and Robert T Knight, "Categorical speech representation in human superior temporal gyrus," *Nature neuroscience*, vol. 13, no. 11, pp. 1428–1432, 2010.

[4] Jean-Rémi King, Laura Gwilliams, Chris Holdgraf, Jona Sassenhagen, Alexandre Barachant, Denis Engemann, Eric Larson, and Alexandre Gramfort, "Encoding and decoding framework to uncover the algorithms of cognition," in *The Cognitive Neurosciences*. The MIT Press, 05 2020.

[5] The Bitbrain team, "What is an EEG artifact?," https://www.bitbrain.com/blog/eeg-artifacts, 17 April 2020, Consulté en juin 2025.

[6] Tijl Grootswagers, Susan G. Wardle, and Thomas A. Carlson, "Decoding dynamic brain patterns from evoked responses : A tutorial on multivariate pattern analysis applied to time series neuroimaging data," *Journal of Cognitive Neuroscience*, vol. 29, no. 4, pp. 677–697, 04 2017.

[7] Roman Kessler, Alexander Enge, and Michael A. Skeide, "How eeg preprocessing shapes decoding performance," 2025.

[8] MNE Developers, "Decoding over time – MNE-Python," https://mne.tools/stable/auto_tutorials/machine-learning/50_decoding.html#decoding-over-time, Consulté en juin 2025.

[9] Thomas A. Carlson, Tijl Grootswagers, and Amanda K. Robinson, "An introduction to time-resolved decoding analysis for m/eeg," 2019.

[10] scikit-learn developers, "StandardScaler," https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#sklearn.preprocessing.StandardScaler, Consulté en juin 2025.

[11] scikit-learn developers, "StratifiedKFold," https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html#sklearn.model_selection.StratifiedKFold, Consulté en juin 2025.