gipsa-lab **Cognitive Robotics, Interactive Systems, & Speech Processing (CRISSP)**	**Proposition de Master Recherche « Modélisation statistique de l'attention dans des interactions multi-parties »**

Contexte général

Ce travail s'inscrit dans le cadre général de la robotique cognitive, visant à doter un robot social de capacités « cognitives » (perception, motricité, cognition spatiale, raisonnement, apprentissage, langage, conscience, etc.) lui permettant de partager avec des partenaires humains une même appréhension de l'environnement d'interaction afin de réaliser des tâches collaboratives.

Nos travaux sont plus particulièrement orientés vers la génération de comportements multimodaux (parole, regard, gestes, etc.) en interaction. L'acquisition de données comportementales et l'apprentissage de ces modèles de comportement sont effectués trois phases :

- Une phase de démonstration, où le robot collecte les signaux multimodaux qu'il devrait exécuter ainsi que ceux de ses partenaires. Nous avons pour ceci développé un système de téléopération immersive [Cambuzat et al , 2018] permettant à un pilote humain d'agir et percevoir via le corps robotique afin que son « don cognitif » intègre les contraintes d'action et de perception du robot. Pendant cette phase, le robot est passif et engrange dans une mémoire comportementale ces expériences d'interaction
- Une phase de modélisation, où des algorithmes d'IA sont mis en œuvre pour calculer au mieux les comportements du robot, étant donné la tâche et les comportements de ses partenaires. A cette fin, des modèles de bout-en-bout de mise en correspondance de séquences ainsi que des modèles probabilistes (Chaines de Markov, réseaux bayésiens dynamiques) ont été développés [Bailly et al, 2018]
- Une phase d'exploitation temps-réel, où les modèles sont utilisés pour piloter le robot de manière automatique en boucle fermée. Grâce à un protocole de notation en ligne [Nguyen et al, 2017], on peut évaluer et améliorer ces modèles.

Figure 1 : Nina anime le jeu Unanimo , où deux partenaires doivent deviner les mots les plus fréquemment associés à un mot cible.

Cadre expérimental

Dans le cadre du projet « Robotrio », nous avons collecté plus d'une vingtaine d'interactions où le robot Nina avait la tâche d'animer un jeu collaboratif impliquant deux autres partenaires humains (ci-contre). Le jeu « Unanimo » consiste à deviner les mots les plus fréquemment associés à un mot cible selon des sondages effectués au sein d'une communauté d'internautes. Le robot a la charge de favoriser la prise de décision commune, d'encourager le débat et bien sûr de délivrer un score à chaque proposition après s'être assuré de la collégialité de celle-ci.

Les actions du pilote et donc du robot (parole, mouvements oculaires, mouvements de tête et des paupières, etc.) ont été enregistrées ainsi que les vidéos des sujets pris à la fois du point de vue des caméras et microphones d'oreille embarquées dans les yeux mobiles et les oreilles du robot et de caméras fixes pointées sur chaque interlocuteur, pour avoir des informations robustes sur les comportements de ces derniers.

En plus de ces données brutes, plusieurs annotations sont disponibles :

- Les phases du jeu sont connues grâce à l'environnement de téléopération.
- Les données audio des interlocuteurs humains ont déjà été transcrites.
- Un premier prétraitement des données a été effectué lors d'un stage ingénieur (Juliette Rengot, PFE ENSC) sur la distribution des directions de regard de tous les partenaires.

Cf. vidéos et données sur http://www.gipsa-lab.fr/projet/RoboTrio

Travail demandé

Le travail consistera à effectuer et évaluer une analyse puis une modélisation statistique de ces distributions, intégrant des niveaux de compréhension de l'interaction de plus en plus élaborés, notamment :

- Le découpage de la tâche en activités élémentaires (production de parole, écoute, prise d'information sur les résultats, etc)
- Le régime conversationnel mis en jeu (implication d'un ou deux interlocuteurs) [Otsuka et al, 2007]
- Le style de chaque interlocuteur (état et profil psychologique, extraversion/intraversion, rapport de dominance, etc)

Les modèles développés serviront au pilotage conjoint des mouvements de tête et des yeux du robot et seront évalués suivant le protocole déjà mentionné [Nguyen et al, 2017].

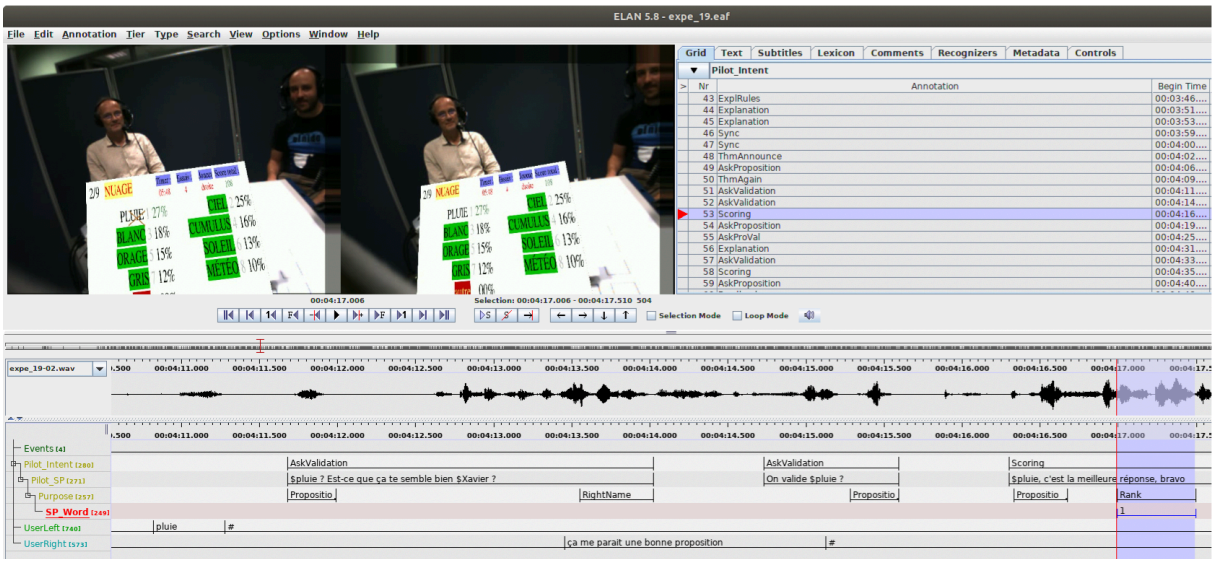

Figure 2. Exemple d'annotation de l'activité du robot : Le sujet de gauche a proposé « pluie ». Le robot a demandé deux fois la validation du choix par le sujet de droite (« ça te semble bien, xavier ? », « on valide pluie ? » et consulte la fiche de résultat (affichée en réalité augmentée) pour attribuer un score à ce mot.

Indemnité de stage

Selon les barêmes en vigueur, environ 600€ par mois

Contacts

Gérard BAILLY, DR CNRS gerard.bailly@gipsa-lab.fr
Frédéric ELISEI, IR CNRS frederic.elisei@gipsa-lab.fr

Références

Bailly, G., A. Mihoub, C. Wolf & F. Elisei (2018). Gaze and face-to-face interaction: from multimodal data to behavioral models. in Advances in Interaction Studies. Eye-tracking in interaction. Studies on the role of eye gaze in dialogue. G. Brône & B. Oben. Amsterdam, NL. John Benjamins: 139-168.

Bailly, G. & F. Elisei (2018) Demonstrating and learning multimodal socio-communicative behaviors for HRI: building interactive models from immersive teleoperation data, AI-MHRI: AI for Multimodal Human Robot Interaction Workshop at the Federated AI Meeting (FAIM), Stockholm - Sweden: pp. 39-43.

Nguyen, D.-C., G. Bailly & F. Elisei (2018) Comparing cascaded LSTM architectures for generating gaze-aware head motion from speech in HAI task-oriented dialogs, HCI International, *Las Vegas, USA: pp. 164-175.*

Cambuzat, R., Elisei, F., Bailly, G., Simonin, O., & Spalanzani, A. (2018) Immersive teleoperation of the eye gaze of social robots, International Symposium on Robotics (ISR), Munich, Germany: pp. 232-239.
Nguyen, D.-C., G. Bailly & F. Elisei (2017) An evaluation framework to assess and correct the multimodal behavior of a humanoid robot in human-robot interaction, Gesture in Interaction (GESPIN), Posznan, Poland: pp. 56-62.
Otsuka, K., Sawada, H., & Yamato, J. (2007). Automatic inference of cross-modal nonverbal interactions in multiparty conversations: " who responds to whom, when, and how?" from gaze, head gestures, and utterances. international conference on Multimodal interfaces: pp. 255-262.