



HAL
open science

Comparer les activités cérébrales lors de conversations humain-humain et humain-agent artificiel avec BrainPredict

Youssef Hmamouche, Magalie Ochs, Laurent Prevot, Chaminade Thierry

► **To cite this version:**

Youssef Hmamouche, Magalie Ochs, Laurent Prevot, Chaminade Thierry. Comparer les activités cérébrales lors de conversations humain-humain et humain-agent artificiel avec BrainPredict. Workshop sur les Affects, Compagnons artificiels et Interactions, CNRS, Université Toulouse Jean Jaurès, Université de Bordeaux, Jun 2020, Saint Pierre d'Oléron, France. hal-02933471

HAL Id: hal-02933471

<https://inria.hal.science/hal-02933471>

Submitted on 8 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparer les activités cérébrales lors de conversations humain-humain et humain-agent artificiel avec *BrainPredict*

Youssef Hmamouche*

youssef.hmamouche@lis-lab.fr

Aix Marseille Université, Université de Toulon, CNRS, LIS,
UMR7020

Marseille, France

Aix Marseille Université, CNRS, LPL
Aix-en-Provence, France

Laurent Prévot

laurent.prevot@univ-amu.fr

Aix Marseille Université, CNRS, LPL
Aix-en-Provence, France

Institut Universitaire de France
Paris, France

Magalie Ochs

magalie.ochs@lis-lab.fr

Aix Marseille Université, Université de Toulon, CNRS, LIS,
UMR7020

Marseille, France

Thierry Chaminade

thierry.chaminade@univ-amu.fr

Aix-Marseille Université, CNRS, INT
Marseille, France

RÉSUMÉ

Dans cet article, nous présentons un outil permettant la prédiction et la visualisation de l'activité cérébrale locale d'un individu au cours d'une conversation. Cet outil a été utilisé en particulier pour comparer l'activité cérébrale de conversation d'individu suivant que l'interlocuteur est un humain ou une tête parlante. Le module de prédiction de cet outil est construit à partir de classifieurs entraînés sur un corpus de conversations humain-humain et humain-machine incluant des données comportementales et neurophysiologiques. Le module prend en entrée des variables comportementales calculées à partir de données brutes, principalement la parole du participant et de l'interlocuteur, la vidéo de l'interlocuteur et les mouvements oculaires du participant, et permet de prédire l'activité cérébrale de l'individu. Le module de visualisation montre en temps réel la dynamique des zones actives cérébrales synchronisées avec les données comportementales brutes, permettant ainsi de faire un lien entre les comportements et l'activité cérébrale. De plus, il montre quelles variables comportementales permettent de prédire l'activité des zones cérébrales.

KEYWORDS

conversation, interactions humain-machine, activité cérébrale, prédictions

ACM Reference Format:

Youssef Hmamouche, Magalie Ochs, Laurent Prévot, and Thierry Chaminade. 2018. Comparer les activités cérébrales lors de conversations humain-humain et humain-agent artificiel avec *BrainPredict*. In *Proceedings of Workshop Workshop sur les Affects, Compagnons Artificiels et Interactions (WACAI 2020)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

L'étude des interactions sociales interpersonnelles à l'aide de l'imagerie par résonance magnétique fonctionnelle (IRMf) est devenue de plus en plus populaire ces dernières années. Dans nos travaux de recherche, récemment, nous avons collecté un corpus de conversations humain-humain et humain-machine en enregistrant l'activité cérébrale des participants en IRMf. Un tel corpus permet d'explorer le comportement du cerveau humain lors des conversations naturelles, et permet aussi de comparer les cas où l'interlocuteur est un humain ou un robot.

Dans la littérature, plusieurs projets ont été développés pour visualiser l'activité cérébrale, et de nombreux travaux ont été proposés pour prédire les signaux IRMf en fonction des variables comportementales [12, 14, 15]. Cependant, à notre connaissance, il n'y a pas de travaux fournissant des outils pour visualiser la prédiction de l'activité cérébrale des conversations non contrôlées, ainsi que les caractéristiques utilisées pour ces prédictions. Dans cet article, nous proposons un nouvel outil pour prédire et visualiser l'activité cérébrale locale lors des conversations naturelles humain-humain et humain-machine. Il se compose de deux modules, le premier module de prédiction sert à générer des prédictions basées sur les données comportementales brutes. Il prend en entrée les fichiers audio de la conversation entre le participant et l'interlocuteur, la vidéo de l'interlocuteur et les coordonnées des mouvements oculaires du participant. Ce module fournit en sortie la prédiction quant à l'activité cérébrale. Le deuxième module de visualisation permet de visualiser l'activation des zones cérébrales au cours de la conversation en fonction des prédictions obtenues.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WACAI 2020, June 03–05, 2020, WACAI, FR

© 2018 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/1122445.1122456>

Le module de prédiction utilisé est basé sur des classifieurs que nous avons entraîné sur un corpus de conversations humain-humain et humain-machine enregistrées lors d'une expérience IRMF réalisée sur 25 participants où l'activité cérébrale et les signaux comportementaux sont enregistrés. Ce module est lui-même divisé en deux traitements, le premier calcule des séries temporelles comportementales caractérisant les signaux bruts de la conversation. L'objectif est d'extraire, à partir des enregistrements bruts de la conversation, des variables descriptives multimodales de haut niveau qui seront utilisées comme variables prédictives des modèles de prédiction. Le deuxième traitement du module de prédiction consiste à appliquer des classifieurs pour prédire si une zone cérébrale locale est active ou non (classification binaire). Deux types de conversations sont pris en compte, humain-humain et humain-machine. Les variables prédictives dépendent du type de conversation (humain-humain ou humain-machine). Ainsi, selon le type d'interaction, le module de prédiction choisit les modèles de prédiction appropriés et sélectionne les variables comportementales pertinentes pour chaque zone du cerveau.

Le second module de notre outil permet de visualiser les prédictions d'activité cérébrale calculées en utilisant deux formes, la première sous forme de série temporelles montrant l'évolution de l'activation des zones cérébrales dans le temps, et la seconde montrant cette activation dans le cerveau en utilisant la bibliothèque de Python Visbrain [6].

Dans cet article, nous nous concentrons sur le fonctionnement global et la partie visualisation de l'outil. Nous proposons d'illustrer l'outil sur un exemple spécifique de prédiction de la zone cérébrale individuelle en fonction des variables comportementales lors des interactions humain-humain ou humain-machine.

Le reste de l'article est organisé comme suit. La section suivante contient une description des signaux comportementaux multimodaux utilisés par l'outil et une description des séries temporelles extraites utilisées comme variables prédictives des modèles de prédiction. Dans la section 3, nous décrivons le fonctionnement, les entrées et les sorties de l'outil proposé. Nous illustrons son utilisation à travers un exemple dans la section 4. Et la section 5 est destinée aux conclusions et travaux futurs.

2 CORPUS DE CONVERSATIONS NATURELLES HUMAIN-HUMAIN ET HUMAIN MACHINE EN IRMF

Comme mentionné dans la section 1, l'outil proposé extrait des séries temporelles multimodales à partir des signaux de conversations. Les modèles de prédiction sont ensuite appliqués sur les séries temporelles, et non sur les données brutes. Ces modèles sont entraînés à partir des séries temporelles extraites (variables prédictives) et les signaux neurophysiologiques (variables cibles) enregistrés lors d'une expérience IRMF (voir Figure 1). Dans cette section, nous décrivons cette expérience, les signaux bruts enregistrés et les séries temporelles extraites.

2.1 Traitement des signaux IRMF

Des procédures standard d'acquisition fonctionnelle d'IRMf ont été utilisées, décrites en détail dans [19]. Les images en 3 dimensions du signal BOLD (Blood Oxygen-Level Dependent) sont enregistrées

dans le cerveau entier toutes les 1.205 secondes. Des procédures de prétraitement SPM12 standard sont utilisées [18], y compris la correction des retards dans l'acquisition de tranche, le réaligement de l'image, la correction des inhomogénéités de champ magnétique, la normalisation vers l'espace MNI standard en utilisant la procédure DARTEL [2] pour le coregistration de l'anatomie des participants individuels, et enfin le lissage spatial avec un noyau gaussien tridimensionnel demi-maximum pleine largeur de 5 mm. L'extraction du signal BOLD dans les régions d'intérêt est effectuée à l'aide de la boîte à outils Conn [24], et comprend plusieurs procédures de débruitage, d'une part une dérivation linéaire utilisant un filtre passe-haut avec un seuil de 128 secondes, d'autre part en utilisant des paramètres de réaligement pour calculer les régresseurs de nuisance liés aux mouvements des participants pendant la numérisation, en troisième lieu en enregistrant les battements cardiaques et respiratoires pour supprimer les artefacts physiologiques avec la boîte à outils PhysIO [13], et enfin en extrayant le signal BOLD dans la substance blanche et le liquide céphalo-rachidien et en utilisant les cinq premières variables propres de la série temporelles comme nuisance représentant les fluctuations du signal dans les tissus cérébraux non corticaux. Une parcelle de 275 sur la base de modèles de connectivité fonctionnelle et anatomique [8] définit des régions d'intérêt (ROI) pour l'ensemble du cerveau, et des régions spécifiques sont choisies en fonction de leur emplacement anatomique. Des séries temporelles continues (385 points temporels) sont extraites pour chaque ROI et chaque session et participant représentant l'activité moyenne après débruitage.

Pour la démonstration actuelle, nous nous concentrons sur les 6 régions d'intérêts choisis afin de valider notre approche en utilisant des zones fonctionnelles bien définies : la région Gyrus Fusiform gauche et droit sont impliquées dans la perception des visages. Les cortex moteur gauche et droit, impliqués dans la production de la parole, et les Sulcus temporaux supérieurs gauche et droit qui sont impliqués dans la perception de la parole.

2.2 Traitement des signaux comportementaux multimodaux

Nous décrivons dans cette partie comment nous avons analysé les signaux comportementaux multimodaux afin d'en extraire des séries temporelles pouvant être utilisées comme variables d'entrée pour les modèles de prédiction.

2.2.1 Traitement des audios. Les conversations sont d'abord transformées en texte, puis nous les annotons et les segmentons mot à mot automatiquement en utilisant SPPAS [5]. À partir des transcriptions obtenues, nous avons extrait des séries temporelles linguistiques correspondant à l'activité de la parole (la présence de la parole), les chevauchements, les rires, les pauses remplies et le temps de réaction (qui représente dans notre cas le temps mis par un interlocuteur pour parler après que l'autre interlocuteur ait terminé son tour).

Nous avons également inclus des éléments dits particules interpersonnelles, se sont des mots qui peuvent exprimer l'attitude du locuteur (*ex.*, mais, bon, peut-être). Des marqueurs de discours, qui sont des expressions utilisées pour organiser le discours (*ex.*, je veux dire, donc, d'accord) [20]. Et des éléments lexicaux de réaction,

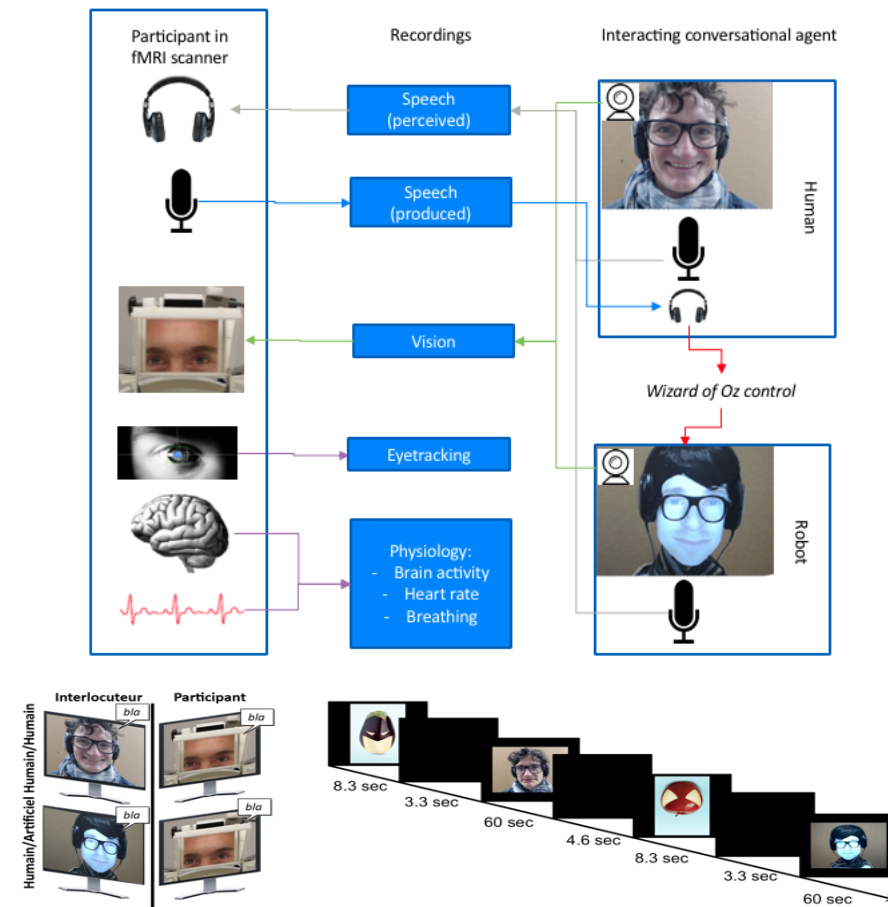


FIGURE 1: L'expérimentation IRMf.

qui sont des mots représentant la réaction ou la saisie de parole (*ex., oui, non, d'accord, ah, mh*) [10].

Nous avons également inclus la richesse lexicale avec deux métriques utilisés dans [16]. La première métrique considère le nombre de mots différents et la deuxième considère le nombre d'adjectifs plus le nombre d'adverbes *resp.*, divisés par le nombre total de mots dans le texte de chaque tour de parole. L'analyse des sentiments est également pris en compte en calculant la polarité et la subjectivité en utilisant la bibliothèque *Pattern* [21]. Le score de polarité varie entre -1 (représentant un comportement négatif) et 1 (comportement positif). La subjectivité varie entre 0 (objectif) et 1 (subjectif).

Le module qu'on utilise pour le calcul de ces deux mesures est celui présenté dans [21]. Il repose d'abord sur une association manuelle de la polarité et des scores de subjectivité à un ensemble d'adjectifs parmi les plus utilisés. Deuxièmement, un autre ensemble est extrait avec les noms les plus fréquents et les adjectifs prédécesseurs comme variables. Enfin, un classifieur kNN est appris pour déterminer les scores des adjectifs voisins de ceux annotés manuellement [21].

2.2.2 Traitement des vidéos. Afin d'identifier les expressions faciales et les mouvements de tête, nous avons utilisé Openface [3]

sur les vidéos pour détecter les coordonnées faciales (*landmarks*), les positions et rotations de tête, et les unités d'actions faciales [4], qui représentent les contractions des mouvement faciaux (des valeurs entre 0 et 1) définies à partir du système d'Ekman [7]. Les séries temporelles associées à ces fonctionnalités sont construites en analysant chaque image des vidéos.

2.2.3 Traitement des données d'eyetracking. Les données d'eyetracking sont enregistrées à l'aide du système "EyeLink 1000 Plus" [22] pendant l'expérience IRMf. Ces données oculaires brutes sont enregistrées dans des fichiers au format EDF (*European Data Format*) avec une fréquence de 1000 Hz. Ils contiennent les coordonnées du mouvement du regard en pixel, plus d'autres informations telles que le début et la fin des conversations, les saccades et les clignotements. À partir des coordonnées obtenues, nous calculons la vitesse des mouvements du regard et nous extrayons les saccades oculaires et les clignotements en tant que séries temporelles binaires pour chaque sujet, par exemple la variable qui représente les saccades prend une valeur 1 lorsqu'il y a un saccade, et 0 sinon. Ensuite, nous combinons les coordonnées des mouvements du regard avec les variables faciales obtenue par Openface, pour détecter où le sujet regarde à chaque instant.

2.2.4 Rééchantillonnage. Les séries temporelles extraites ont des fréquences différentes, tandis que les modèles d'apprentissage automatique utilisés nécessitent des variables avec le même nombre d'observations. Nous les rééchantillons pour avoir la même fréquence du signal BOLD, c-à-d, une observation chaque 1.2s. Nous avons utilisé deux stratégies de rééchantillonnage. Pour les entités continues, nous les rééchantillons en considérant leur pourcentage d'occurrence. entre chaque deux points de temps successifs. Et pour les fonctionnalités discrètes, nous considérons le nombre d'occurrences au lieu du pourcentage.

3 DESCRIPTION DE L'OUTIL BRAINPREDICT

L'outil *BrainPredict* est composé de deux modules distincts principaux : un module d'apprentissage automatique pour calculer les prédictions, et un module de visualisation. L'outil est conçu de manière modulaire, permettant une mise à jour rapide pour intégrer de nouveaux modèles de prédiction ou de nouvelles variables.

L'interface de l'outil est développée en utilisant Qt Creator en C++ [23], tandis que le module de prédiction est implémenté en Python. Ce projet est toujours en cours de développement, mais une première version est fournie sur GitHub [11], y compris le code du module de prédiction et l'interface, et incluant une démonstration montrant l'utilisation de l'outil.

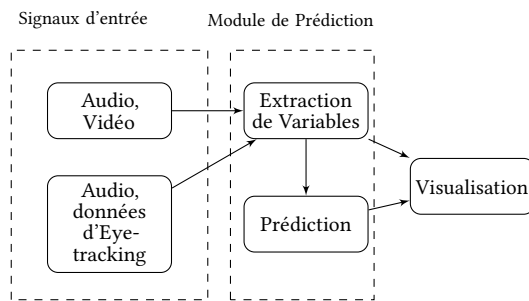


FIGURE 2: Schéma de l'outil *BrainPredict*.

3.1 Entrées

L'entrée de l'outil est une conversation entre un participant et un interlocuteur. L'interlocuteur peut être un humain ou un robot. La conversation se compose de la vidéo de l'interlocuteur, d'un fichier contenant les coordonnées des mouvements oculaires du participant et les fichiers audios correspondant à la parole du participant et de l'interlocuteur. La vidéo du participant n'est pas utilisée, car la configuration expérimentale actuelle n'a pas permis son enregistrement à l'intérieur du scanner IRM (seule la vidéo de l'interlocuteur en dehors de l'IRMf a été considérée).

3.2 Sorties

Les résultats consistent en une visualisation des prédictions sous forme de séries temporelles, montrant l'évolution de l'activation des zones cérébrales pendant la conversation. Le temps est en abscisse, tandis que l'ordonnée montre les prédictions des activations des régions, 0 pour inactif et 1 pour actif. Une visualisation du cerveau est également fournie en utilisant la bibliothèque Visbrain [6] en

construisant en temps réel une vidéo où chaque image contient une visualisation des zones prédites au moment associé de la conversation. L'outil affiche également les noms des séries temporelles comportementales utilisées pour la prédiction de la zone cérébrale sélectionnée.

3.3 Fonctionnement

Le but final de l'outil *BrainPredict* est de visualiser les prédictions de l'activité cérébrale de zones cérébrales spécifiques sur la base des valeurs précédentes des variables comportementales extraites de la conversation. Cet outil est spécifique aux prédictions du signal BOLD suivant le protocole d'acquisition d'IRMf utilisé dans l'étape d'entraînement, où une image du cerveau entier est enregistrée toutes les 1.205 secondes. Sachant que la réponse du signal BOLD à un événement comportemental atteint son pic à environ 5 secondes après l'événement [9], nous calculons chaque prédiction en fonction des valeurs des variables comportementales des 5 secondes précédentes.

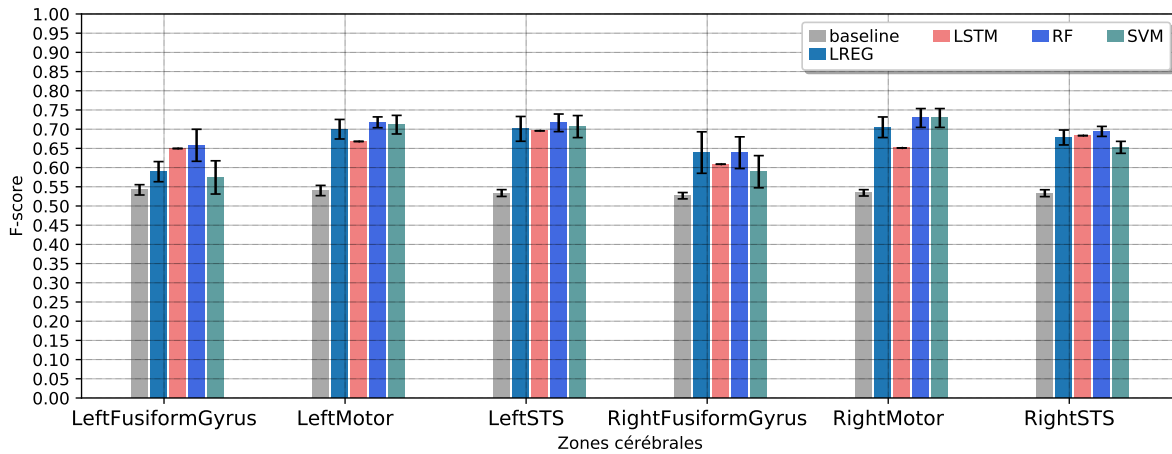
3.3.1 Extraction des variables. L'extraction des variables est effectuée en utilisant la même approche que celle appliquée à l'étape d'entraînement du module de prédiction. La sous-section 2.2 décrit cette approche, y compris les outils utilisés pour calculer les caractéristiques à partir de signaux bruts. Concrètement, la table 1 contient les noms de toutes ces caractéristiques.

Signaux brutes	Variables extraites
Audio	Activité de la parole, éléments de particules, marqueurs de discours, chevauchement de parole, temps de réaction, pauses remplies, éléments de réaction, rires, richesse lexicale, polarité, subjectivité.
Vidéo	Unités d'action faciale, coordonnées de pose de tête et coordonnées de regard, émotions, sourires.
Eyetracking	Coordonnées du mouvement du regard du participant et des variables binaires catégorisant <i>resp.</i> , la présence de saccades, et si le participant regarde le visage, les yeux de la bouche de l'interlocuteur.

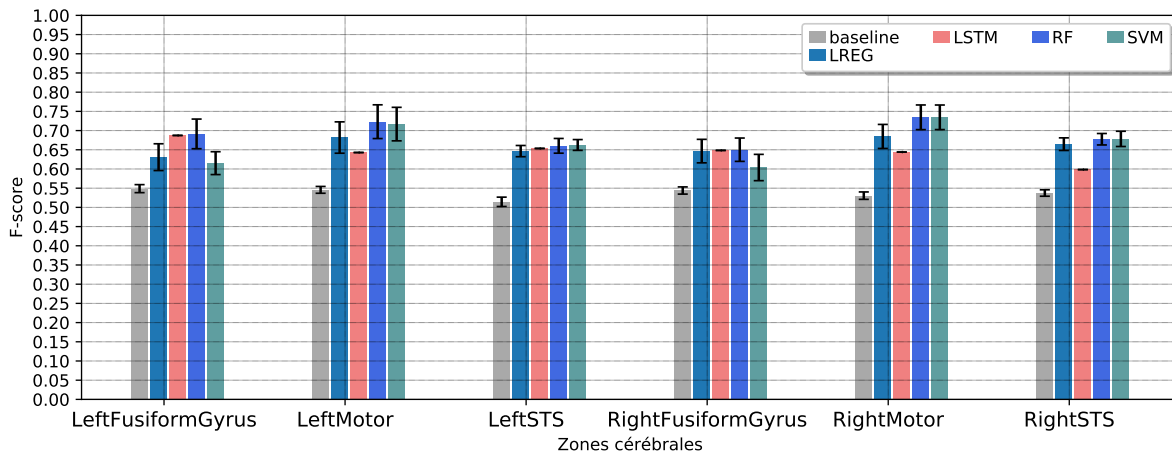
TABLE 1: Les variables comportementales extraites.

3.3.2 Prédiction. Deux types de conditions sont traitées indépendamment, où l'interlocuteur est soit un humain ou soit un robot (tête parlante). Au total, pour chaque condition (humain-humain et humain-robot), les variables extraites contiennent 13248 observations. La méthodologie de prédiction que nous avons adoptée est basée sur la prédiction si une zone cérébrale est active ou non, en discrétisant d'abord le signal BOLD dans chaque région d'intérêt en une variable binaire. Cette approche nous semble plus cohérente car ce qui est important dans ce problème n'est pas de prédire exactement la valeur continue du signal BOLD, mais juste de prédire si une région est active ou non.

La discrétisation est effectuée après avoir normalisé le signal BOLD en 0 et 1, puis en testant plusieurs seuils de discrétisation



(a)



(b)

FIGURE 3: Comparaison entre les performances des classifieurs évalués sur la base des F-scores pour les conditions humain-humain (a) et homme-robot (b).

proches de la moyenne, et en sélectionnant celui qui permet d’avoir les meilleures prédictions pour chaque zone cérébrale.

Le module de prédiction comprend un ensemble de classifieurs pour prédire l’activité de chaque zone cérébrale. Les classifieurs proviennent de la bibliothèque *Sickit-learn* [17] : Machine à vecteur de support (SVM), Forêt aléatoire (RF), régression logostique (LREG). Nous utilisons également le réseau LSTM (*Long Short Term Memory*) de la bibliothèque *Tensorflow* [1].

Ces modèles sont entraînés sur les données de 20 participants. Puis ils sont testés sur les données de 4 sujets. Nous avons appliqué une validation de type *10-fold-cross-validation* pour trouver les paramètres des modèles avec une étape de sélection pour réduire le nombre de variables prédictives. La validation croisée est appliquée à tous les classificateurs, à l’exception du réseau LSTM, car cela prend énormément de temps. Pour ce modèle spécifique, nous avons appliqué une passe entraînement-test directement avec une

architecture fixe composée d’une couche cachée LSTM et d’une couche de sortie entièrement connectée contenant un seul neurone avec la fonction d’activation *sigmoïde*. Ce réseau LSTM est entraîné en utilisant l’algorithme de descente de gradient stochastique et utilise l’entropie croisée binaire comme fonction objective.

Enfin, sur la base des F-scores obtenus sur l’ensemble de test, nous sélectionnons le meilleur modèle et les meilleurs variables comportementaux pour chaque zone cérébrale. Globalement, les F-scores obtenus sur les données de test varient entre 0.6 et 0.74 dans les zones cérébrales étudiées. Ces résultats sont montrés dans la figure 3, avec une comparaison d’un modèle *baseline* qui génère des prédictions de manière aléatoire avec trois stratégie : uniforme, selon la classe majoritaire, et stratifié : selon la distribution des données d’entraînement.

Dans la partie utilisation, les résultats obtenus concernant les modèles de prédiction et les prédicteurs de chaque zone cérébrale

sont utilisés par l'interface pour montrer les prédictions pour une conversation donnée dans les régions d'intérêt.

4 EXEMPLE

Illustrons les sorties de l'outil *BrainPredict* en utilisant l'exemple fourni dans la figure 5. Dans cet exemple, les entrées sont constituées de la vidéo de l'interlocuteur (la vidéo de l'agent conversationnel robotique dans sa version féminine dans ce cas, affichée dans la fenêtre "interlocuteur") et l'audio du participant et de l'interlocuteur. Les régions d'intérêt sélectionnées pour cet exemple correspondent aux zones cérébrales impliquées dans la production de la parole (cortex moteur primaire ventral gauche et droit), la perception de la parole (sillon temporal supérieur gauche et droit) et la perception du visage (gauche et gyrus fusiforme droit). Les séries temporelles sont affichées dans l'ordre indiqué dans la fenêtre de sélection en haut à gauche, et la couleur correspond aux couleurs des zones du cerveau affichées à gauche. Comme le taux d'échantillonnage de l'IRMf est une image chaque 1.205 seconde, nous avons 50 d'observations pour une conversation de 1 minute. L'illustration montre la prédiction approximativement vers le milieu de la conversation d'une minute. Il est prévu que le cortex moteur primaire ventral gauche, mais pas le droit, sera activé, ce qui est affiché en gris dans la première ligne de la fenêtre "Séries temporelles" et sur le rendu du cerveau (en haut, à droite). De même, le gyrus fusiforme gauche mais pas le droit est représenté en vert sur la vue de dessous du rendu du cerveau, et le sillon temporal supérieur devrait être activé dans les deux hémisphères (lignes roses et cyan).

Les variables prédictives obtenues des zones cérébrales sélectionnées sont stockées dans un fichier csv, mais peuvent être consultées dans un tableau séparé via un bouton. Une illustration de ce tableau est présentée dans la figure 4, où les variables sont organisées par modalités : *speech* (variables de la parole de l'interlocuteur), *speech_left* (variables de la parole du participant) et *facial_features* pour les variables faciales. Pour chaque région du cerveau, chaque modalité contient un sous-ensemble de ses variables comportementales (voir Tableau 1). Par exemple, l'IPU représente l'activité de parole, les AUi sont les unités d'action faciale, et {*pose_Rx*, *pose_Ry*, *pose_Rz*} sont les rotations de tête.

Predictors Table					
LeftMotor	RightMotor	LeftFusiformGyrus	RightFusiformGyrus	LeftSTS	RightSTS
{'speech_left...', ['IPU_left']}	{'speech_ts': ['IPU'], 'speech_left... ['IPU_left']}	{'facial_features_ts': ['AU01_r', 'AU02_r', 'AU04_r', 'AU05_r', 'AU06_r', 'AU07_r', 'AU10_r', 'AU12_r', 'AU14_r', 'AU15_r', 'AU17_r', 'AU20_r', 'AU23_r', 'AU25_r', 'AU26_r', 'pose_Rx', 'pose_Ry', 'pose_Rz', '...']}	{'facial_features_ts': ['AU01_r', 'AU02_r', 'AU04_r', 'AU05_r', 'AU06_r', 'AU07_r', 'AU10_r', 'AU12_r', 'AU14_r', 'AU15_r', 'AU17_r', 'AU20_r', 'AU23_r', 'AU25_r', 'AU26_r', 'pose_Rx', 'pose_Ry', 'pose_Rz', '...']}	{'speech_ts': ['IPU'], 'facial_featu... ['AU01_r', 'AU02_r', 'AU04_r', 'AU05_r', 'AU06_r', 'AU07_r', 'AU10_r', 'AU12_r', 'AU14_r', 'AU15_r', 'AU17_r', 'AU20_r', 'AU23_r', 'AU25_r', 'AU26_r', 'pose_Rx', 'pose_Ry', 'pose_Rz', '...']}	{'speech_ts': ['IPU', 'FilledBreaks', 'Feedbacks', 'Discourses', 'Particles', 'Laughters']}

FIGURE 4: Exemple de prédicteurs sélectionnés.

5 CONCLUSION

Dans cet article, nous avons proposé un nouvel outil pour la prédiction et la visualisation de l'activité cérébrale locale. Les données brutes neurophysiologiques et des conversations humain-humain

et humain-machine enregistrées au cours de l'expérience IRMf de manière non contrôlée représentent la base pour l'entraînement du module de prédiction utilisé. La prédiction implique le calcul de séries temporelles comportementales à partir de ces enregistrements bruts. L'outil contient également un module de visualisation permettant de visualiser les prédictions de l'activité cérébrale lors d'une conversation donnée. Comme extension de ce travail, nous visons à inclure des zones de l'ensemble du cerveau, ainsi que d'extraire d'autres séries temporelles comportementales comme prédicteurs pour améliorer la précision des modèles de prédiction.

6 REMERCIEMENTS

Cette recherche est soutenue par les subventions ANR-16-CONV-0002 (ILCB), ANR-11-LABX-0036 (BLRI) et AAP-ID-17-46-170301-11.1 par l'Initiative d'Excellence d'Aix-Marseille Iniversité (A*MIDEX), un programme français d'Investissement d'Avenir".

RÉFÉRENCES

- [1] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow : A System for Large-Scale Machine Learning. 265–283. <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>
- [2] John Ashburner. 2007. A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 1 (Oct. 2007), 95–113. <https://doi.org/10.1016/j.neuroimage.2007.07.007>
- [3] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L. Morency. 2018. OpenFace 2.0 : Facial Behavior Analysis Toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*. 59–66. <https://doi.org/10.1109/FG.2018.00019>
- [4] Marian Stewart Bartlett, Paul A Viola, Terrence J Sejnowski, Beatrice A Golomb, Jan Larsen, Joseph C Hager, and Paul Ekman. 1996. Classifying facial action. In *Advances in neural information processing systems*. 823–829.
- [5] Brigitte Bigi. 2015. SPPAS - MULTI-LINGUAL APPROACHES TO THE AUTOMATIC ANNOTATION OF SPEECH. *The Phonetician* 111-112, ISSN :0741-6164 (2015), 54–69. <https://hal.archives-ouvertes.fr/hal-01417876>
- [6] Etienne Combrisson, Raphael Vallat, Christian O'Reilly, Mainak Jas, Annalisa Pascarella, Anne-lise Saive, Thomas Thiery, David Meunier, Dmitrii Altukhov, Tarek Lajnef, Perrine Ruby, Aymeric Guillot, and Karim Jerbi. 2019. Visbrain : A Multi-Purpose GPU-Accelerated Open-Source Suite for Multimodal Brain Data Visualization. *Frontiers in Neuroinformatics* 13 (2019), 14. <https://doi.org/10.3389/fninf.2019.00014>
- [7] Paul Ekman and Wallace V Friesen. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior* 1, 1 (1976), 56–75.
- [8] Lingzhong Fan, Hai Li, Junjie Zhuo, Yu Zhang, Jiaojian Wang, Liangfu Chen, Zhengyi Yang, Congying Chu, Sangma Xie, Angela R. Laird, Peter T. Fox, Simon B. Eickhoff, Chunshui Yu, and Tianzi Jiang. 2016. The Human Brainnetome Atlas : A New Brain Atlas Based on Connectional Architecture. *Cereb Cortex* 26, 8 (2016), 3508–3526. <https://doi.org/10.1093/cercor/bhw157>
- [9] C Gössl, L Fahrmeir, and DP Auer. 2001. Bayesian modeling of the hemodynamic response function in BOLD fMRI. *NeuroImage* 14, 1 (2001), 140–148.
- [10] Agustin Gravano, Julia Hirschberg, and Štefan Beňuš. 2011. Affirmative cues words in task-oriented dialogue. *Computational Linguistics* 38, 1 (2011), 1–39.
- [11] Youssef Hmamouche, Magalie Ochs, Laurent Prévot, and Thierry Chaminade. 2019. A Tool for Predicting and Visualising Local Brain Activity. <https://github.com/Hmamouche/BrainPredict>. Accessed on 02.12.2019.
- [12] Alexander G. Huth, Wendy A. de Heer, Thomas L. Griffiths, Frédéric E. Theunissen, and Jack L. Gallant. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532, 7600 (April 2016), 453–458. <https://doi.org/10.1038/nature17637>
- [13] Lars Kasper, Steffen Bollmann, Andreea O. Diaconescu, Chloe Hutton, Jakob Heinze, Sandra Iglesias, Tobias U. Hauser, Miriam Sebold, Zina-Mary Manjaly, Klaas P. Pruessmann, and Klaas E. Stephan. 2017. The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods* 276 (2017), 56–72. <https://doi.org/10.1016/j.jneumeth.2016.10.019>
- [14] André Knops, Bertrand Thirion, Edward M. Hubbard, Vincent Michel, and Stanislas Dehaene. 2009. Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. *Science* 324, 5934 (June 2009), 1583–1585. <https://doi.org/10.1126/science.1171599>
- [15] Tom M. Mitchell, Svetlana V. Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L. Malave, Robert A. Mason, and Marcel Adam Just. 2008. Predicting

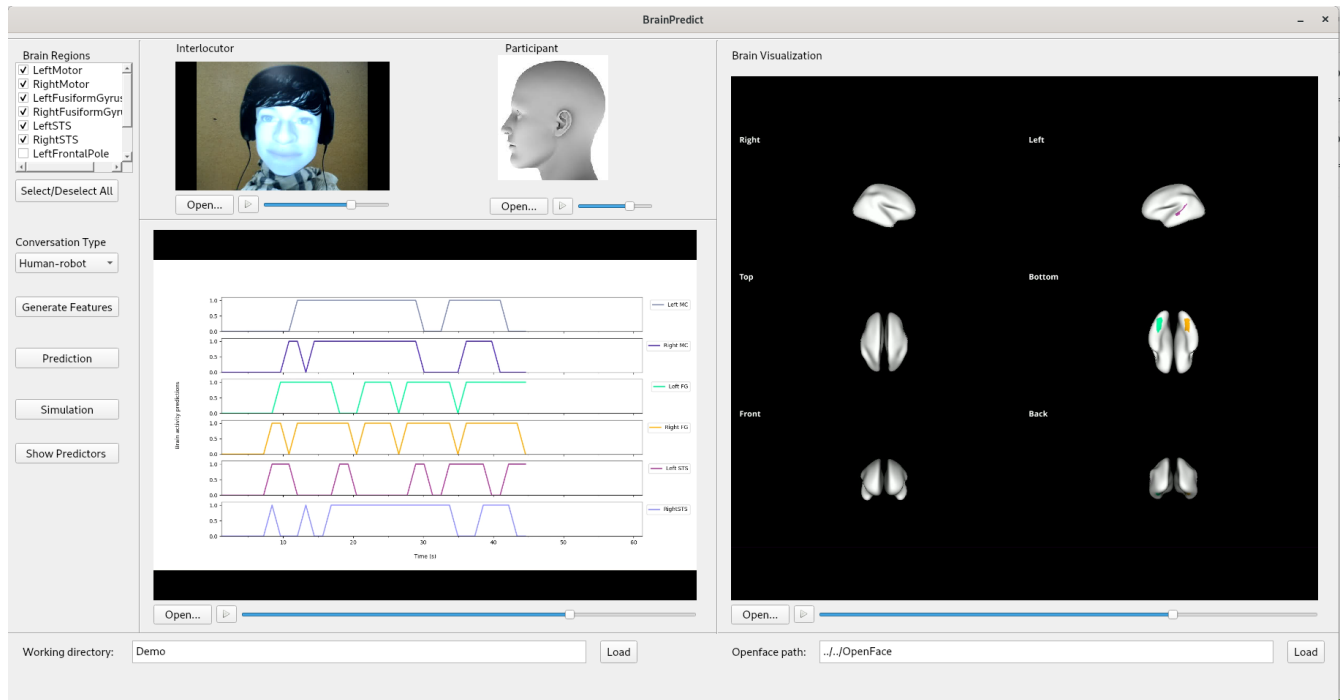


FIGURE 5: Illustration de la sortie de l’outil *BrainPredict*.

- Human Brain Activity Associated with the Meanings of Nouns. *Science* 320, 5880 (May 2008), 1191–1195. <https://doi.org/10.1126/science.1152876>
- [16] Magalie Ochs, Sameer Jain, and Philippe Blache. 2018. Toward an automatic prediction of the sense of presence in virtual reality environment. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. ACM, 161–166.
- [17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn : Machine Learning in Python. *Journal of Machine Learning Research* 12, Oct (2011), 2825–2830. <http://www.jmlr.org/papers/v12/pedregosa11a>
- [18] William D. Penny, Karl J. Friston, John T. Ashburner, Stefan J. Kiebel, and Thomas E. Nichols. 2011. *Statistical Parametric Mapping : The Analysis of Functional Brain Images*. Elsevier.
- [19] Birgit Rauchbauer, Bruno Nazarian, Morgane Bourhis, Magalie Ochs, Laurent Prévot, and Thierry Chaminade. 2019. Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B : Biological Sciences* 374, 1771 (April 2019), 20180033. <https://doi.org/10.1098/rstb.2018.0033>
- [20] Deborah Schiffrin. 1987. *Discourse markers*. Number 5. Cambridge University Press.
- [21] Tom De Smedt and Walter Daelemans. 2012. Pattern for python. *Journal of Machine Learning Research* 13, Jun (2012), 2063–2067.
- [22] SR Research. 2019. EyeLink 1000 Plus - The Most Flexible Eye Tracker. <https://www.sr-research.com/products/eyelink-1000-plus/>. Accessed on 22.11.2019.
- [23] The Qt Company. 2019. Qt | Cross-platform software development for embedded & desktop. <https://www.qt.io>. Accessed on 22.11.2019.
- [24] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. 2012. Conn : A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity* 2, 3 (May 2012), 125–141. <https://doi.org/10.1089/brain.2012.0073>