AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mme **LAAIDI Naouar**
Soutiendra : **le Vendredi 06/02/2026** à **10H00**
Lieu : ***FSDM – Centre Visioconférence***

Une thèse intitulée :

Conception et optimisation d'un système embarqué de reconnaissance vocale en langue arabe basé sur l'intelligence artificielle pour des applications industrielles

En vue d'obtenir le ***Doctorat***

FD : **Sciences et Technologies de l'Information et de la Communication**
Spécialité : **Génie électrique et Intelligence Artificielle**

Devant le jury composé comme suit :

Nom et prénom	Etablissement	Grade	Qualité
LAMRINI Mohamed	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
OUANAN Mohammed	Faculté des Sciences, Meknès	PES	Rapporteur
NOUAITI Ayoub	Ecole Supérieure de Technologie, Meknès	MCH	Rapporteur
BOUMHIDI Jaouad	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Rapporteur
MERRAS Mostafa	Ecole Supérieure de Technologie, Meknès	MCH	Examinateur
EL MOUTAOUAKIL KARIM	Faculté Polydisciplinaire, Taza	PES	Examinateur
SATORI Hassan	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse

Résumé :

La reconnaissance automatique de la parole constitue aujourd'hui un levier essentiel pour le développement d'interfaces vocales en milieu industriel, notamment lorsque les commandes sont exprimées en arabe standard moderne. Cette thèse s'intéresse à la conception et à l'implémentation de systèmes capables d'interpréter ces commandes vocales tout en répondant aux exigences des plateformes embarquées à ressources limitées, telles que le Raspberry Pi 4.

Pour répondre à cet objectif, deux approches sont étudiées et comparées. La première repose sur des modèles HMM-GMM implémentés avec CMU Sphinx, choisis pour leur stabilité et leur faible coût computationnel. La seconde s'appuie sur une architecture neurale combinant réseaux convolutionnels (CNN) et convolutions temporelles (TCN), afin de capturer les dépendances temporelles du signal sans recourir aux modèles récurrents.

Les deux systèmes sont entraînés sur un corpus original de 35 commandes industrielles enregistrées par 30 locuteurs issus de différents secteurs tels que l'énergie, le ferroviaire ou la chimie, puis évalués sur Raspberry Pi selon plusieurs critères : précision, robustesse au bruit, latence et consommation mémoire. Les résultats montrent que l'architecture CNN-TCN offre un bon compromis entre fiabilité et efficacité embarquée, tout en mettant en perspective les apports et les limites de l'approche HMM-GMM pour des applications industrielles réelles.

Mots clés :

Reconnaissance automatique de la parole, industrie 4.0, arabe standard moderne (MSA), CMU Sphinx, HMM-GMM, CNN-TCN, Raspberry Pi, systèmes embarqués, apprentissage profond..

Design and Optimization of an Embedded Arabic Speech Recognition System Based on AI for Industrial Applications

Abstract:

This thesis focuses on the design and implementation of automatic speech recognition (ASR) systems for industrial voice commands expressed in Modern Standard Arabic (MSA), with deployment on low-resource embedded platforms such as the Raspberry Pi 4 as a primary target. The objective is to provide reliable voice-controlled interfaces that remain compatible with the computational and memory constraints of embedded devices used in industrial environments.

To achieve this goal, two complementary ASR approaches are investigated and compared. The first relies on a probabilistic framework based on Hidden Markov Models with Gaussian Mixtures (HMM-GMM), implemented using the CMU Sphinx toolkit, chosen for its robustness and low computational cost on constrained hardware. The second explores a neural architecture that combines Convolutional Neural Networks (CNN) with Temporal Convolutional Networks (TCN), in order to capture temporal dependencies in the speech signal without resorting to recurrent layers, thus improving suitability for real-time embedded inference.

Both systems are trained on a custom speech corpus of 35 industrial commands, collected from 30 native Moroccan speakers working in sectors such as energy, rail transport, and chemical processing. The implementations are deployed and evaluated on a Raspberry Pi platform with respect to recognition accuracy, noise robustness, inference latency, and resource consumption. The experimental results show that the CNN–TCN architecture offers an effective trade-off between reliability and embedded efficiency, while the comparison with the more traditional HMM–GMM approach highlights the respective strengths and limitations of each method for real-world industrial applications.

Key Words :

Automatic
speech recognition, Industry 4.0, Modern Standard Arabic, CMU Sphinx, HMM-GMM, CNN-TCN, Raspberry Pi, embedded systems, deep learning.