

# PRESENTATION DU PROJET DE RECHERCHE DOCTORALE

KOUASSI JEAN-CLAUDE, Ingénieur Informaticien, option Réseaux et Télécoms, Groupe EDHEC-Abidjan, Côte d'Ivoire  
Ingénieur Machine Learning, Udacity

## **THÈME PROPOSÉ: Système de traduction vocale automatisée de vidéo par l'apprentissage profond et cognition humaine.**

*Mots-clés: natural language processing, speech recognition, machine learning, deep learning*

### **DESCRIPTION**

Depuis plusieurs décennies, avec l'avènement de l'internet, l'utilisation de vidéos pour l'échange d'information s'est généralisée. Nous remarquons sa présence dans des domaines très variés : les milieux universitaires, professionnels et aussi chez le grand public. Mais cela n'est pas un effet du hasard. En effet, plusieurs études approfondies ont démontré que l'utilisation de la vidéo apporte plus de valeur cognitive chez les auditeurs, contrairement à la simple lecture. «Son utilisation en tant qu'élément d'illustration, par exemple, permet de donner à l'apprentissage une dimension plus proche de la réalité dans la mesure où une situation complexe à décrire peut être rendue plus compréhensible par la présentation d'images fixes ou de séquences animées.» (De Lièvre et al., 2000) (*voir références*).

Le caractère universel des informations véhiculés par ce media a entraîné l'invention du sous-titrage qui permet à un grand nombre d'internautes de langues diverses de suivre les vidéos diffusées dans des langues qui ne sont pas les leurs. Certes cela permet d'atteindre une plus large audience incluant ceux qui acceptent de lire les sous-titres. Mais force est de reconnaître que cette méthode est fatidique, surtout pour les longues séries de vidéos allant jusqu'à plusieurs heures.

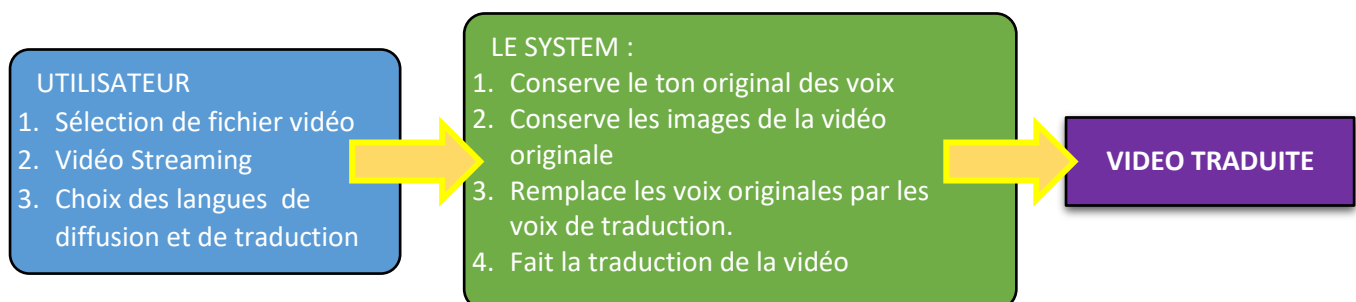
Alors pour augmenter les chances d'assimilation de l'utilisateur qui lit une vidéo diffusée originellement dans une langue qui n'est pas la sienne, pourquoi ne pas inventer un système de traduction automatisée qui traduirait instantanément, ou presque, la vidéo lue, dans sa langue préférée ? Cela devrait avoir pour répercussion immédiate une amélioration nette de la capacité d'assimilation des internautes utilisant des vidéos sous-titrées.

Un tel système devrait être aussi orienté vers le bien-être de l'humain, c'est-à-dire qu'il est souhaitable de prendre en compte les différentes possibilités de son interaction avec la cognition humaine, en vue d'augmenter les capacités d'assimilation des individus, en fonction du rendu final de l'application.

### **OBJECTIF DE RECHERCHE**

L'objectif principal de ce travail de recherche est de créer un système qui assure la traduction vocale automatisée d'une vidéo dans une autre langue désirée, autre que celle de diffusion. Il s'effectuera à l'aide des dernières avancées en matière d'intelligence artificielle, d'apprentissage automatique et profond.

Son principe de fonctionnement global est le suivant :



**Figure 1: Description fonctionnelle du système de traduction**

Par la suite, ce projet pourra servir à divers sous-objectifs à déterminer.

Il pourra alors être très utile pour les échanges :

- Entre universitaires
- Dans les cours en ligne (CLOM ou MOOC)
- La diffusion d'informations générales en ligne
- Etc.

---

## METHODOLOGIE DE RECHERCHE (PROPOSITION)

La procédure globale consistera à inventorier dans un premier temps les diverses technologies intervenant dans ce champ d'action. Puis suivra la détermination de celles qui seront utilisées en fonction du type d'approche pour l'intégration de la vidéo : immersive (action sur l'environnement extérieur de la vidéo) ou intrusive (modification du contenu du fichier).

**Pour les six (06) premiers mois :** Collecte et analyse des diverses technologies intervenant dans ce champ d'action.

**Pendant un an :** Conception et tests de fonctionnalité restreints.

**Pour les six (06) mois suivants :** Tests avec avis des individus et phases d'amélioration.

**La dernière année :** Publications Scientifiques sur les travaux effectués et rédaction de la thèse.

---

## CONTRIBUTION A L'AVANCEMENT DES CONNAISSANCES

Ce projet est une contribution dans le domaine des Interactions Homme-Machine (IHM). Jusqu'à présent les systèmes de traduction de vidéo se limitent pour la majorité à un simple sous-titrage de leur contenu, ou à une traduction séparée des fichiers vidéo, un fichier vidéo par langue.

Ce travail sera une ouverture vers une autre possibilité qui se propose d'être plus efficace d'une part en matière d'augmentation des capacités d'assimilation des individus (valeur cognitive) sans trop d'effort, et d'autre part en matière d'accroissement de l'audience. Cette façon de faire rend la vidéo accessible, à priori, à toute autre personne quelque soit sa langue. Et de plus, c'est une technique qui ne cessera de s'améliorer sans cesse par la suite.

---

## QUELQUES RÉFÉRENCES

1. **Advanced Video Coding Systems**, Wen Gao & Siwei Ma, ISBN 978-3-319-14243-2 (eBook), 2014.
2. **Pourquoi utiliser la video en formation**
3. **TricorNet: A Hybrid Temporal Convolutional and Recurrent Network for Video Action Segmentation**, Li Ding & Chenliang Xu, 2017
4. **An Improved Video Analysis using Context based Extension of LSH**, Angana Chakraborty & Sanghamitra Bandyopadhyay, 2017
5. **AENet: Learning Deep Audio Features for Video Analysis**, Naoya Takahashi, Michael Gygli, Luc Van Gool, 2017
6. **Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation**, Melvin Johnson, Mike Schuster et al., 2016
7. **Wavenet: a Generative Model for raw Audio**, Aaron van den Oord, Sander Dieleman et al., 2016
8. **Sequence to Sequence – Video to Text**, Subhashini Venugopalan, Marcus Rohrbach et al., 2015
9. **Translating Videos to Natural Language Using Deep Recurrent Neural Networks**, Subhashini Venugopalan, Huijuan Xu et al., 2015
10. **Movie/Script: Alignment and Parsing of Video and Text Transcription**, Timothee Cour, Chris Jordan et al., 2008