



## ***AVIS DE SOUTENANCE DE THESE***

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

Mme EL MANAA Khaoula  
Soutiendra : le Samedi 07/02/2026 à 10H00  
*Lieu : FSDM - Centre Visioconférence*

## *Une thèse intitulée :*

# AI ARABIC SPEECH RECOGNITION FOR REAL-TIME EMBEDDED DRONE CONTROL SYSTEMS

## *En vue d'obtenir le Doctorat*

**FD : Sciences et Technologies de l'Information et de la Communication**  
**Spécialité : Systèmes Embarqués et Intelligence Artificielle**

*Devant le jury composé comme suit :*

<b>Nom et prénom</b>	<b>Etablissement</b>	<b>Grade</b>	<b>Qualité</b>
SATORI Khalid	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Président
OUANAN Mohammed	Faculté des Sciences, Meknès	PES	Rapporteur
EL AMRANI Aumeur	Ecole Supérieure de Technologie, Meknès	PES	Rapporteur
BOUMHIDI Jaouad	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Rapporteur
EL AKKAD Nabil	Ecole Nationale des Sciences Appliquées, Fès	MCH	Examinateur
TAZI El Bachir	Faculté Polydisciplinaire, Taza	PES	Examinateur
SATORI Hassan	Faculté des Sciences Dhar EL Mahraz, Fès	PES	Directeur de thèse



### Résumé :

Cette thèse intègre la Reconnaissance Automatique de la Parole (RAP) aux systèmes embarqués afin de concevoir un drone commandé par la voix en temps réel pour la langue arabe. Il répond à un double défi : la rareté des corpus spécialisés pour les variantes arabes peu dotées et les contraintes computationnelles des plateformes embarquées. Pour combler ces lacunes, une plateforme de reconnaissance vocale open source, hautement configurable et adaptée à la fois aux spécificités linguistiques et à l'efficacité matérielle, a été développée. Un corpus interne et spécifique au domaine des commandes de drones a été constitué, intégrant les principales caractéristiques phonétiques, les structures syllabiques représentatives et un ensemble de bruits réels liés à l'environnement des UAV. Deux systèmes complémentaires de reconnaissance vocale ont ensuite été développés. La première repose sur un modèle stochastique de Markov caché (HMM) modulaire et efficace, optimisé pour les plateformes embarquées à ressources limitées. Le second est un modèle de bout en bout, prêt pour informatique en périphérie de réseau, intégrant l'arabe standard moderne (MSA) et l'arabe marocain (Darija), enrichi par une attention légère au niveau des canaux pour renforcer la robustesse face aux interférences acoustiques des drones. La conception du système s'appuie sur une analyse phonologique et syllabique approfondie, mettant en avant les patrons accentués et les structures syllabiques équilibrées, offrant ainsi un cadre rigoureux pour la modélisation de l'arabe dans des conditions à faibles ressources. Un protocole d'évaluation unifié, exécuté sur CPU, GPU et plateformes embarquées, mesure la précision de reconnaissance, le temps d'inférence et la consommation mémoire. Les résultats montrent que le modèle stochastique atteint une précision satisfaisante avec un coût computationnel minimal, tandis que le modèle de bout en bout offre une précision supérieure au prix d'une légère augmentation de la latence et de la mémoire. Les deux systèmes conservent des performances en temps réel et une compacité élevée, démontrant la faisabilité d'une reconnaissance vocale arabe robuste dans des environnements embarqués contraints. Cette thèse propose ainsi une solution pratique et déployable, reliant la conception de corpus, la modélisation statistique et l'apprentissage profond compact pour faire progresser la reconnaissance vocale de l'arabe dans des contextes à faibles ressources et à contraintes matérielles.

### Mots clés :

**Reconnaissance Automatique de la Parole (RAP) ; Control de vol des drones ; Langues Arabes à Faibles Ressources ; Modèles Stochastiques et d'Apprentissage Profond ; Mécanismes d'Attention ; Robustesse Linguistique et au Bruit ; Intelligence Artificielle en Périphérie; Systèmes Embarqués en Temps Réel ; Plates-formes à Ressources Limitées.**



## AI ARABIC SPEECH RECOGNITION FOR REAL-TIME EMBEDDED DRONE CONTROL SYSTEMS

### Abstract:

This work integrates Automatic Speech Recognition (ASR) with embedded systems to build a real-time, Arabic voice-controlled drone. It addresses the dual challenges of; the scarcity of specialized corpora for under-resourced Arabic languages drone commands and the computational limitations of embedded platforms. To bridge this gap, we design and realize a highly configurable, open-source speech-recognition platform that is both scalable and tuned for linguistic specificity and hardware efficiency. We started by creating an in-house, domain-specific datasets of Arabic drone commands that capture critical phonetic features, representative syllabic structures, and a real-world drone-related noise corpus. In this study two complementary ASR systems are developed. The first is a modular, versatile, and efficient phoneme-level stochastic Hidden Markov Model (HMM) designed for embedded platforms with limited computational resources. The second is an end-to-end, edge-ready ASR model that unifies Modern Standard Arabic (MSA) and Moroccan Arabic Darija (MAD), enhanced with lightweight channel-wise attention to improve robustness under UAV acoustic interference. The system design embeds linguistic insights through a targeted phonological and syllabic analysis emphasizing stressed, high-energy patterns and balanced phoneme–syllable coverage, offer principled guidance for Arabic modeling under resource constraints and align with observed performance gains. A comprehensive Evaluation protocol across CPUs, GPUs, and embedded low-powered platforms assesses recognition accuracy, inference time, and memory consumption. Results show that the stochastic model achieves solid accuracy with minimal computational cost, while the end-to-end model delivers higher accuracy at a modest increase in latency and memory. Both systems retain real-time performance and compactness, demonstrating the feasibility of robust Arabic speech recognition for UAV command and control. This thesis contributes a practical, deployable ASR framework that advances under-resourced language processing by linking corpus design, statistical modeling, and compact deep learning for real-time embedded applications.

### Keywords :

Automatic Speech Recognition (ASR); Drone Flight Control; Low-Resource Arabic Languages; Stochastic and Deep Learning Models; Attention Mechanisms; Linguistic and Noise-Aware Robustness; Edge AI; Embedded Real-Time Systems; Resource-Constrained Platforms.