

## Animation réaliste de l'activité de communication

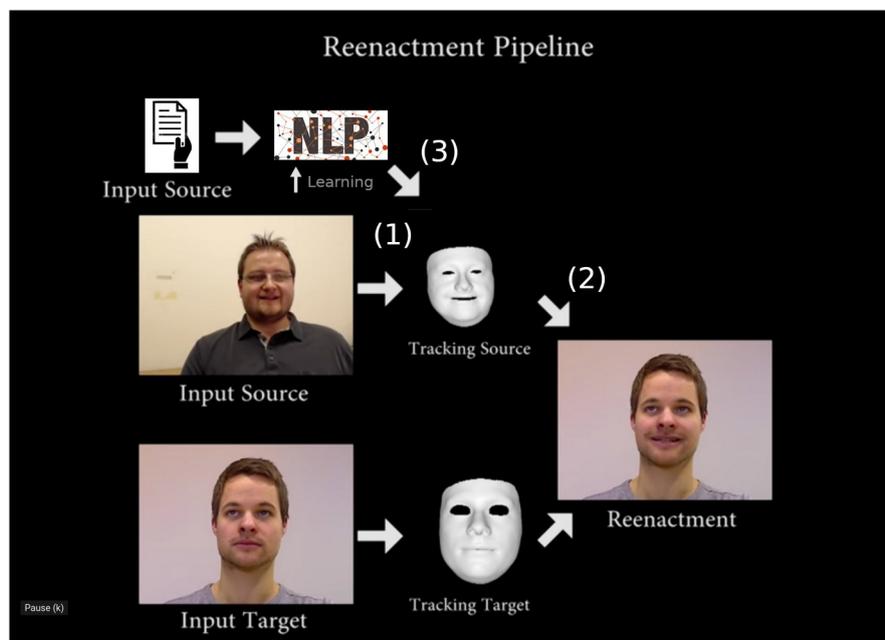
**mots-clefs** : Rendu visuel réaliste, Linguistique incarnée, Natural Language Processing / DeepLearning

**Accueil** : Laboratoire LIRIS

**Période** : Janvier/Fevrier 2022 - Juillet 2022

**Candidatures (CV, lettre de motivation) et renseignements** : frederic.armetta(a)univ-lyon1.

**Encadrement** : Didier Bottineau (linguistique enactive, lab. ICAR), Alexandre Meyer, Hamid Ladjal (capture et animation réaliste, lab. LIRIS), Frédéric Armetta (NLP, lab. LIRIS)



tâches (1), (2), (3) pour une animation réaliste de l'activité de communication (image adaptée depuis [1])

### Objectifs du stage

Ce stage a pour objet la réalisation d'une version simplifiée des tâches (1), (2) et (3).

(1) Capturer des points d'expression afin de réaliser un premier dataset

(2) Générer une animation réaliste à partir de ces points

(3) Générer l'animation à partir d'un texte quelconque (utiliser (1) pour apprendre -> générer des points expressifs pour un texte quelconque -> générer l'animation (2))

### Description générale du projet

Dans le cadre de ce projet, on s'intéresse à reproduire par une génération visuelle précise les gestes articulatoires et expressifs de l'activité langagière tels qu'envisagés par les travaux sur les modes et bases articulatoires (articulatory settings : [1], [2], [3], [4]). En effet, le réalisme d'une animation est corrélé au rendu fidèle de ces mouvements (mâchoire, langue, et par extension regard, micro-expressions), qui permettent la formation des syllabes et qui sont spécifiques aux langues, à la culture, et nécessaires à l'expressivité suivant le contexte du discours (interrogatif, affirmatif, interjectif, chanté, etc.).

L'objectif de ce projet est de pouvoir générer un rendu réaliste de l'acte de communication parlé d'un personnage, à partir d'un texte fourni, suivant les spécificités de la langue (français, anglais, espagnol, etc.). L'outil résultant permettant entre autres de faciliter l'apprentissage des langues (prononciation par imitation, etc.) et de former des enseignants.

Le projet se dégage ainsi en plusieurs thèmes :

(1) Acquisition : Constitution d'un jeu de données (dataset), permettant de corréler la configuration de la face du personnage avec le contexte du texte lu. Dans un premier temps, on s'intéresse à capturer la position des lèvres et de la mâchoire à partir d'une vidéo de personnage parlant. Différentes méthodes pourront être envisagées [2].

(2) Animation : Générer à partir de positions capturées, un rendu visuel réaliste du parlé d'un personnage. Une première approche pourra utiliser directement le flux de donnée capturée (1) pour reproduction. Dans un deuxième temps, l'animation pourra être générée à partir d'un texte libre (3).

(3) Génération de l'expression à partir de texte : Ce thème s'intéresse à exploiter les données acquises (dataset), et d'exploiter des méthodes (NLP, DeepLearning [2]) permettant de générer la position des points capturés pour un texte quelconque.

## **Profil recherché pour ce stage**

- bonne maîtrise des langages de programmation, autonomie
- capacités d'analyse
- curiosité scientifique, un intérêt pour le domaine de l'intelligence artificielle et la linguistique

## **Bibliographie**

- [1] Thies J., Zollhöfer M., Stamminger M., Theobalt C. and Nießner M. Face2Face: Real-time Face Capture and Reenactment of RGB Videos, arXiv, <https://www.youtube.com/watch?v=ohmajJTcpNk>, 2020
- [2] Goodfellow I. J., Bengio Y. and Courville A. Deep Learning, MIT Press, 2016
- [3] Delattre, P. The general phonetic characteristics of languages, final report, Technical report, US Department of Health, Education and Welfare. ERIC: ED051725, 1969
- [4] Gick B., Wilson I. and Derrick D. Articulatory Phonetics, Chichester (UK), Wiley-Blackwell, 2013
- [5] Jenner, B. 'Articulatory setting: genealogies of an idea', *Historiographia Linguistica* 28:121-141, 2001
- [6] Laver J. , The phonetic description of voice quality, Cambridge,1980