



HAL
open science

BrainPredict: a Tool for Predicting and Visualising Local Brain Activity

Youssef Hmamouche, Laurent Prevot, Magalie Ochs, Chaminade Thierry

► **To cite this version:**

Youssef Hmamouche, Laurent Prevot, Magalie Ochs, Chaminade Thierry. BrainPredict: a Tool for Predicting and Visualising Local Brain Activity. Proceedings of The 12th Language Resources and Evaluation Conference, Nov 2020, Marseille, France. pp.703-709. hal-02612819

HAL Id: hal-02612819

<https://hal.science/hal-02612819>

Submitted on 19 May 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.

BrainPredict: a Tool for Predicting and Visualising Local Brain Activity

Youssef Hmamouche^{1,2}, Laurent Prévot^{1,3}, Magalie Ochs², Thierry Chaminade⁴

¹Aix Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France

²Aix Marseille Université, Université de Toulon, CNRS, LIS, UMR7020, Marseille, France

³Institut Universitaire de France, Paris, France

⁴Aix-Marseille Université, CNRS, INT, Marseille, France

firstname.lastname@univ-amu.fr

Abstract

In this paper, we present a tool allowing dynamic prediction and visualization of an individual's local brain activity during a conversation. The prediction module of this tool is based on classifiers trained using a corpus of human-human and human-robot conversations including fMRI recordings. More precisely, the module takes as input behavioral features computed from raw data, mainly the participant and the interlocutor speech but also the participant's visual input and eye movements. The visualisation module shows in real-time the dynamics of brain active areas synchronised with the behavioral raw data. In addition, it shows which integrated behavioral features are used to predict the activity in individual brain areas.

Keywords: conversation, human-robot interactions, brain activity, predictions

1. Introduction

Studying human social interactions using functional Magnetic Resonance Imaging (fMRI) has become increasingly popular in recent years. Recently, we recorded a corpus of human-human and human-robot conversations while participants brain activity was recorded with fMRI (Rauchbauer et al., 2019). This corpus has the potential to bring new insights about how the human brain behaves during natural conversations, and allows comparisons between the cases where the interlocutor is a human or a robot. Unsurprisingly, visualization of brain activity is necessary when dealing with this type of tasks.

In the literature, several projects have been developed to visualize brain activity, and many works have been proposed to predict the fMRI signals based on behavioral features (Mitchell et al., 2008; Huth et al., 2016; Knops et al., 2009). However, we did not find works providing tools for displaying together brain activity prediction of non-controlled conversations, the raw material used in this prediction as well as the features used for these predictions. In this paper, we tackle this problem, and we propose a new tool for predicting and visualising the local brain activity during conversations. It consists of two modules, the first one is for generating predictions based on the conversation's raw behavioral data. It includes the audios of the participant and the interlocutor, the video of the interlocutor and the eye movement recordings of the participant. And the second part is for visualising the activation of the brain areas during the conversation based on the obtained predictions.

The prediction module used is based on classifiers that we trained on a corpus of human-human and human-robot conversations recorded during an fMRI experiment performed on 24 participants in which the brain activity and the behavioral signals were recorded synchronously. This module consists of two processing steps. In the first step,

the integrated behavioral time series from the raw signals of the conversation are computed. The idea here is to extract, from raw recordings of the conversation, high level multi-modal descriptive features that will be used as predictor variables of the prediction models. The second processing of the prediction module consists in applying classifiers to predict if a local brain area is active or not, i.e., a binary classification. Two types of conversations are taken into account, human-human and human-robot, thus, depending on the type of the interaction, the prediction module chooses the appropriate prediction models, and selects the relevant behavioral features.

The second part of our tool allows the visualization of the brain activity predictions during the conversation using two forms, the first one as a time series showing the evolution of the brain areas activation over time, and the second one showing the activation of the studied areas in the brain directly using the Visbrain Python library (Combrisson et al., 2019).

In this paper, we focus on the global functioning and the visualisation part of the tool since the prediction module is "plug and play" and can be updated depending on the task. We propose to illustrate the tool on a specific example of individual brain area prediction depending on behavioral cues during human-human or human-robot interactions.

The rest of the paper is organized as follows. The next section contains a description of the multi-modal behavioral signals used by the tool, and a description of the extracted time series used as predictive features of the prediction models. In Section 3, we describe the functioning, the inputs and the outputs of the proposed tool. We illustrate its use in Section 4. And Section 5 is for conclusions and future work.

2. Datasets

As mentioned in Section 1, the proposed tool extracts multi-modal time series from the conversation signals before applying prediction models. These models are trained on conversational data and neurophysiological signals recorded during a fMRI experiment (see Figure 1). In this section, we describe this experiment, the recorded raw signals, and the extracted time series.

2.1. Processing fMRI signals

Standard functional MRI acquisition procedures were used, described in details in (Rauchbauer et al., 2019). Blood Oxygen-Level Dependent (BOLD) signal 3-dimensional images are recorded in the whole brain every 1.205 seconds. Standard SPM12 preprocessing procedures are used (Penny et al., 2011), including correction for time delays in slice acquisition ("slice timing"), image realignment, magnetic field inhomogeneities correction, normalization to the standard MNI space using the DARTEL procedure for coregistration of individual participants' anatomy (Ashburner, 2007), and finally spatial smoothing with a 5-mm full-width half-maximum 3-dimensional Gaussian kernel. Extraction of the BOLD raw time series in regions of interest is performed using the conn toolbox (Whitfield-Gabrieli and Nieto-Castanon, 2012), and includes several denoising procedures, firstly a linear detrending using a high-pass filter with a threshold of 128 seconds, secondly using realignment parameters to calculate nuisance regressors related to participants' movement during scanning, thirdly taking heartbeat and breathing recordings to remove physiological artifacts with the PhysIO toolbox (Kasper et al., 2017), and finally extracting BOLD signal in the white matter and cerebrospinal fluid and using the 5 first eigenvariates of the time-series as nuisance representing signal fluctuations in non-cortical brain tissues. A parcellation based on functional and anatomical connectivity patterns (Fan et al., 2016) defines 275 regions of interest (ROIs) for the whole brain, and specific regions are chosen based on their anatomical locations and known functions. Continuous time-series (385 time points) are extracted for each ROI, and each session and participant, representing the mean activity after denoising.

For the current demonstration, we focus on 6 ROIs chosen in order to validate our approach using well-defined functional areas: the left and right Fusiform Gyrus ROIs corresponds to the Fusiform Face Area involved in face perception, the left and right Motor Cortex ROIs support speech production, and the left and right Superior Temporal Sulcus ROIs which are involved in speech perception, In the rest of this paper, we use the following abbreviations of the names of the ROIs:

- Motor Cortex: MC
- Superior Temporal Sulcus: STS
- Fusiform Gyrus: FG

2.2. Processing multimodal behavioral signals

We describe in this part how we analysed the multimodal behavioral signals in order to extract time series that can be

used as input features for prediction models.

The speeches of the conversations are first transcribed into text, then we annotated and segmented them word-by-word automatically using SPPAS (Bigi, 2015). From the obtained transcriptions, we extracted linguistic time series that consist of the speech activity (the presence of the speech), overlaps, laughters, filled pauses and the reaction time, which represents in our case the amount of time taken by an interlocutor to speak after the other interlocutor finishes his turn (Rauchbauer et al., 2020).

We also focused on interpersonal particles items, *i.e.*, words that may express the mood of the speaker (*e.g.*, *but*, *well*, *maybe*), discourses markers, which are expressions used to make the discourse organized (*e.g.*, *I mean*, *so*, *therefore*, *okay*) (Schiffrin, 1987), and feedback lexical items, which are words representing reaction and perception (*e.g.*, *yes*, *no*, *okay*, *right*) (Gravano et al., 2011). The time series describing these items are resampled according to the fMRI acquisition frequency by calculating the percentage of their existence in each 1.205s time bin.

We also included lexical richness based on two metrics from (Ochs et al., 2018), that consider the number of the different words (type-token ratio) and the number of adjectives plus the number of adverbs *resp.*, divided by the number of total words in the text of each speaking turn. Sentiment analysis is also considered by calculating the polarity and the subjectivity using the Pattern library (Smedt and Daelemans, 2012). The polarity score fluctuates between -1 (negative behavior) and 1 (positive behavior). The subjectivity is between 0 (objective) and 1 (personal). The method of their calculations is based first on a manual association of the polarity and the subjectivity scores to a set of adjectives among the most used. Second, another set is extracted with the most frequent nouns and the preceding adjectives as features. Finally, a kNN classifier is learned to determine the scores of neighbor adjectives to those manually annotated (Smedt and Daelemans, 2012).

For facial features, in the experiment we recorded the videos of the interlocutors. The Openface (Baltrusaitis et al., 2018) software is used to detect facial landmark coordinates (based on the pixel coordinate system), head pose translations and rotations, and facial action units (Bartlett et al., 1996), which represent the existence and the intensity of movement of well-defined facial expressions based on the Facial Action Coding System (Ekman and Friesen, 1976). The time series associated with these features are constructed by analyzing each image of the videos.

Eyetracking data of the participants are recorded using the "EyeLink 1000 Plus" system (SR Research, 2019) during the fMRI experiment. The raw data are recorded in European Data Format files with a frequency of 1000HZ. They contain the coordinates of the gaze on the screen using the pixel coordinate system, and other information like the beginning and end of conversations, fixations, saccades, and blinks. From the obtained coordinates, we compute the speed of the gaze movements, and we extract ocular saccades and blinks as separate features. Then, we combine the gaze movements coordinates with Openface output, more precisely the landmarks, to detect where the sub-

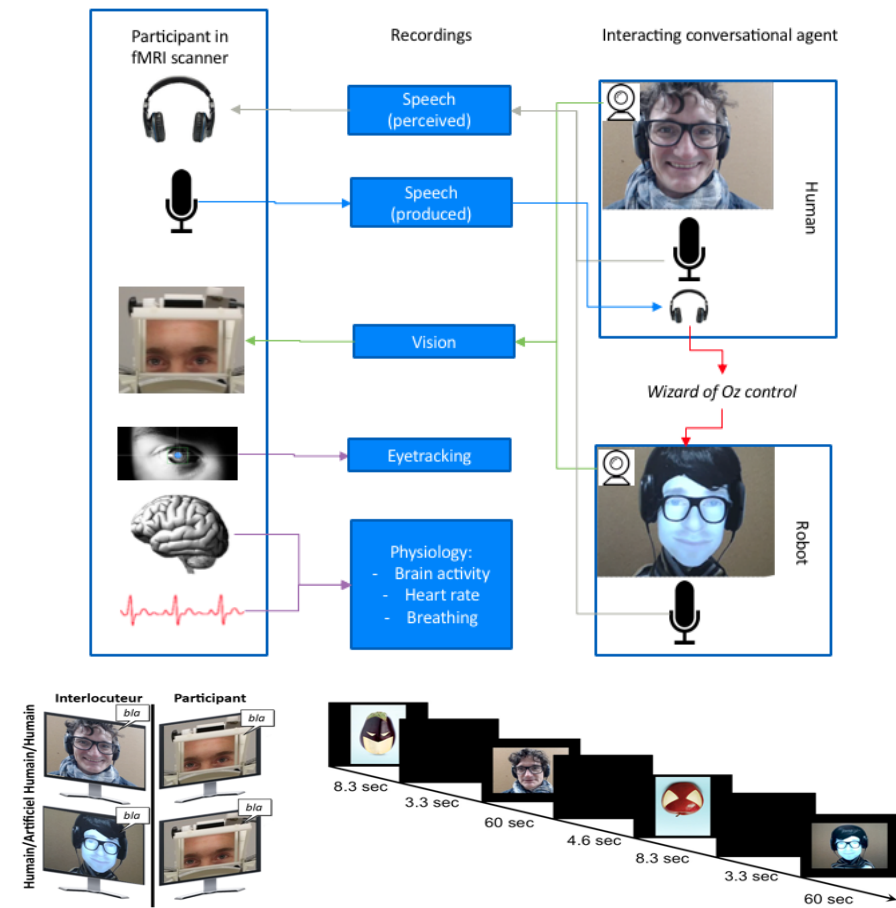


Figure 1: The fMRI experiment.

ject is looking in at each time step (face, eyes, mouth). Finally, the computed behavioral features are re-sampled and gathered together to build multivariate time series with the same number of observations for each subject.

3. Description of the tool

The tool is composed of two main separate modules: a machine learning part, dealing with computing features and prediction, and a visualisation part. It is designed in a modular way, allowing fast update in case new prediction models have to be added or new features have to be extracted. The interface of the tool is developed using Qt Creator with C++ (The Qt Company, 2019), while the prediction module is implemented in Python.¹

3.1. Inputs

The inputs of the tool are raw recordings of a conversation between a participant and an interlocutor. The interlocutor can be a human or a robot. The recordings consist of the video of the interlocutor, a file containing the eye movement coordinates of the participant, and the audios of both the participant and the interlocutor. The video of the partic-

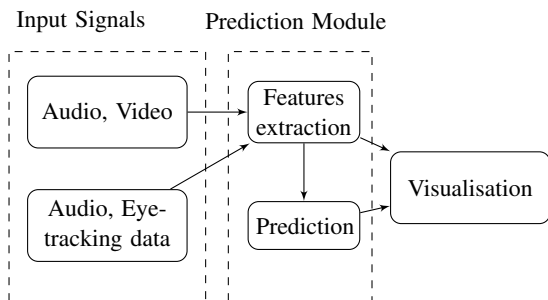


Figure 2: Schema of the BrainPredict tool.

ipant is not used, as the current experimental setup didn't allow its recording inside the MRI scanner.

3.2. Outputs

The outputs consist of a time series format of the predictions, showing the evolution of brain area activation during the conversation. Time is in abscissa, while the ordinate shows the binary 0 for inactive and 1 for active. A brain visualisation is also provided using the Visbrain library (Combrisson et al., 2019) by constructing in real-time a video where each image contains a visualisation of the predicted areas at the corresponding time of the conversation. The color code matches the time-series predictions and the saturation of the color matches the activity (low for inactive, high for active). The tool also shows the names

¹The BrainPredict tool is still undergoing development, but a first version is provided in <https://github.com/Hmamouche/BrainPredict>, including the code source of the prediction module and the interface, the detailed prediction results, and a demo showing its utilization.

of the integrated behavioral time series used for the prediction of the selected brain area. They are used by the tool, but they are also recorded in an external file as well as the obtained predictions, which can be useful to the user for further investigations.

3.3. Functioning

The goal is to visualise the brain activity of specific brain areas based on previous values of the extracted behavioral features from the conversation. This tool is specific to predictions of the BOLD signal following the fMRI acquisition protocol used in the training step, where a whole brain image is recorded every 1.205 seconds. We detail in the rest of this part the prediction procedure used by the tool, including feature extraction, and the way the used classifiers are evaluated and trained.

3.3.1. Features extraction

Feature extraction is performed using the same approach applied in the data processing step. Subsection 2.2. describes this approach including the tools used to compute the features from raw signals. Concretely, Table 1 contains the names of all the extracted behavioral features.

Raw signals	Extracted Features
Speech	Speech activity, Particles items, Discourse Markers, Overlap, Reaction Time, Filled-breaks, Feedbacks, Laughters, Lexical Richness, Polarity, Subjectivity.
Video	Facial Action Units, Head pose coordinates, Gaze coordinates.
Eyetracking data	Coordinates of the gaze movement of the participant, and binary variables categorizing <i>resp.</i> the presence of saccades, and if the participant is looking at the face, the eyes of the mouth of the interlocutor.

Table 1: The extracted behavioral features.

3.3.2. Prediction

Since we focus on predicting if a brain area is active or not, we first discretize the BOLD signal in each ROI into a binary variable. The BOLD time series of each participant are normalized, then a threshold for each ROI is used for discretization. Different thresholds are tested, and the most appropriate discretization threshold in terms of prediction scores is used for each ROI.

Knowing that the BOLD signal response to a behavioral event reaches its peak around 5 seconds after the event (Gössl et al., 2001), we compute each prediction based on the values of behavioral features in the previous 5 seconds. This delay (5s) is not fix for all ROIs and participants. Therefore, our approach consists in considering more points near to this delay, i.e., we model the BOLD signal at time t based on sequences of lagged variables of each feature between $t - 7s$ and $t - 3s$. This model has an auto-regressive form, and can be written as follows:

$$Y(t) = f(X_1(t-\tau_1 : t-\tau_2), \dots, X_k(t-\tau_1 : t-\tau_2)) + U(t), \quad (1)$$

where f is the function between the BOLD variable and the behavioral features that we aim to determine, $Y(t)$ represents the discretized BOLD variable, $\{X_1, \dots, X_k\}$ are k behavioral features, $X_i(t - \tau_1 : t - \tau_2)$ is a sequence of temporal variables of the feature X_i where $\tau_1 = 7s$ and $\tau_2 = 3s$, and $U(t)$ represents the error of the model.

This problem can be modeled using binary prediction. The classifiers used are from the Scikit-learn library (Pedregosa et al., 2011): Support-Vector Machine (SVM), Random Forest (RF), and the Logistic Regression (LREG). We also used the Long Short Term Memory (LSTM) network from the Tensorflow library (Abadi et al., 2016), since the predictive variables are in form of sequences of original behavioral features. We also included a baseline classifier generating random predictions based on 3 strategies: a stratified way, by generating predictions regarding the distribution of the training data, a uniform way by generating random predictions uniformly, and a last strategy that consists in generating constant predictions based on the most frequent label.

The evaluation procedure: Two evaluations preformed independently, according to the type of the conversations where the interlocutor is a human or a robot. In total, for each condition, we have the data of 24 participants, which consists of 13248 observations, where the predictive variables represent the temporal variables associated to the behavioral features, and the target variables are the discretized BOLD signal of 6 ROIs: left and right Motor Cortex (MC), Left and right Superior Temporal Sulcus (STS), and left and right Fusiform Gyrus (FG).

The data of all subjects are concatenated in order to have generic prediction models independent from participants. One particular remark that we should mention is that when handling multi-subjects tasks, it is important not to use the data of a given subject in the training and test sets, even with different observations, because this may cause over-fitting and inflate predictions.

Therefore, we kept the data of 4 participants as test set, which is almost 17% of all data. And data from the rest of participants (20) are used as a training set.

To evaluate the models, we considered 3 classification measures, the weighted recall, precision and F-score. We concentrate here on the weighted F-score since it considers the recall and the precision. Moreover, it is more preferable than the accuracy for imbalanced data, because it takes into account the frequencies of both classes.

A 10-fold-cross-validation is applied on the training data to find the best parameters of each classifier with a feature selection step to find the most relevant predictive variables for each ROI, where each time, the data of 2 participants are used as a validation set. Then, the best parameters of each classifier are chosen based on the mean of the obtained 10 F-scores. The 10-fold-cross-validation is applied on all classifiers except the LSTM network because it takes a huge amount of time. For this specific model, we ap-

plied a training-test pass directly with a fixed architecture composed of one LSTM hidden layer and a fully connected output layer containing one neuron to provide one prediction each time using the sigmoid activation function. The network is trained using the stochastic gradient descent algorithm and uses the binary cross-entropy as a loss function.

Finally, the classifiers with the best parameters found in the 10-fold-cross-validation are trained on the training set, then evaluated on test data.

The results: Figure 3 provides results for the baseline and the classifiers for each brain area, for both conditions, i.e., human-human and human-machine conversations. The values represent the obtained scores of the test data set, while the error bars represent the standard deviations of the 10-fold-cross-validation results. The results show that globally the Random Forrest is almost the best or very close to the best classifier for all ROIs. The LSTM network does not provide good predictions for some ROIs. This is logical because generally artificial neural networks need many examples to learn, while in our case we still do not have a lot of observations considering the fMRI recording frequency. The baseline classifier provides F-scores between 0.5 and 0.55, while the best F-scores vary between 0.65 and 0.75. Logically, it is hard to get scores close to 1, because the ROIs activation's depends on other factors that are not recorded during the fMRI experiment, like the personality of each participant for example. But what is important is that we can detect the most relevant set of behavioral factors that allows predicting each brain area.

We have also conducted a statistical test using the Student's T-test to test the equality (null hypothesis) of the means of the F-scores between the best and the baseline classifiers obtained by the 10-fold-cross-validation on the training data. This statistical test is investigated in (Dietterich, 1998) and it is one of recommended methods to compare the performance of machine learning algorithms. For the hardest ROI for which we got the worst predictions, which is the right FG, the test provides a p-value (the probability of the null hypothesis) equal to 0.0005 for human-human, and 0.003 for human-robot conditions. Therefore, the obtained F-scores of the best algorithms are significantly better than the baseline classifier at threshold of statistical significance less than 0.01.

In the utilization part, based on the obtained F-scores, we select the best classifier and the best set of behavioral predictors for each brain area that will be used by the prediction module.

4. Example

Let's illustrate the outputs of the BrainPredict tool using the example provided in Figure 5. In this example, the inputs are the visual input of the interlocutor (the video of the robotic conversational agent in its feminine version in this case, shown in the "interlocutor" window) and the audio from both the scanned participant and the interlocutor. The regions of interest are selected for the current proof-of-concept as they correspond to brain areas involved in speech production (left and right MC), speech perception

(left and right STS), and face perception (left and right FG). The time series are displayed in the order given in the top left selection window, and the colour corresponds to the brain areas colors displayed on the left. As the sampling rate of the fMRI is one volume every 1.205 seconds, we have 50 observations for a conversation of 1 min.

The illustration shows the predictions until approximately the time 45s of the conversation, when it is predicted that both the left and the right ventral primary motor cortex will not be activated, which is shown in grey in the first line in the window "Time series" and on the brain render (top, right). For the FG, only the left area that is activated and shown in green on the bottom view of the brain render. Similarly, the left STS is predicted to be activated but not the right STS (pink and cyan lines).

The obtained predictor features of the selected brain areas are stored in a csv file, but can be viewed in a separate table via the button "Show Predictors". An illustration of this table is shown in Figure 4, where the features are organized by modalities: speech (speech features of the interlocutor), speech_left (speech features of the participant), and facial features. For each brain region, each modality contains a subset from its behavioral features (see Table 1). For example, the IPU represents the speech activity, the AU_i are the facial action units, and {pose_Rx, pose_Ry, pose_Rz} are the head rotation coordinates.

5. Conclusion

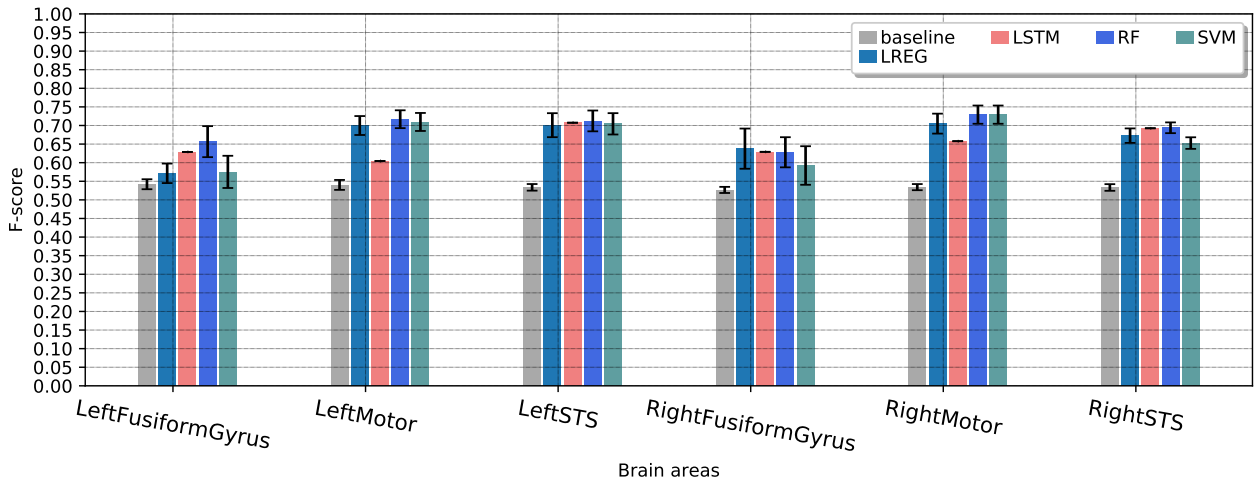
In this paper, we have proposed a new tool for the prediction and visualisation of local brain activity. The raw data recorded during the fMRI experiment investigating natural human-human and human-robot conversations are the primary sources of the prediction module of this tool. Prediction entails calculating integrated behavioral time-series on the basis of these raw recordings. The tool also contains a visualisation module allowing to see the brain activity unfold during a given conversation.

We have developed this project with the goal to provide a machine learning tool that can be used by social cognitive neuroscience researchers to analyse the dependencies between behavioral features and local brain areas, and to study the differences between human-human and human-robot interactions during unconstrained interactions.

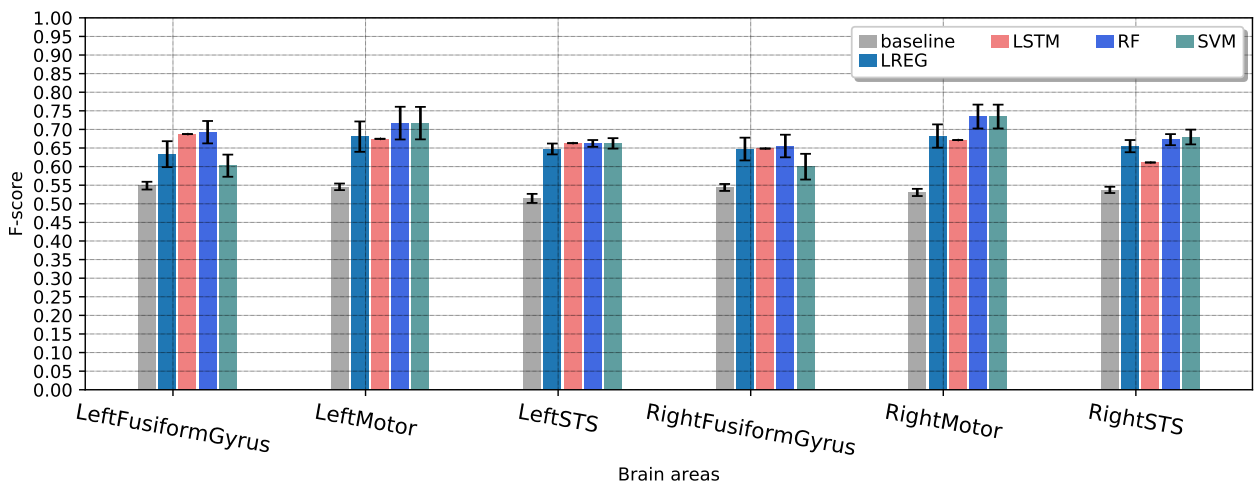
In the future, we aim to include the possibility to visualize areas from the whole brain, as well as extracting more behavioral time series as predictors to improve the prediction models.

6. Bibliographical References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. pages 265–283.
- Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113.
- Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. (2018). OpenFace 2.0: Facial Behavior Analysis



(a)



(b)

Figure 3: A comparison between the performance of the evaluated prediction models based on the F-score for human-human (a) and human-robot (b) conditions.

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', '...]	{'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', AU05_r', AU06_r', AU07_r', AU10_r', AU12_r', AU14_r', AU15_r', AU17_r', AU20_r', AU23_r', AU25_r', AU26_r', AU10_r', '...]	{'speech_ts': [IPU], 'FilledBreaks', 'Discourses', 'Particles', 'Laughters']}]

Figure 4: Example of the predictor features output

Toolkit. In *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pages 59–66.

Bartlett, M. S., Viola, P. A., Sejnowski, T. J., Golomb, B. A., Larsen, J., Hager, J. C., and Ekman, P. (1996). Classifying facial action. In *Advances in neural information processing systems*, pages 823–829.

Bigi, B. (2015). SPPAS - multi-lingual approaches to the automatic annotation of speech. *The Phonetician*, 111-

112(ISSN:0741-6164):54–69.

Combrisson, E., Vallat, R., O'Reilly, C., Jas, M., Pascarella, A., Saive, A.-I., Thiery, T., Meunier, D., Altukhov, D., Lajnef, T., Ruby, P., Guillot, A., and Jerbi, K. (2019). Visbrain: A multi-purpose gpu-accelerated open-source suite for multimodal brain data visualization. *Frontiers in Neuroinformatics*, 13:14.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation*, 10(7):1895–1923.

Ekman, P. and Friesen, W. V. (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior*, 1(1):56–75.

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., and Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb Cortex*, 26(8):3508–3526.

Gössl, C., Fahrmeir, L., and Auer, D. (2001). Bayesian modeling of the hemodynamic response function in bold

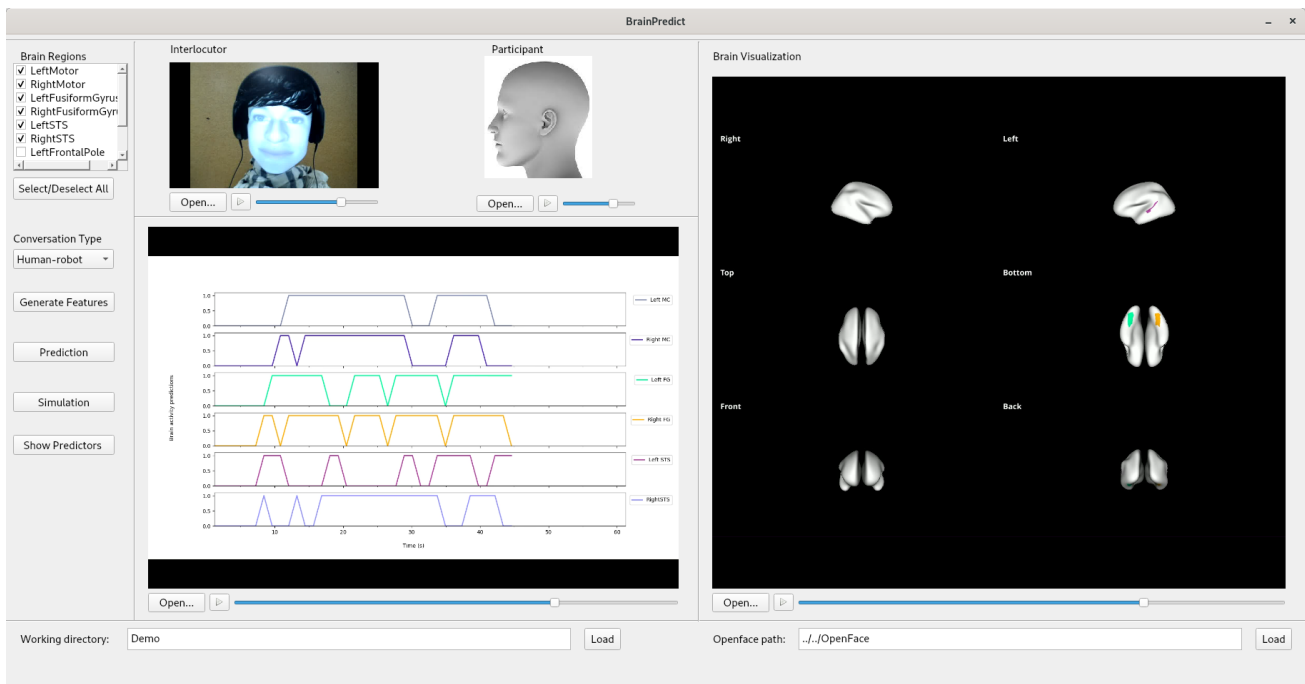


Figure 5: Illustration of the output of the BrainPredict tool.

- fmri. *NeuroImage*, 14(1):140–148.
- Gravano, A., Hirschberg, J., and Beňuš, Š. (2011). Affirmative cue words in task-oriented dialogue. *Computational Linguistics*, 38(1):1–39.
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., and Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458.
- Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., and Stephan, K. E. (2017). The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *Journal of Neuroscience Methods*, 276:56–72.
- Knops, A., Thirion, B., Hubbard, E. M., Michel, V., and Dehaene, S. (2009). Recruitment of an Area Involved in Eye Movements During Mental Arithmetic. *Science*, 324(5934):1583–1585.
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., and Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, 320(5880):1191–1195.
- Ochs, M., Jain, S., and Blache, P. (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*, pages 161–166. ACM.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(Oct):2825–2830.
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.
- Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., and Chaminade, T. (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):20180033.
- Rauchbauer, B., Hmamouche, Y., Brigitte, B., Prévot, L., Ochs, M., and Chaminade, T. (2020). Multimodal corpus of bidirectional conversation of human-human and human-robot interaction during fmri scanning. In *Proceedings of the twelfth international conference on Language Resources and Evaluation, LREC 2020*. European Language Resources Association (ELRA).
- Schiffrin, D. (1987). *Discourse markers*. Number 5. Cambridge University Press.
- Smedt, T. D. and Daelemans, W. (2012). Pattern for python. *Journal of Machine Learning Research*, 13(Jun):2063–2067.
- 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.
- The Qt Company. (2019). Qt | Cross-platform software development for embedded & desktop. <https://www.qt.io>. Accessed on 22.11.2019.
- Whitfield-Gabrieli, S. and Nieto-Castanon, A. (2012). Conn: A Functional Connectivity Toolbox for Correlated and Anticorrelated Brain Networks. *Brain Connectivity*, 2(3):125–141.