Transcription et Alignement de la Parole Théâtrale par Analyse Prosodique

Contexte :

La transcription automatique de la parole dans des contextes théâtraux pose des défis majeurs. La richesse du langage théâtral, la diversité des accents et des registres, les variations prosodiques marquées, ainsi que les caractéristiques acoustiques propres aux captations en salle (réverbérations, bruits de scène) rendent cette tâche particulièrement complexe. Ce sujet de stage vise à explorer un système de transcription et d'alignement automatique spécifiquement adapté aux enregistrements théâtraux. Il s'appuiera sur des corpus audio non annotés et exploitera les scripts originaux des œuvres pour guider la modélisation linguistique et prosodique.

Objectifs :

L'objectif principal de ce stage est de concevoir et d'adapter un système de reconnaissance automatique de la parole (ASR) à des corpus théâtraux, en intégrant des techniques d'alignement avec le script original basé sur des informations prosodiques. La première étape consistera à exploiter des corpus audio non annotés issus de captations de pièces de théâtre pour former ou affiner des modèles existants, en utilisant des approches d'apprentissage auto-supervisé comme Wav2Vec ou HuBERT. Le script de chaque pièce sera utilisé comme support pour enrichir la modélisation linguistique et contextualiser la transcription. Une attention particulière sera portée aux variations prosodiques propres à l'interprétation théâtrale (intonation, pauses, rythme), qui serviront à aligner la transcription produite avec le texte de la pièce et à détecter les éventuelles divergences dues à des improvisations ou omissions.

Pour aller plus loin, des approches multimodales pourront être explorées. Par exemple, l'utilisation des signaux visuels tels que les mouvements des lèvres ou les expressions faciales des comédiens pourrait améliorer la précision de la transcription, particulièrement dans les environnements acoustiquement complexes. Enfin, des techniques d'adaptation stylistique seront mises en œuvre pour mieux gérer les variations de registre, qu'il s'agisse de langue classique, contemporaine ou poétique.

Encadrement et motivation :

Ce stage est proposé à des étudiants inscrits en M2 d'informatique et intelligence artificielle.

Il sera encadré par Rémi Ronfard, directeur de recherche INRIA, directeur scientifique de l'équipe ANIMA du laboratoire LJK et du centre INRIA de l'université Grenoble Alpes, et responsable de l'action exploratoire ITHEA (informatique théâtrale) ; et Benjamin Lecouteux, professeur de l'Université Grenoble Alpes, membre de l'équipe GETALP du Laboratoire d'Informatique de Grenoble (LIG), et chercheur associé de l'action exploratoire ITHEA.

L'équipe ANIMA est spécialisée en informatique graphique et vision par ordinateur. Elle a constitué depuis plusieurs années un corpus de captations vidéo de pièces de théâtre, indexées et analysées à l'aide d'algorithmes de vision par ordinateur (détection, suivi et reconnaissance des acteurs) et accessibles en ligne sur le site http://kinoai.inria.fr à l'intention des chercheurs en études théâtrales.

L'équipe GETALP est spécialisée dans le traitement de la parole et de la langue naturelle. Elle s'intéresse en particulier à la parole théâtrale, qui est incarnée, expressive et multi-modale.

Ce stage de M2 s'inscrit dans une collaboration à long terme entre nos deux équipes sur le sujet de la compréhension, de l'analyse et de la diffusion des mises en scène de théâtre. Dans une première étape, nous cherchons à constituer un corpus de textes de théâtre alignés avec les captations vidéo de leurs mises en scène, qui sera mis à disposition de la communauté des chercheurs en sciences cognitives intéressés par le sujet de la communication théâtrale. Une première étude (Martinez 2023) a montré que les méthodes de reconnaissance vocales disponibles « sur étagère » étaient insuffisantes pour créer un tel corpus et que des approches plus spécifiques devaient être développées. C'est l'objet de ce stage.

Le stage se déroulera dans les locaux de l'action exploratoire ITHEA d'Inria à Grenoble (MINATEC). En cas de succès, il pourra être suivi par une thèse de doctorat sur le même sujet, sous réserve d'obtention d'une allocation de recherche.

Références :

Max Bain, Jaesung Huh, Tengda Han, Andrew Zisserman. WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. INTERSPEECH 2023.

Adela Barbulescu, Rémi Ronfard, Gérard Bailly. Characterization of Audiovisual Dramatic Attitudes. Interspeech 2016 - 17th Annual Conference of the International Speech Communication Association, Sep 2016.

Chow and Brown. A Musical Approach to Speech Melody. Frontiers in Psychology, Section : Cognition, Volume 9, Article 247, March 2018.

Katsalis, A. *et al.* (2023). NLP-Theatre: Employing Speech Recognition Technologies for Improving Accessibility and Augmenting the Theatrical Experience. In: Arai, K. (eds) Intelligent Systems and Applications. IntelliSys 2022. Lecture Notes in Networks and Systems, vol 543. Springer, Cham.

Emma Martinez. Conception d'un système de reconnaissance de la parole pour le théâtre. Mémoire de master Sciences du Langage, Univ. Grenoble Alpes. Sous la direction de Benjamin Lecouteux et Rémi Ronfard. Septembre 2023.

Gabriele Sofia, « Mémoire phonique « incarnée » du théâtre. Prolégomènes d'une approche cognitive », *Revue Sciences/Lettres* [En ligne], 5 | 2017.