

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE

*Université de Mohamed El-Bachir El-Ibrahimi - Bordj Bou Arreridj*

**Faculté des Sciences et de la technologie**

**Département d'Electronique**

# Mémoire

*Présenté pour obtenir*

LE DIPLOME DE MASTER

FILIERE : **Industries Electroniques**

Spécialité : **Électronique Industrielle (MCIL)**

Par

- GHALEM Dalila
- ZEGADI Yousra

*Intitulé*

*Identification des appareils électriques basée sur les modèles GMM*

*Evalué le : 10/09/2021*

*Par la commission d'évaluation composée de\* :*

<i>Nom &amp; Prénom</i>	<i>Grade</i>	<i>Qualité</i>	<i>Etablissement</i>
<i>Mme A.Laouamri</i>	<i>MAA</i>	<i>Président</i>	<i>Univ-BBA</i>
<i>Mr S.Mezaache</i>	<i>MCB</i>	<i>Examineur</i>	<i>Univ-BBA</i>
<i>Mr Hacine Gharbi Abdenour</i>	<i>MCA</i>	<i>Encadreur</i>	<i>Univ-BBA</i>
<i>Mr Ghazali Fateh</i>	<i>Doctorant</i>	<i>Co-encadreur</i>	<i>Univ-BBA</i>

*Année Universitaire 2020/2021*

\* Conformément à :

- L'arrêté n°055 du 21 janvier 2021 Fixant dispositions exceptionnelles autorisées en matière d'organisation et gestion pédagogique, de l'évaluation et de la progression des étudiants, durant la période COVID-19 au titre de l'année universitaire 2020-2021 ;
- Procès-verbal de la réunion de l'équipe du domaine des Sciences et Technologies du mois de Mai 2021.

# **Remerciement**

*Nous tenons d'abord à remercier ALLAH tout puissant, qui nous à éclairer le bon chemin pour réaliser ce travail.*

*Nous adressons à remercier nos familles pour leur encouragement.*

*Nous tenons à remercier à exprimer un profond remerciement à l'encadreur Dr. Hacine Gharbi Abdenour et co-encadreur Mr. Ghazali Fateh qui nous ont donné du temps et de l'énergie pour aboutir ce résultat.*

*Nous remercierons également tous les membres de jury pour avoir bien voulu évaluer notre travail.*

*Nous voudrions exprimer nos profonds respects à l'ensemble des enseignants et enseignantes pour leur effort pendant les années d'études.*

*Finalement, nous adressons notre remerciement pour toutes les personnes de proche ou loin qui nous souhaitent du succès.*

## Dédicace

Avant tout, je remercie ALLAH le tout puissant de m'avoir donné le courage  
durant toutes ses années d'étude malgré toutes les difficultés.

Je dédie ce travail à :

L'homme de ma vie, mon bonheur, la source de ma volonté qui sacrifié pour  
me voir réussir que ALLAH te garde dans son vaste paradis PAPA.

Ma mère la lumière de mes yeux qui m'a toujours soutenu durant toutes les  
périodes, qu'ALLAH la protège.

Mes chères sœurs.

Mes grands-parents pour leurs fiertés.

Mes oncles et mes tantes pour leurs encouragements.

Mes enseignants et enseignantes depuis le primaire.



Yousra

## Dédicaces

A l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir.

### Mon père.

A la lumière de mes jours, la source de mes efforts, la flamme de mon cœur,  
ma vie et mon bonheur.

### Ma maman.

A mes très chers frères

A mes très chères sœurs

A toutes mes amis avec qui j'ai partagé les plus beaux et inoubliables  
moments de ma vie

Toute personne qui m'a aidé un jour à réussir jusqu'ici

**Je vous dis merci**



Dalila

## Résumé

La demande croissante de l'énergie électrique mondiale exige aux fournisseurs ainsi aux distributeurs d'électricité de mettre en place une stratégie de gestion de la consommation énergétique efficace et fiable permettant aux clients de mieux maîtriser leurs consommations électriques. Cette situation provoque l'apparition de nouveaux compteurs électriques intelligents, qui intègrent dans leur fonctionnement une tâche d'identification en temps réel de la classe d'appareil en état de marche à partir des mesures faites au niveau du compteur d'énergie électrique principal. Un travail publié récemment consiste à proposer un système IAED basé sur l'extraction des paramètres statistiques des coefficients harmoniques et le classificateur KNN combiné avec la stratégie de règle de vote. Dans le but de valider la pertinence des paramètres statistiques des harmoniques et d'améliorer la complexité de calcul, nous proposons un système IAED de même architecture du système précédent, mais qui utilise le classificateur GMM combiné avec la règle de vote. Les performances du système sont évaluées sous la base PLAID.

Les résultats obtenus nous ont confirmé la pertinence des paramètres statistiques des coefficients harmoniques, ainsi que, l'efficacité du classificateur GMM dans l'amélioration de la complexité du point de vue temps de calcul.

### ملخص

إن الطلب المتزايد على الطاقة الكهربائية العالمية يتطلب من الموردين وموزعي الكهرباء وضع إستراتيجية فعالة وموثوقة لإدارة استهلاك الطاقة ، مما يسمح للمتعاملين التحكم بشكل أفضل في استهلاكهم للكهرباء. هذا ما ساعد في ظهور عدادات كهرباء ذكية جديدة ، هاته الأخيرة تحتوي في تشغيلها وظيفه جد مهمة و هي تحديد فئة الأجهزة العاملة في الوقت الفعلي ، من القياسات التي تم إجراؤها على مستوى عداد الطاقة الكهربائية. بحث نشر مؤخرا اقترح نظام تحديد الأجهزة الكهربائية المنزلية يعتمد على استخدام ميزات التوافقيات الإحصائية و المصنف KNN مع طريقة قواعد التصويت. من اجل التحقق من كفاءة استخدام هذا النوع من الميزات. في هذا العمل نقترح استخدام هاته الميزات و المصنف GMM مع إستراتيجية قواعد التصويت. يتم تقييم أداء النظام باستخدام قاعدة بيانات PLAID.

أكدت النتائج التي تم الحصول عليها أهمية استخدام ميزات التوافقيات الإحصائية . بالإضافة إلى ذلك , أظهرت لنا النتائج كفاءة المصنف GMM في تحسن أداء النظام من حيث زمن التشغيل.

## Sommaire

<b>Introduction générale</b> .....	1
<b>Chapitre I : Généralités sur l'identification des appareils électriques domestiques</b> .....	3
I.1) INTRODUCTION .....	3
I.2) CONTEXTE ET OBJECTIF DE L'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES .....	3
I.3) APPROCHES D'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES .....	4
I.4) PRINCIPE DE FONCTIONNEMENT D'UN SYSTEME D'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES (SIAED) .....	6
I.4.1) DETECTION D'EVENEMENTS .....	8
I.4.2) EXTRACTION DES PARAMETRES.....	8
I.5) PHASE D'APPRENTISSAGE ET DE TEST .....	9
I.5.1) PHASE D'APPRENTISSAGE .....	9
I.5.2) PHASE DE TEST .....	10
I.6) ETAT DE L'ART SUR LES METHODES D'IDENTIFICATION IAED.....	10
I.7) CLASSIFICATION GMM (Modèle de Mélange de Gaussiennes).....	12
I.7.1) MODELE DE MELANGE DE GAUSSIENNES (GMM).....	12
I.7.2) DISTRIBUTIONS GAUSSIENNE ET DISTRIBUTION GMM.....	13
I.8) ETAT DE L'ART SUR LES BASES DE DONNEES .....	14
I.9) CONCLUSION.....	14
<b>Chapitre II :Identification des appareils électriques basée sur l'analyse harmonique</b> ....	15
II.1) INTRODUCTION .....	15
II.2) SYSTEME IAED BASE SUR L'EXTRACTION DES PARAMETRES STATISTIQUES COMBINEE AVEC LE CLASSIFICATEUR KNN.....	15
II.2.1) ANALYSE HARMONIQUE .....	16
II.2.2) PARAMETRES STATISTIQUES.....	18
II.3) BASE DE DONNEES PLAID (plug loadappliance identification dataset).....	19
II.4) CONCLUSION .....	20
<b>Chapitre III :Système proposé : implémentation et résultats</b> .....	21
III.1) INTRODUCTION.....	21
III.2) OBJECTIF PRINCIPAL .....	21

III.3) ARCHITECTURE DU SYSTEME PROPOSE.....	22
III.3.1) PREPARATION DE DONNEES.....	22
III.3.2) EXTRACTION DES PARAMETRES STATISTIQUES DES COEFFICIENTS STFS.....	23
III.4) EXPERIENCES ET RESULTATS.....	26
III.4.1) NOMBRE OPTIMAL DE COMPOSANTES GAUSSIENNES DES MODELES GMM .....	26
III.4.2) L'INFLUENCE DE L'ORDRE DES PARAMETRES STATISTIQUES ( PAIR/ IMPAIR).....	26
III.5) CONCLUSION .....	30
<b>Conclusion générale.....</b>	<b>31</b>

## Liste des Figures

Figure I.1 : principe de mesures de l'approche non intrusive NILM [5] .....	5
Figure I.2: Schéma général d'un système NILM. En rouge les trois critères selon lesquels on peut classer les approches NILM [2] .....	5
Figure I.3 : Schéma général d'un système d'identification des appareils électriques domestiques.....	6
Figure I.4 : Etapes de fonctionnement d'un système (SIAED).....	8
Figure II.1 : Schéma synoptique de fonctionnement du système d'identification présenté dans [1] .....	16.
Figure III.1 : Structure générale du système d'identification proposé.....	22

## *Liste des Tableaux*

Tableau I.1 : Etat de l'art sur les algorithmes de classification et les techniques d'extraction de paramètres.....	11
Tableau I.2 : Etat de l'art sur les bases de données [2].....	14
Table II.1 : Résumé sur les appareils de la base PLAID.....	19
Tableau III.1 : Taux de classification en fonction du nombre de gaussiennes.....	26
Tableau III.2 : Influence de l'ordre harmonique sur les performances du système.....	27
Tableau III.3 : Taux de classification TCV et TCS après l'application de la règle de vote...	28
Tableau III.4 : Comparaison des performances entre les classificateurs KNN et GMM.....	29

## *Abréviations*

**SIAED** Système d'identifications des Appareils Electriques Domestiques

**IAED** Identification des Appareils Electriques Domestiques

**KNN** K-Nearest Neighbors

**GMM** Gaussian Mixture Models (Modèles de Mélange de Gaussiennes)

**HMM** Hidden Markov Model (Modèles de Markov cachés)

**NILM** Non-Intrusive Load Monitoring

**ILM** Intrusive Load Monitoring

**SILM** Semi-intrusive Load Monitoring

**SVM** Support Vector Machines

**EM** Expectation-Maximization

**STFS** Short-time Fourier Series

**COOLL** Controlled On/Off Loads Library

**PLAID** Plug Load Appliance Identification Dataset

**fdp** Fonction de densité de probabilité

**TF** transformée de Fourier

**TFD** transformée de Fourier discrète

**FFT** Fast Fourier Transform (Transformée de Fourier Rapide)

**SF** Séries de Fourier

**NBR** Nombre de Paramètres

## *Introduction générale*

Les réseaux intelligents sont des réseaux d'électricité utilisant des technologies informatiques dans le but de contribuer à une adéquation entre production, distribution et consommation d'électricité. Ces réseaux utilisent des compteurs d'électricité intelligents dans des installations domestiques, pour que les fournisseurs puissent mieux gérer l'offre à la demande en temps réel et pour que les clients puissent gérer simplement et précisément leurs consommations électriques. Ces compteurs intègrent généralement une fonction d'identification des appareils électriques domestiques par l'utilisation des méthodes de surveillances des courbes de charges des appareils domestiques [25]. Ces méthodes se divisent en deux catégories de types intrusif et non intrusif.

Généralement, le fonctionnement d'un système d'identification des appareils électriques domestiques (IAED) exige une étape d'extraction de paramètres caractérisant les différentes classes d'appareils ainsi qu'une étape de classification des signaux électriques. La conception d'un tel système se base sur une phase d'apprentissage pour modéliser les classes d'appareils et une phase de test basée sur une étape de classification pour évaluer les performances du système. Ces deux phases exigent une étape d'extraction de paramètres permettant de convertir chaque signal électrique en une séquence de vecteurs de paramètres. Les deux phases exigent ainsi, une base de données à répartir en une base d'apprentissage et une base de test. Différentes bases de données sont utilisées pour l'évaluation des systèmes d'identification des appareils électriques, telles que les bases de données PLAID et COOLL, etc. Dans [9,1], les auteurs ont proposé des systèmes d'identification basés respectivement sur les classificateurs HMM et KNN combinés avec la méthode d'extraction des paramètres harmoniques. Dans [1], le classificateur KNN est combiné avec la stratégie de règle de vote pour classier des signaux du courant électriques. Les performances de ce système d'identification sont évaluées sous la base de données PLAID, qui est constituée de 1074 signaux électriques échantillonnés à une fréquence de 30 KHz.

Cependant, le classificateur KNN est lent du point de vue temps de calcul, si le nombre de vecteurs de références enregistrés durant la phase de test est assez grand. Ainsi, notre travail consiste à proposer un système d'identification des appareils électriques basé sur la classification GMM combinée avec la règle de vote. Ce système utilise les modèles GMM

pour modéliser les classes d'appareils électriques en évitant ainsi l'enregistrement des vecteurs de références durant la phase d'apprentissage.

Notre objectif est de valider le descripteur de paramètres statistiques utilisé dans [1] et d'améliorer la complexité du système en utilisant les modèles GMM avec l'évaluation de performances sous la base de données PLAID.

Notre manuscrit est organisé en trois chapitres :

- Le premier chapitre permet de donner des généralités sur l'identification des appareils électriques domestiques et de décrire son principe de fonctionnement, avec un état de l'art sur les systèmes d'identification des appareils électriques existants.
- Le deuxième chapitre est consacré à la description d'un système IAED, relié à notre travail (related work). Ce système présenté dans [1] est basé sur le classificateur KNN.
- Le troisième chapitre présente le système IAED proposé et son implémentation avec les résultats obtenus à partir des expériences. Une étude comparative entre les performances du système proposé et celles du système présenté dans [1], sera menée.
- Nous terminons finalement notre manuscrit par une conclusion générale et des perspectives.

*Chapitre I*

*Généralités sur l'identification des  
appareils électriques domestiques*

---

## **I.1) INTRODUCTION**

Un compteur électrique est un organe électrique sert à mesurer la quantité d'énergie d'électricité afin de facturer la consommation d'énergie au client. A l'origine ces appareils étaient de conception électromécanique. Ils sont remplacés ensuite par des modèles électroniques. Les nouvelles versions des compteurs électriques sont des compteurs électriques intelligents qui permettent l'identification automatique des appareils électriques domestiques pour connaître leurs intervalles et leurs quantités de consommation de l'électricité. Ces compteurs utilisent des outils d'analyse du signal électrique ainsi que des outils de classification automatique (GMM, KNN, HMM, ...) [1]. Ces compteurs intelligents sont dotés de capacités de communication permettant de renvoyer les informations de consommation au système central du service public.

## **I.2) CONTEXTE ET OBJECTIF DE L'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES**

La nécessité d'un équilibre entre la production et la consommation d'électricité pour des raisons environnementales et économiques, reste une préoccupation quotidienne pour les fournisseurs d'électricité. L'optimisation de la production d'électricité peut passer par des délestages électriques, une facturation adaptée aux usages des clients ou des échanges d'électricité avec d'autres fournisseurs.

Pour les clients, avoir accès aux informations de consommation offre la possibilité d'optimiser la consommation d'électricité et la facture d'énergie. Ces informations peuvent être obtenues en utilisant des systèmes de surveillance non intrusive de charge (NILM) [1]. Ce type de systèmes consiste à désagréger la consommation d'énergie au niveau du compteur principal d'électricité, ce qui nécessite l'identification des appareils électriques domestiques (**IAED**) à partir du signal électrique. Cette tâche d'identification a pour objectif d'identifier la classe d'appareil correspondante au signal électrique d'entrée du système.

L'identification des appareils électriques exige généralement une étape d'extraction de paramètres (caractéristiques) ainsi qu'une étape de classification pour identifier la classe d'appareil correspondante au signal électrique. Plus particulièrement, la méthode d'extraction de paramètres basée sur l'analyse de Fourier, à partir du signal de courant électrique a été largement utilisée dans les systèmes **IAED** [2][1].

Dans les sections suivantes, on donnera des généralités sur les systèmes d'identification des appareils électriques domestiques. Plus particulièrement, on décrira brièvement les systèmes IAED basés sur l'analyse harmonique du signal du courant électrique.

### **I.3) APPROCHES D'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES**

Les méthodes d'identification des appareils électriques consistent à surveiller les courbes de charges afin de définir chaque consommation par usage à partir d'une consommation générale, Où les informations peuvent être obtenues grâce à des systèmes de surveillance de charge qui sont classés en trois approches peuvent être envisagées : approches intrusives, semi-intrusives et intrusives [4].

- 1) Approches intrusives ILM (Intrusive Load Monitoring) : sont basées sur l'utilisation de compteurs à détection distribuée pour prendre en charge le contrôle, la surveillance, les différentes formes d'ILM peuvent être trouvées (Sous-comptage électrique : un compteur d'énergie est utilisé pour chaque appareil, appareils intelligents : qui ont des puces communicantes qui leur permettent d'auto-déclarer leur consommation d'énergie à une unité centrale, sondag eélectrique.
- 2) La surveillance de charge semi-intrusive SILM (Semi-intrusive Load monitoring) est en quelque sorte une version détendue de l'ILM. SILM partage les mêmes inconvénients de l'ILM mais avec moins d'intrusion.
- 3) Non intrusive NILM (Non Intrusive Load Monitoring) Le but du système de surveillance de charge est de classer la consommation d'énergie au niveau du compteur principal d'électricité. Ce qui nécessite l'identification des appareils électriques, en utilisant les caractéristiques du courant, détecte les événements de commutation des appareils, extrait les caractéristiques distinctives et les utilise pour identifier les appareils ou les utilisations finales, également appelée désagrégation de l'énergie. D'autre part, le système a besoin d'algorithmes pour traiter et analyser le signal, tout en fournissant la nature et le nombre de charges qui existent [9].

Les méthodes des NILM doivent être capables de détecter les changements d'état provoqués par une mise en marche ou en arrêt d'une charge, ainsi que de renseigner sur la nature et le nombre des charges en présence [4]. Les grandeurs à mesurer sont les signaux

correspondant à la tension d'alimentation totale  $u(t)$  et le courant total de consommation  $i(t)$  [5] (voir la figure I.1).

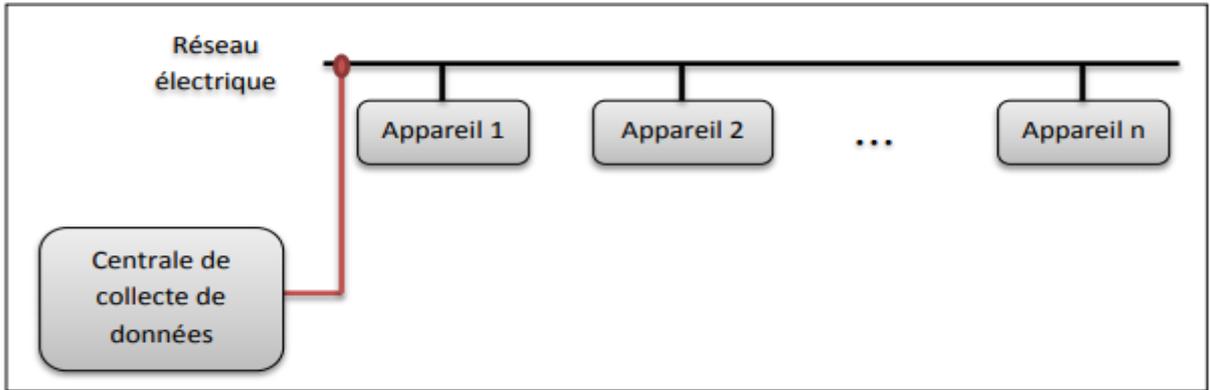


Figure II.1 : principe de mesures de l'approche non intrusive NILM [5].

### Qui, quand et combien ?

Le système NILM essaye de répondre à trois questions principales : qui consomme ? quand ? et combien consomme-t-il ? Savoir qui consomme relève de l'identification du type de consommateur. Savoir quand est-ce qu'il consomme revient à faire une détection du moment de début et de fin de consommation. Et combien, constitue en général le dernier étage d'un système NILM et est conditionné par les étages de détection et d'identification [2].

La Figure I.2 représente le Schéma général d'un système NILM. Les trois critères selon lesquels on peut classer les approches NILM sont indiqués en rouge.

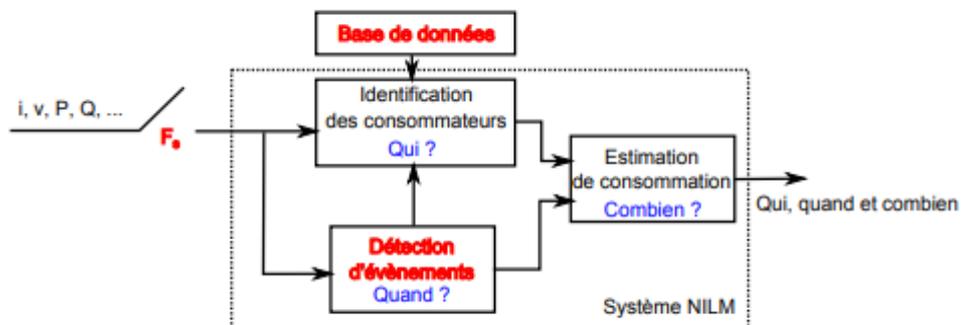


Figure I.2: Schéma général d'un système NILM. En rouge les trois critères selon lesquels on peut classer les approches NILM [2].

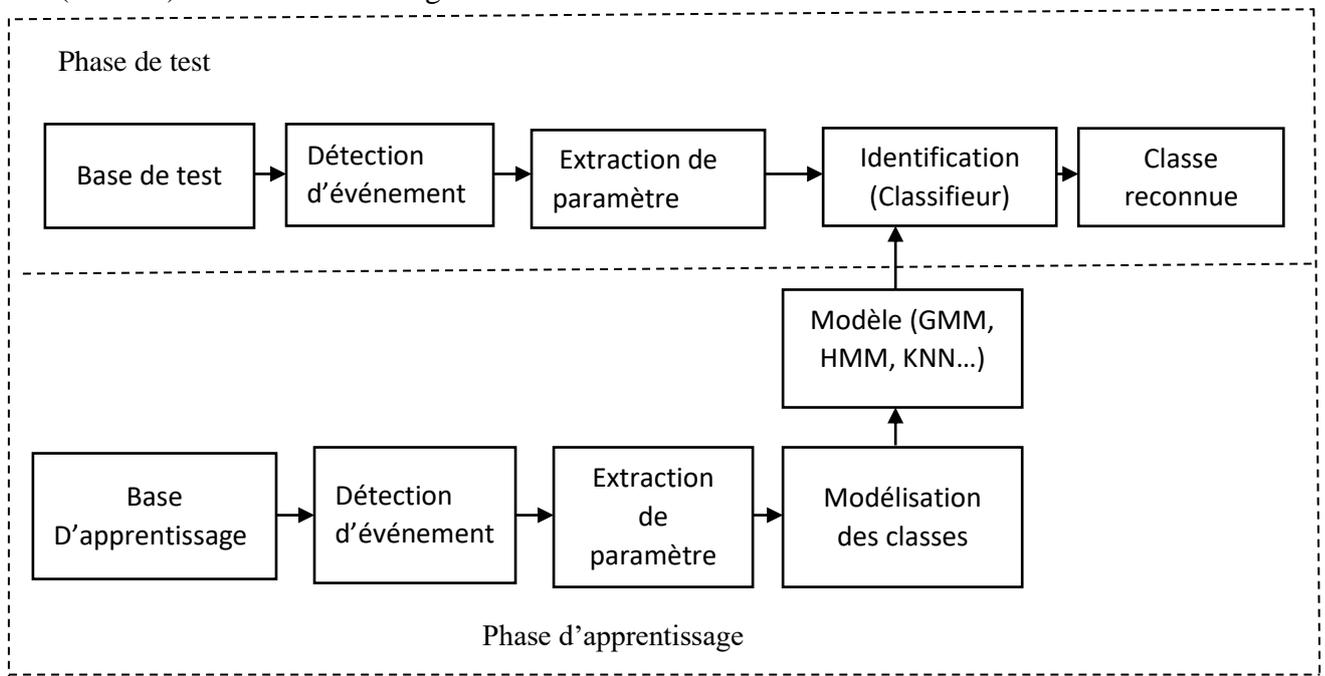
On trouve essentiellement trois catégories possibles pour classer les approches NILM.

- Approche supervisée / non supervisée.
- Approche basse / Haute fréquence d'échantillonnage.
- Approche basée /non basée sur le détecteur d'évènement.

#### **I.4) PRINCIPE DE FONCTIONNEMENT D'UN SYSTEME D'IDENTIFICATION DES APPAREILS ELECTRIQUES DOMESTIQUES (SIAED)**

Le schéma général d'un système d'identifications des appareils électriques domestiques

(SIAED) est illustré sur la figure I.3 :



**Figure I.3 : Schéma général d'un système d'identification des appareils électriques domestiques.**

Un système IAED exige préalablement une étape de détection de l'évènement de la mise en marche de l'appareil à identifier dans le but de savoir l'intervalle d'identification de l'appareil.

La conception d'un système IAED se déroule en une phase d'apprentissage et une phase de test. La première phase a pour objectif de modéliser les différentes classes d'appareils électriques en utilisant des références ou des modèles à entraîner par des algorithmes d'apprentissage standards (GMM, ANN, SVM, ...), appliqués sur des signaux d'entraînement extraits à partir d'une base de données d'apprentissage. La deuxième phase consiste à évaluer les performances du système IAED en utilisant des mesures de performances telles que le

taux d'identification, la complexité du point de vue temps de calcul et d'espace mémoire. Cette phase exige une étape d'identification qui consiste à utiliser les références ou les modèles entraînés durant la phase d'apprentissage, pour classifier chaque signal de la base de test en utilisant des algorithmes de classification standards tels que KNN, SVM, ANN et HMM [1][6][7][8].

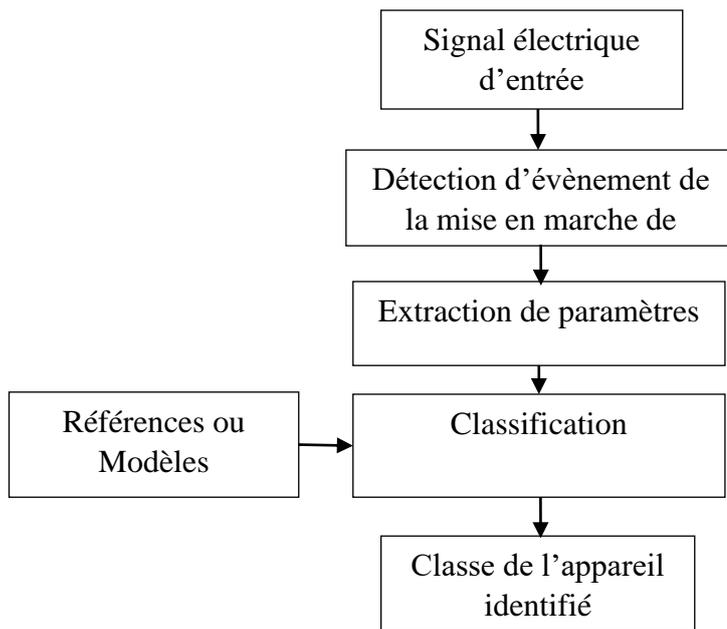
Chacune des phases exige une étape d'extraction de paramètres permettant de convertir chaque signal en une séquence de vecteurs dont chacun est composé d'un ensemble de paramètres extrait sur une fenêtre d'analyse de courte durée [9]

Plusieurs bases de données très connues dans le domaine de recherche scientifique, contenant des signaux du courant et de la tension électriques, sont préparées et destinées pour évaluer les performances des systèmes IAED. Parmi ces bases de données, on peut citer la base PLAID [19] et la base COOLL [10].

Le fonctionnement d'un système IAED pour l'identification de l'appareil électrique domestique à partir du signal électrique d'entrée, se déroule en plusieurs étapes :

- Détection de l'événement de la mise en marche de l'appareil à identifier.
- Extraction de paramètres (caractéristiques) en transformant le signal d'entrée du système en une séquence de vecteurs de paramètres.
- Classification de la séquence de vecteurs des paramètres extraits pour obtenir la classe de l'appareil à identifier, en utilisant les références ou les modèles entraînés durant la phase d'apprentissage.

Le schéma synoptique de fonctionnement d'un système d'identification des appareils électriques est illustré sur la figure I.4.



**Figure I.4 : Etapes de fonctionnement d'un système (SIAED).**

#### **I.4.1) DETECTION D'ÉVÈNEMENTS**

Un évènement est défini par une transition qui survient, sur le courant et/ou la tension, lorsqu'un appareil s'allume (transition ON ou enclenchement) ou s'éteint (transition OFF ou déclenchement). Cette étape consiste à détecter chaque mise en route ou extinction d'appareil qui survient sur le réseau, c'est-à-dire déterminer les instants de transition [5].

#### **I.4.2) EXTRACTION DES PARAMÈTRES**

La phase d'apprentissage et la phase de test exigent chacune une étape d'extraction des paramètres. Généralement, cette étape découpe chaque signal en une séquence de fenêtres d'analyse chevauchées, ensuite elle convertit chaque fenêtre d'analyse en un vecteur composé d'un ensemble de paramètres en appliquant des techniques d'analyse telles que l'analyse de Fourier [2] [1][11] et l'analyse par ondelettes [12].

#### **I.4.3) MÉTHODES DE CLASSIFICATION DES APPAREILS ÉLECTRIQUES**

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains traits descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée [13].

La plupart des algorithmes de classification tentent de trouver un modèle (une fonction mathématique) qui explique la relation entre les données d'entrée et les classes de sortie, Les méthodes de classification de formes peuvent être réparties en deux types. La classification supervisée et la classification non supervisée [13].

Généralement, la plupart des systèmes IEAD se basent sur des algorithmes de classification supervisée tels que KNN, HMM, SVM, etc.

#### **I.4.4) EVALUATION DES PERFORMANCES**

L'évaluation d'un système de classification se base généralement sur un partage de la base de données globale en trois bases :

- Base de d'apprentissage utilisée pour la construction des modèles des classes.
- Base de données de validation utilisée pour ajuster et valider les paramètres des modèles de classes, obtenus durant la phase d'apprentissage.
- Base de données de test utilisée pour évaluer les performances du système de classification.

Dans certains cas, la taille de la base de données ne permet pas d'avoir un partage en trois phase, dans ce cas, on utilise uniquement un partage en deux : phase d'apprentissage et phase de tests [13].

#### **I.5) PHASE D'APPRENTISSAGE ET DE TEST**

##### **I.5.1) PHASE D'APPRENTISSAGE**

L'apprentissage automatique est un domaine dont l'intérêt majeur est le développement des algorithmes permettant à une machine d'apprendre à partir d'un ensemble des données. La motivation originale de ce domaine était de mettre en œuvre des systèmes artificiels intelligents [14].

La classification est considérée comme un exemple d'apprentissage supervisé et de reconnaissance de formes, c'est-à-dire un apprentissage dans lequel un ensemble de formations d'observations correctement identifiées est disponible. La procédure non supervisée correspondante est connue sous le nom de clustering. Elle consiste à regrouper des données en catégories en fonction d'une mesure de la similarité inhérente ou de la distance [4].

L'objectif, dans le cas d'apprentissage supervisé, est déterminé explicitement par la prédiction d'une cible. Dans ce cas, pour chaque exemple d'apprentissage  $x_i$  de  $\Omega$ , une cible (étiquette)  $y_i$  est associée. La tâche d'un algorithme d'apprentissage est alors d'entraîner un modèle qui puisse prédire, pour une entrée  $x$  quelconque, la valeur de la cible  $y$ . La nature de l'ensemble de cibles (noté  $Y$ ) dépendra du type de problèmes à résoudre [14].

Deux types de problèmes fréquents sont les problèmes de classification et de régression. Pour un problème de classification,  $Y$  correspond à un ensemble fini de classes auxquelles peuvent appartenir les différents  $x$ . Par contre pour celui de la régression,  $Y$  correspond à un ensemble de valeurs continues [14].

### **I.5.2) PHASE DE TEST**

La base de test est un ensemble d'exemples ayant les mêmes caractéristiques que ceux de la base d'entraînement et qui sont écartés au départ de l'entraînement pour effectuer les tests. La méthode de prédiction utilisée dépend essentiellement du type d'information prédite c'est-à-dire le type de l'attribut de classe. Si l'attribut est catégoriel ou symbolique (appartient à un ensemble fini), il s'agit de classification. Par contre si cet attribut est continu (numérique) il s'agit d'un problème de régression [15].

### **I.6) ETAT DE L'ART SUR LES METHODES D'IDENTIFICATION IAED**

Un système IAED se base sur des techniques d'extraction de paramètres ainsi que sur des algorithmes de classification. Dans cette section, on présente un état de l'art sur ces techniques et ces algorithmes. Le tableau (I.1) illustre un état de l'art sur les algorithmes de classification supervisée.

Tableau I.1 : Etat de l'art sur les algorithmes de classification et les techniques d'extraction de paramètres.

Références	Le classificateur utilisé	Méthode d'extraction utilisé
[10, 11]	HMM	Analyse de Fourier
[23]	HMM	Analyse par ondelette
[1]	KNN combiné avec la règle de vote	Analyse de Fourier
[12]	KNN combiné avec la méthode des règles de vote	Analyse par ondelette

Dans [9], les auteurs ont appliqué la méthode de classification HMM (hidden Markov model) combinée avec l'analyse de Fourier pour l'identification de 11 appareils à partir des signaux du courant électrique. Les performances du système sont évaluées sous la base de données PLAID, en utilisant le taux de classification comme une mesure de performance. La méthode d'extraction des paramètres basée sur l'analyse de Fourier consiste à convertir chaque signal du courant électrique en une séquence de vecteurs de 249 Coefficients STFS (short-time Fourier series) extraits sur des fenêtres d'analyse chevauchées de durée égale à la période de (1/60 s). De plus, chaque appareil est modélisé par un modèle HMM durant la phase d'apprentissage. Les résultats obtenus ont montré la pertinence des harmoniques impairs avec un nombre réduit de paramètres. Dans [23], les auteurs ont amélioré les performances du système précédent en utilisant l'analyse par ondelette. Dans [1], les auteurs ont proposé un système IAED basé sur l'extraction de paramètres statistiques de coefficients d'harmonique ainsi que le classificateur KNN combiné avec la règle de vote. De plus, ils ont étudié l'étape de sélection de paramètres pour la réduction de la dimensionnalité. Dans [11], les auteurs ont étudié la pertinence et la redondance des paramètres harmoniques ainsi que le nombre optimal de paramètres. Dans [12], les auteurs ont amélioré le système proposé dans [1] en utilisant une représentation compacte de donnée basée sur l'analyse par Ondelettes.

Dans la section suivante, on présentera la classification GMM qui sera utilisée dans notre système proposé (voir chapitre 3). Plus particulièrement, notre système IAED se base sur l'analyse de Fourier et la classification GMM combinée avec la règle de vote.

### **I.7) CLASSIFICATION GMM (Modèle de Mélange de Gaussiennes)**

Les modèles GMM sont des modèles probabilistes qui constituent de nos jours un outil très important dans la modélisation statistique de données multidimensionnelles (vecteur de paramètres, observation, objet). Couramment, ces modèles trouvent leurs applications dans le domaine de classification supervisée. La classification GMM consiste à modéliser chaque classe de données par un modèle statistique GMM durant la phase d'apprentissage. Cette modélisation se base sur l'algorithme de regroupement (clustering) EM (Expectation-Maximization) qui consiste à regrouper l'ensemble de donnée d'une classe en groupes dont chacun est représenté par une fdp gaussienne.

Dans les systèmes de classification supervisée, la modélisation GMM consiste à représenter toute fonction de densité de probabilité (fdp) décrivant l'ensemble de données d'une classe par une somme pondérée de fonctions fdp gaussiennes durant la phase d'apprentissage[3][24].L'étape de classification d'un vecteur de paramètres d'entrée  $V_e$  consiste à trouver le modèle parmi les modèles de classe  $c$  ( $c=1, 2, \dots, N$ ), qui maximise la probabilité à postériori  $P(C=c|V_e)$ , ou  $c$  appartient à l'ensemble de  $N$  classes. Le problème critique dans la modélisation GMM est le choix du nombre de gaussienne optimal à fixer au début de l'algorithme GMM.

Dans notre travail, on modélise les vecteurs de paramètres statistiques des harmoniques de chaque classe d'appareil électrique par un modèle GMM dont le choix du nombre de gaussiennes est déterminé expérimentalement et sera décrit dans le chapitre 3.

#### **I.7.1) MODELE DE MELANGE DE GAUSSIENNES (GMM)**

Un modèle de mélange de gaussienne GMM sert à estimer para métriquement la distribution des variables aléatoires en les modélisant comme une somme de plusieurs gaussiennes (appelées noyaux). Il s'agit alors de déterminer la variance, la moyenne et l'amplitude de chaque gaussienne. Ces paramètres sont optimisés selon un critère de maximum de Vraisemblance pour approcher le plus possible la distribution recherchée. Cette

procédure se fait le plus souvent itérativement via l'algorithme EM (Expectation-Maximization).

### **I.7.2) DISTRIBUTIONS GAUSSIENNE ET DISTRIBUTION GMM**

La distribution gaussienne ou la distribution normale a une courbe en forme de cloche, avec les points de données répartis symétriquement autour de la valeur moyenne.

La fonction de densité de probabilité d'une distribution gaussienne d'une variable aléatoire mono-dimensionnelle est donnée par :

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (I - 1)$$

$\mu$  : est la moyenne.

$\sigma^2$  : est la variance

La fonction de densité de probabilité d'une distribution gaussienne d'une variable aléatoire multidimensionnelle  $x$  de dimension  $d$  est donnée par :

$$f(x \mid \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp \left[ -\frac{1}{2} (x - \mu)' \Sigma^{-1} (x - \mu) \right] \quad (I - 2)$$

Où,  $\mu$  et  $\Sigma$  sont respectivement le vecteur de moyennes et la matrice de covariance.

La fonction de densité de probabilité d'une distribution de probabilité GMM (Gaussian Mixture Model) d'une variable aléatoire multidimensionnelle de dimension  $d$  est donnée comme suit :

$$f(x) = \sum_{i=1}^k \frac{\alpha_i}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp \left( -\frac{1}{2} (x - \mu_i)' \Sigma_i^{-1} (x - \mu_i) \right) \quad (I.3)$$

$k$  : est le nombre de composantes gaussiennes.

$\mu_i$  et  $\Sigma_i$  sont respectivement le vecteur de moyennes et la matrice de covariance de la  $i^{\text{ème}}$  gaussienne,  $\alpha_i$  est le poids de la  $i^{\text{ème}}$  gaussienne vérifiant la condition :  $\sum_{i=1}^k \alpha_i = 1$ .

$|\Sigma_i|$  est le déterminant de la matrice  $\Sigma_i$ .

### **I.8) ETAT DE L'ART SUR LES BASES DE DONNEES**

Pour la conception du système d'identification on utilise une base de données de référence, Le Tableau (I.2) illustre un état de l'art sur les bases de données utilisées pour l'évaluation des systèmes IAED...

Tableau I.2 : Etat de l'art sur les bases de données [2].

Références	Auteurs La base de données	Fréquence	D'échantillonnage
[17]	J. Zico Kolter Matthew J. Johnson	REDD	15KHZ
[18]	Anderson, K., Ocneanu, A., Benitez	BLUED	12KHZ
[19]	Gao J, Giri S, Kara EC, Berges M	PLAID	30KHz
[20]	Gulati	HFED	10KHz à 5MHz
[21]	Kelly, J. and Knottenbelt	UK-DALE	16 KHz
[22]	Kahl, M., Haq, A. U., Kriechbaumer	WHITED	44.1 KHz
[10]	Thomas P et all 2016	COOLL	100 KHz

### **I.9) CONCLUSION**

Dans ce chapitre, nous avons présenté des généralités sur les systèmes d'identification des appareils électriques domestique en décrivant brièvement cette tâche d'identification ainsi que son principe de fonctionnement. On a présenté également les deux phases de conception d'un système IAED. Chaque phase exige une étape d'extraction de paramètres. La première phase appelée phase d'apprentissage a pour objectif de modéliser les différentes classes, alors que la phase de test basée sur une étape de classification permet d'évaluer les performances du système. Finalement, nous avons présenté un état de l'art sur les méthodes de classification et les techniques d'extraction de paramètres. De plus, on a donné un état de l'art sur les bases de données destinées à la conception des systèmes IAED.

## *Chapitre II*

# *Identification des appareils électriques basée sur l'analyse harmonique*

---

## **II.1) INTRODUCTION**

L'identification des appareils électriques domestiques est largement discutée dans plusieurs travaux de recherches [9,12,23,1,25]. Dans [1] les auteurs ont proposé un système d'identification des appareils électriques basé sur l'extraction des paramètres statistiques des coefficients harmoniques du signal de courant électrique, et la classification KNN combinée avec la stratégie de la règle de vote. Les paramètres statistiques sont : la moyenne, l'écart-type, asymétrie, et l'aplatissement). La phase d'apprentissage consiste à enregistrer un grand nombre de vecteurs statistiques de référence représentant les classes d'appareils dans une matrice, afin de les utiliser dans la l'étape de classification des vecteurs statistiques de test. Cette classification se base sur la comparaison de chaque vecteur d'une séquence de test avec l'ensembles des vecteurs de références, exigeant ainsi un grand temps de calcul et un grand d'espace mémoire si le nombre de vecteurs de référence est grand. Ainsi, dans le but de réduire la complexité du système du point de vue de temps de calcul et d'espace mémoire, nous allons utiliser un système IAED de même architecture mais qui se base sur des modèles paramétriques à entraîner dans la phase d'apprentissage et à utiliser dans l'étape d'identification.

Dans notre travail, on s'intéresse à proposer d'utiliser les modèles GMM pour la modélisation des classes d'appareils électrique pour la tâche d'identification IAED. Une étude comparative des performances du système précédent, sera menée dans le but de mieux valider les résultats basés sur ce type de descripteurs statistiques en utilisant le classificateur GMM. Ainsi, dans les sections suivantes on décrira brièvement le principe de fonctionnement du système présenté dans [1], qui peut être considéré comme un travail relié à notre étude (related Work).

## **II.2) SYSTEME IAED BASE SUR L'EXTRACTION DES PARAMETRES STATISTIQUES COMBINEE AVEC LE CLASSIFICATEUR KNN**

Le schéma synoptique de fonctionnement du système proposé dans [1] est illustré sur la Figure II.1.

La conception de ce système se déroule en une phase d'apprentissage et une phase de test. Chacune des phases exige une étape d'extraction des paramètres statistiques permettant de convertir chaque signal du courant électrique en une séquence de vecteurs de paramètres statistiques. Cette étape consiste premièrement à découper le signal en fenêtres d'analyse

chevauchées de 50% et de durée de 16.7 ms (1/60). Chaque fenêtre d'analyse est convertie en un vecteur de 250 coefficients harmoniques STFS en utilisant le module des coefficients de la décomposition en Série de Fourier (DSF). Ensuite, les paramètres statistiques sont estimés sur chaque séquence de 15 vecteurs de paramètres STFS en utilisant les estimateurs de la moyenne et de l'écart type de chaque coefficient STFS. Ainsi, chaque séquence de vecteurs de 250 paramètres STFS, représentant un signal d'entrée, sera convertie en une séquence de vecteurs de 500 paramètres statistiques STFS incluant les moyennes et les écart type. Les résultats présentés en [1], ont montré qu'un nombre de 10 paramètres statistiques est suffisant pour obtenir approximativement les mêmes performances en utilisant tous les paramètres statistiques. Ainsi, dans notre travail, on n'utilise que des vecteurs de 10 paramètres statistiques.

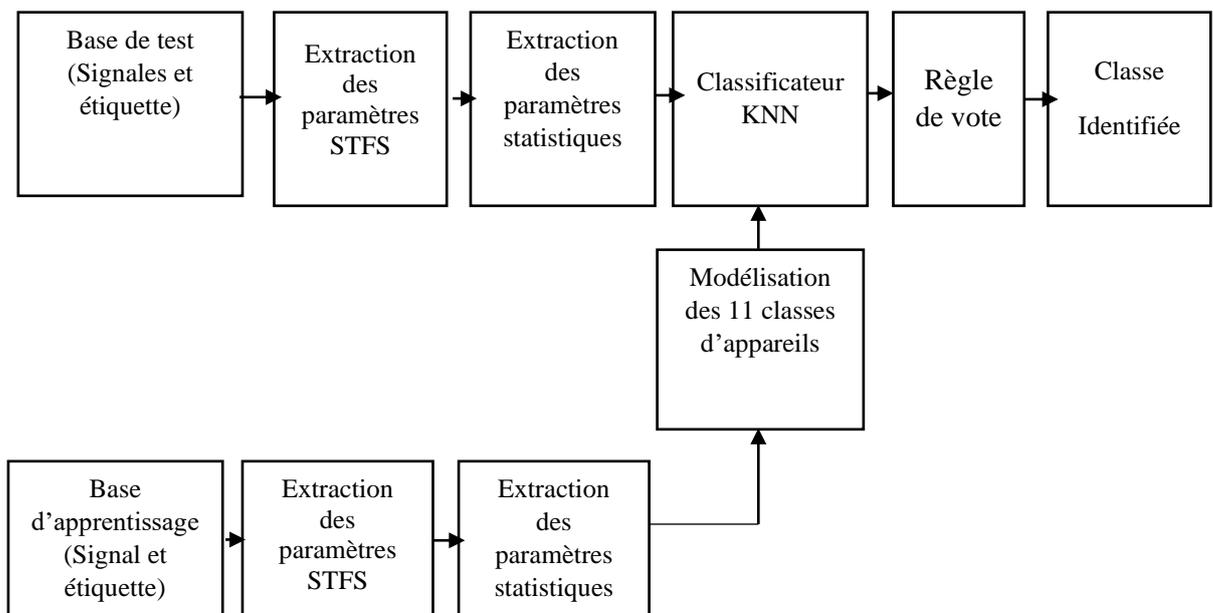


Figure II.1. Schéma synoptique de fonctionnement du système d'identification présenté dans [1].

### II.2.1) ANALYSE HARMONIQUE

La méthode universelle de l'analyse harmoniques est la transformée de Fourier (TF). Dans le cas discret le vecteur du signal de courant  $s[n]$  de longueur  $N$  peut être analysé par la transformée de Fourier discrète (TFD). Dont, sa formule mathématique pour un signal à temps discret  $s[n]$  est :

$$s[k] = \sum_{n=0}^{N-1} s[n] \exp\left(\frac{-j2\pi kn}{L}\right) \quad , \quad k = 0, \dots, N - 1 \text{ (II-1)}$$

$N$  : est la longueur de  $s[n]$ .

$k$  : représente les composantes fréquentielles.

Cette approche d'extraction d'une chaîne de vecteurs est basée sur la division de chaque signal électrique de courant en fenêtres chevauchées par une demi période, puis l'application de la transformée de Fourier discrète (TFD) sur chaque fenêtre.

Cependant, le contenu harmonique de  $s[n]$  en utilisant l'équation (II-1) devient de plus en plus lourd que la valeur de  $N$  est élevée, même en utilisant des algorithmes rapides tel que la Fast Fourier Transforme (FFT). Ce problème devient important, en particulier, lorsque la base de données est grande et qu'un grand nombre de signaux doit être analysé. Étant donné que les signaux de courant électrique sont quasi-périodiques, nous proposons d'exploiter cette caractéristique pour améliorer le temps de calcul en utilisant la Décomposition en Séries de Fourier (DSF) au lieu de la TF. La DSF donne des instances particulières de la TF (certaines fréquences). Cela signifie que pour obtenir la TF d'un signal périodique, il suffit de calculer ses coefficients de DSF. Pour un signal périodique  $s[n]$  de période  $N$ , la décomposition en Séries de Fourier Discrètes (SFD) est donnée par :

$$s[n] = \sum_{k=0}^{N-1} C_k \exp\left(\frac{j2\pi kn}{N}\right) \quad n = 0, \dots, N-1 \quad (II-2)$$

Où

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} s[n] \exp\left(\frac{-j2\pi kn}{N}\right), \quad k = 0, \dots, N-1 \quad (II-3)$$

$C_k$  : sont les coefficients de la série de Fourier représentant les composantes fréquentielles.  $k$  varie de 0 à  $N-1$ , où  $N$  représente la fréquence d'échantillonnage  $f_e$ .

Le nombre de coefficients correspondants aux fréquences normalisées  $f_k = \frac{k}{N}$  est pris égal au nombre d'échantillons de la fenêtre d'analyse  $N$ . La bande fréquentielle maximale du signal est  $[0, \frac{f_e}{2}]$ , et le nombre de points fréquentiels représentant cette bande fréquentielle est égal à  $\frac{N}{2}$ .

Pratiquement, la base de données PLAID utilisée dans [1], contient des signaux électriques échantillonnés sur une fréquence  $f_e$  égale à 30 KHz. Ainsi, le nombre N est égal à 500. De plus, le nombre d'harmoniques ou le nombre de paramètres STFS représentant la bande fréquentielle  $[0, \frac{F_e}{2}]$  est égal à 250.

Généralement, on tient en compte le module de ces coefficients.

Les coefficients obtenus appelés les coefficients STFS. Ce vecteur de coefficients est appelé observation. Donc, chaque signal de courant électrique sera représenté par une séquence de vecteurs de coefficients STFS, regroupée dans une matrice de paramètres harmoniques MH(R,P) qui représente le signal complet. Où P est le nombre des harmoniques et R est le nombre des fenêtres.

## **II.2.2) PARAMETRES STATISTIQUES**

Les types de paramètres statistiques les plus performants sont la moyenne et l'écart type. Les paramètres statistiques sont estimés sur chaque séquence de 15 vecteurs de 250 paramètres STFS.

Les paramètres statistiques de moyenne et d'écart type estimés sur une séquence de vecteurs de paramètres STFS sont donnés comme suit :

- La moyenne  $\mu$  est donnée théoriquement par :

$$\mu = E[MH] \quad (II - 4)$$

Pratiquement, la moyenne de chaque harmonique  $p$  est estimée sur chaque séquence de Rs vecteurs de paramètres STFS comme suit :

$$\mu(p) = \frac{1}{R_s} \sum_{r=1}^{R_s} MH(r, p), p = 1, \dots, P \quad (II - 5)$$

- L'écart Type  $\sigma$  de chaque harmonique  $p$  est estimée théoriquement par :

$$\sigma = \sqrt{E|(MH - \mu)^2|} \quad (II - 6)$$

Cet écart est estimé pratiquement par :

$$\sigma(p) = \sqrt{\frac{1}{R_s} \sum_{r=1}^{R_s} (MH(r, p) - \mu(p))^2} \quad p = 1, \dots, P \quad (II - 7)$$

### **II.2.3) COMBINAISON DE LA REGLE DE VOTE AVEC L'ALGORITHME DE CLASSIFICATION**

La méthode de vote est l'une des techniques les plus efficaces utilisée pour la prise de décision finale en reconnaissance de formes. Dans [1] les auteurs ont proposé d'appliquer la règle de vote sur une séquence d'indices de classification obtenue par l'application de la classification KNN sur chaque vecteur de paramètres statistiques d'une séquence de vecteurs représentant un signal d'entrée. Cette règle consiste à attribuer au signal d'entrée la classe la plus votée parmi les classes appartenant à la séquence d'indices.

Le problème critique dans de la combinaison, est le choix du nombre de vecteurs de la séquence d'estimation des paramètres statistiques. Dans [1], le nombre optimal  $R_s$  obtenu est égal à 15 avec une amélioration des résultats de 2.6% par rapport à l'utilisation d'un seul vecteur de paramètres statistiques représentant complètement le signal d'entrée.

### **II.3) BASE DE DONNEES PLAID (plug load appliance identification dataset)**

PLAID est une base des données publiques des signatures électriques. Ces signatures sont des mesures des courants et tensions électriques. Cette base de données contient 11 types d'appareils dont chaque appareil, trois à six exemples ont été enregistrés. Cette base de données contient en total de 1074 signatures. Chaque signature de la base de données fait quelques secondes de durée. Ces signaux sont échantillonnés à une fréquence de 30 kHz.

L'objectif de PLAID est de fournir une bibliothèque publique pour les mesures d'appareils à haute résolution qui peuvent être intégrées dans des algorithmes d'identification d'appareils existants ou nouveaux [19].

Le tableau II.1 résume les noms et le nombre de chaque type d'appareils constituant la base de données.

Table II.1 : Résumé sur les types d'appareils de la base PLAID.

Type d'appareil	Nombre d'exemples
Lampe fluo-compact	175
Ordinateur portable	172
Sèche cheveux	156
Micro-ondes	139
Ventilateur	115
Lampe à incandescence	114
Climatiseur	66
Réfrigérateur	38
Aspirateur	38
Radiateur	35
Machine à laver	26
Total	1075

Dans [1], La base de données est divisée en une base d'apprentissage et une base de test, de manière d'attribuer le même nombre d'appareils (50% pour l'apprentissage et 50% pour le test). Ce processus de répartition permet d'obtenir 537 signaux pour chaque base de données.

#### **II.4) CONCLUSION**

Dans ce chapitre, nous avons décrit brièvement le système d'identification des appareils électriques basé sur l'extraction de paramètres statistiques et la classification KNN combinée avec la stratégie de règle de vote, présenté dans [1], et qui est considéré comme un travail relié à notre étude (related work). Le principe de fonctionnement de ce système est décrit brièvement en présentant les différentes étapes de conception.

## *Chapitre III*

### *Systeme proposé : implémentation et résultats*

---

### **III.1) INTRODUCTION**

L'identification des appareils électriques domestiques est une tâche très importante, utilisée pour la conception des compteurs intelligents dans le but d'économiser l'énergie. Elle consiste à identifier la classe d'appareil électrique à partir du signal électrique.

Dans ce chapitre nous allons décrire notre système d'identification des Appareils Electriques domestiques (IAED) basée sur l'extraction des paramètres statistiques des coefficients STFS (Short-Time Fourier Series) et le classificateur GMM combiné avec la règle de vote. Ce système se base sur la même architecture du système IAED présenté dans [1], à l'exception de l'utilisation du classificateur GMM au lieu du classificateur KNN. L'objectif de notre travail est d'évaluer les performances de ce système basée sur le classificateur GMM dans le but de valider l'efficacité des paramètres statistique des coefficients STFS pour la tâche d'identification des appareils électriques domestiques. La conception d'un tel système exige de chercher sa bonne configuration en choisissant les paramètres optimaux de chaque étape de fonctionnement du système. La configuration optimale du système peut être obtenue en répondant aux questions suivantes :

- Quel est le meilleur nombre de paramètres STFS permettant d'obtenir les meilleures performances ?
- Quel est l'ordre optimal des coefficients STFS (pair / impair) ?
- Est-ce que l'utilisation de la règle de vote contribue à l'amélioration des performances du système d'identification ?
- Quel est le meilleur nombre de composantes gaussiennes des modèles GMM.
- Le classificateur GMM est-il efficace par rapport au classificateur KNN ?

Plusieurs expériences peuvent être effectuées pour répondre sur ces questions. L'implémentation du système IAED proposé s'effectue en utilisant le langage de programmation MATLAB.

### **III.2) OBJECTIF PRINCIPAL**

Notre travail consiste à mesurer les performances du système IAED basé sur l'extraction de paramètres statistiques des coefficients STFS et la classification GMM combinée avec la stratégie de la règle de vote pour identifier la classe de l'appareil correspondant au signal du

courant électrique d'entrée. La base de données PLAID est utilisée pour évaluer ces performances en utilisant le taux de classification comme mesure de performances.

### III.3) ARCHITECTURE DU SYSTEME PROPOSE

Le schéma synoptique du système IAED basé sur le classificateur GMM est illustré sur la figure (III.1).

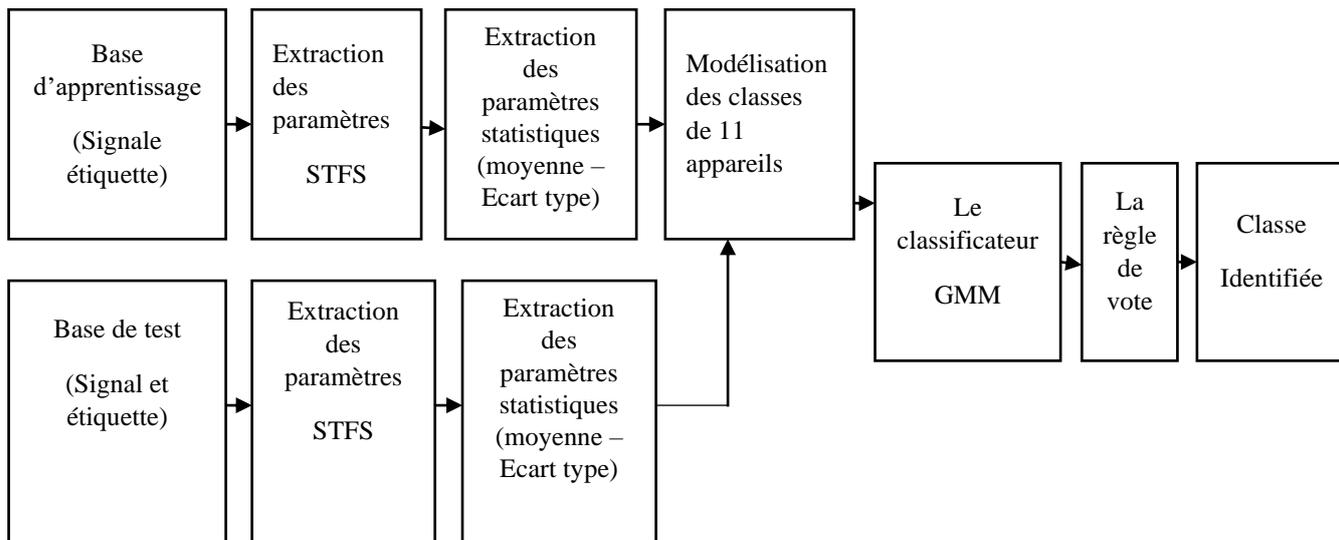


Figure III.1 : Structure générale du système d'identification proposé

Les différentes étapes de conception de ce système sont données comme suit :

- Préparation de données d'apprentissage et de test.
- Extraction des paramètres statistiques des coefficients STFS durant la phase d'apprentissage et la phase de test.
- Modélisation GMM des classes des différents appareils électriques durant la phase d'apprentissage.
- Classification des signaux de test en appliquant l'algorithme de classification GMM ainsi que la règle de vote.

#### III.3.1) PREPARATION DE DONNEES

Dans notre système nous avons choisi la base de données PLAID utilisée dans [1]. Cette base de données contient des signaux de courant et de tension électriques des différents appareils. Dans notre travail, on n'utilise que les signaux du courant électrique. Cette base

contient 1074 signaux du courant électrique appartenant à 11 classes d'appareils. Ces signaux sont échantillonnés sur une fréquence d'échantillonnage de  $f_e$  égale à 30 KHz. L'ensemble de signaux sont répartis en une base d'apprentissage de 537 signaux et une base de test de 537 signaux. Les signaux utilisés sont des signaux obtenus après l'étape de détection de mise en marche des appareils [9].

### **III.3.2) EXTRACTION DES PARAMETRES STATISTIQUES DES COEFFICIENTS STFS**

L'étape d'extraction de paramètres est une étape nécessaire dans la phase d'apprentissage et la phase de test. Elle consiste à convertir le signal du courant électrique en une séquence de vecteurs de paramètres STFS, obtenue par l'application de l'analyse harmonique sur des fenêtres d'analyse chevauchées de 50%. Chaque fenêtre d'analyse est transformée ainsi en un vecteur de 250 coefficients STFS représentant la bande fréquentielle  $[0, f_e/2]$  avec une fréquence fondamentale égale à 60Hz [9].

Les opérations permettant d'effectuer cette extraction sont données comme suit :

- Découpage du signal du courant électrique en une séquence de fenêtres d'analyse chevauchées de 50% et de durée égale à  $1/60$  s (16.7 ms).
- Application de la décomposition en série de Fourier (DSF) sur chaque fenêtre d'analyse pour extraire les 250 coefficients STFS en tenant compte que le module de chaque coefficient.

Dans le but de réduire le temps de calcul et de l'espace mémoire, chaque séquence de vecteurs de coefficients STFS représentant un signal est convertie en une séquence plus réduite de vecteurs de paramètres statistiques estimé chacun sur une sous séquence de 15 vecteurs de paramètres STFS [1] (voir le chapitre 2).

Les résultats obtenus dans [1] ont montré la pertinence des paramètres statistiques estimés à partir de la moyenne et l'écart type des paramètres STFS. Ainsi, chaque sous séquence de 15 vecteurs de paramètres STFS est convertie en un seul vecteur de 500 paramètres statistiques. Les résultats obtenus ont montré également qu'un nombre inférieur à 10 paramètres statistiques est suffisant pour obtenir de bonnes performances. De plus, les premiers paramètres statistiques de rang harmonique impair ont montré leur efficacité par rapport à ceux du rang pair. Ainsi, dans notre travail, on tient en compte des vecteurs de 10

paramètres statistiques répartis en 5 paramètres de type ‘moyenne’ et 5 paramètres de type ‘écart type’.

La question qui se pose : Est-ce que le système proposé donne les mêmes conclusions du système présenté dans [1], en ce qui concerne la pertinence des paramètres statistiques ? Plusieurs expériences seront présentées dans les sections suivantes dans le but de répondre sur cette question.

### **III.3.3) PHASE D’APPRENTISSAGE**

La phase d’apprentissage a pour objectif de modéliser chaque classe d’appareils par un modèle GMM de  $N_g$  composantes gaussiennes. Les différentes étapes de cette phase sont données comme suit :

- Extraction de paramètres statistiques en convertissant chaque signal en une séquence de vecteurs de 10 paramètres statistiques. Cette séquence est enregistrée dans une matrice dont chaque ligne est un vecteur de 10 paramètres statistiques. Cette étape est appliquée sur l’ensemble de 537 signaux d’apprentissage. Les matrices représentant chacune une séquence (ou signal) sont ensuite concaténées verticalement dans une matrice unique. Dans le but garder des informations d’appartenance, la classe et le numéro du signal de chaque vecteur statistique sont enregistrés respectivement dans deux tableaux séparés.
- Modélisation de chaque classe d’appareil par un modèle GMM de type ‘diagonal’ en appliquant la commande Matlab « `fitgmdist` ». Chaque modèle est représenté par un vecteur de poids, un vecteur de moyennes et un vecteur de variances de taille  $N_g$ .

### **III.3.4) PHASE DE TEST**

La phase de test consiste à évaluer les performances du système d’identification en termes de taux de classification. Cette phase exige une étape d’extraction identique à celle de la phase d’apprentissage. Ainsi, l’ensemble des signaux de test seront représentés par une matrice unique contenant les vecteurs de paramètres statistiques de test. La deuxième étape consiste à classer chaque signal en suivant les opérations suivantes :

- Classification de chaque vecteur statistique de test en utilisant l’algorithme de classification GMM. La commande Matlab « `pdf` » est appliquée sur chaque vecteur pour calculer la valeur de vraisemblance correspondant à un modèle GMM. Ainsi, le

résultat de classification de chaque séquence de vecteurs statistiques correspondant à un signal de test, est une séquence d'indices de classes.

- Application de la règle de vote sur chaque séquence d'indices de classes obtenue dans l'étape précédente, dans le but d'attribuer au signal la classe d'appareil la plus votée.

### **III.3.5) MESURE DE PERFORMANCES DU SYSTEME IAED PROPOSE**

Les performances du système d'identification proposé sont évaluées globalement en termes de taux de classification TCV des vecteurs statistiques de la base de test.

Le taux de classification TCV est défini par la relation suivante :

$$TCV = \frac{Nv_c}{Nv_T} \cdot 100 \quad (III-1)$$

- $Nv_T$  : est le nombre global des vecteurs de paramètres statistiques de tous les signaux de la base de test.
- $Nv_c$  : est le nombre de vecteurs de paramètres correctement classifiés.

Le taux TCV donne une évaluation globale des performances de classification sans tenir compte l'appartenance des vecteurs aux signaux. Réellement, l'objectif est de classer chaque signal pour identifier sa classe d'appareil. Ainsi, les performances d'identification sont évaluées en utilisant le taux de classification des signaux TCS, donné par :

$$TCS = \frac{Ns_c}{Ns_T} \cdot 100 \quad (III - 2)$$

- $Ns_T$  : est le nombre total des signaux de la base de test.
- $Ns_c$  : est le nombre de signaux correctement identifiés.

Le nombre des signaux de système proposé de test est égal à 537 signaux et le nombre de classes considérées correspond aux 11 types d'appareils.

### **III.4) EXPERIENCES ET RESULTATS**

Dans cette partie, nous présentons les différentes expériences que nous avons menées pour évaluer les performances du système proposé. Ces expériences nous permettront de répondre sur les questions posées dans la section III-1.

#### **III.4.1) NOMBRE OPTIMAL DE COMPOSANTES GAUSSIENNES DES MODELES GMM**

Le problème critique de la modélisation GMM est le choix du nombre de composantes gaussiennes. Ce choix peut être obtenu en cherchant expérimentalement le nombre de composantes permettant d'avoir les meilleures performances d'identification d'un SIAED. Ainsi, cette expérience a pour objectif de déterminer le nombre optimal des composantes gaussiennes des modèles GMM permettant d'obtenir le meilleur taux de classification GMM des vecteurs statistiques. Ainsi, nous utilisons le taux TCV comme mesure de performances. Le nombre de paramètres statistiques considérés dans cette expérience est égal à dix (5 paramètres de type 'moyenne' et 5 paramètres de type 'écart-type' [1]).

Tableau III.1 : Taux de classification en fonction du nombre de gaussiennes

Nombre de Gaussienne	1	2	4	8
TCV	38.36	53.44	67.41	71.32

Les résultats obtenus à partir du tableau III-1, montrent que le meilleur taux de classification des vecteurs TCV obtenu est égal à 71.32%, en prenant un nombre de gaussiennes égal à 8. Dans les expériences suivantes, on choisira ce nombre de gaussiennes.

#### **III.4.2) L'INFLUENCE DE L'ORDRE DES PARAMETRES STATISTIQUES ( PAIR/ IMPAIR)**

Cette expérience permet d'étudier l'influence de l'ordre (impair/pair) des paramètres statistiques sur les performances du système IAED.

Le tableau III.2 montre le taux de classification TCV obtenu pour les 10 premiers paramètres statistiques avec les trois types d'ordre harmonique, à savoir : pair, impair et tout (pair et impair).

Tableau III.2 : Influence de l'ordre harmonique sur les performances du système

L'ordre harmonique des paramètres statistiques	Taux de Classification TCV %
Pair	<b>43.94</b>
Impair	<b>75.05</b>
Pair et impair (tout)	<b>71.32</b>

Les résultats montrent clairement la pertinence des paramètres statistiques STFS d'ordre impair avec un taux de classification TCV de 75.05%. Les paramètres statistiques STFS d'ordre pair donnent les mauvaises performances avec un taux classification égal à 43.94%. Ces résultats confirment ceux obtenus dans [1] qui ont montré également la pertinence des paramètres statistiques d'ordre impair.

### **III.4.3) COMBINAISON DE LA CLASSIFICATION GMM AVEC LA REGLE DE VOTE**

Dans les expériences précédentes, nous avons évalué les performances du classificateur GMM en utilisant le taux TCV comme mesure de performances. Dans cette expérience, nous appliquons la classification GMM combinée avec la stratégie de règle de vote afin de donner une décision sur la classe de l'appareil correspondant au signal d'entrée. Cette combinaison permet de classer chaque signal d'entrée en une classe d'appareils. Ainsi, dans cette expérience, nous utilisons le taux de classification de signaux TCS comme mesure de performances pour la tâche d'identification des appareils électriques.

Le tableau suivant montre le taux TCV et le taux TCS en fonction du nombre de paramètres statistiques. Le nombre de paramètres varie de 1 à 10.

Tableau III.3 : Taux de classification TCV et TCS après l'application de la règle de vote

Nombre de Paramètres STFS	TCV	TCS
1	36.12	43.94
2	54.18	67.22
3	68.52	87.33
4	68.15	87.33
5	71.32	89.75
6	68.71	87.33
7	70.39	88.64
8	69.27	86.21
9	71.69	88.08
10	71.32	87.89

A partir de ce tableau, on peut remarquer les points suivants :

- L'utilisation des 5 premiers paramètres donne le même taux TCV de 71.32% avec l'utilisation de 10 paramètres et un meilleur taux TCS de 89.75%.
- Les mesures de TCS présentent de grandes valeurs par rapport aux mesures de TCV, justifiant l'efficacité de la règle de vote pour cette tâche d'identification.
- A chaque ajout d'un paramètre d'ordre impair, il y a souvent une amélioration des taux TCV et TCS.

#### **III.4.4) COMPARAISON DES PERFORMANCES DES CLASSIFICATEUR KNN et GMM**

Cette expérience a pour objectif de comparer les performances du système IAED basé sur le classificateur KNN avec celles du système basé sur le classificateur GMM. Cette comparaison tient en compte des mesures de performances du point de vue taux de classification TCS ainsi que le temps d'exécution de l'étape de classification. Cette

expérience est effectuée en utilisant un PC portable I3 de 6 Go de RAM d'une horloge de 2 GHz.

Le tableau III.4 montre le taux TCS et le temps d'exécution  $T_e$  des deux classificateurs KNN et GMM combinés avec la règle de vote.

Tableau III.4 : Comparaison des performances entre les classificateurs KNN et GMM

NBR \ Classificateur	KNN		GMM	
	TCS (%)	Te [s]	TCS (%)	Te [s]
1	66.48	0.255	43.94	0.125
2	83.79	0.261	67.22	0.129
3	94.41	0.268	87.33	0.131
4	94.59	0.275	87.33	0.135
5	95.71	0.284	89.75	0.147
6	95.71	0.289	87.33	0.150
7	95.90	0.292	88.64	0.160
8	95.90	0.295	86.21	0.171
9	95.53	0.301	88.08	0.180
10	95.53	0.312	87.89	0.192

Les résultats de ce tableau nous permettent de remarquer les points suivants :

- Le classificateur KNN donne les meilleurs taux de classification par rapport au classificateur GMM. Ce classificateur donne au taux TCS maximal de 95.90% en utilisant 7 paramètres. Alors que le classificateur GMM donne son taux TCS maximal de 89.75% en utilisant 5 paramètres. Le résultat faible du classificateur GMM par rapport au classificateur KNN peut être justifié par la limitation du nombre de vecteurs statistiques de la base d'apprentissage.
- Le classificateur KNN est plus lent par rapport au classificateur GMM justifiant l'efficacité de ce dernier du point de vue complexité du système IAED.

Ces résultats montrent l'efficacité du classificateur KNN du point de vue taux de classification et l'efficacité du classificateur GMM du point de vue complexité du système. La réalisation d'un compromis entre le taux de classification et la complexité dépend du cahier de charges de conception d'un système IAED.

### **III.5) CONCLUSION**

Dans ce chapitre nous avons présenté le système proposé IAED ainsi que les différentes expériences menées pour répondre sur des questions posées dans la conception du système. Les résultats obtenus durant ces expériences, nous ont permis de choisir le nombre de gaussiennes des modèles GMM, de confirmer la pertinence des paramètres statistiques d'ordre harmonique impair et de montrer l'efficacité de la stratégie de la règle de vote. De plus, une étude comparative entre le classificateur KNN et GMM a montré leurs avantages et leurs inconvénients.

## *Conclusion générale*

L'identification des appareils électriques domestiques est une fonction primordiale intégrée dans les compteurs intelligents d'électricité. Cette tâche consiste à identifier la classe d'appareil en état de marche à partir du signal électrique. Deux catégories des méthodes d'identification des appareils électriques peuvent être envisagées, à savoir : l'approche NILM (non intrusive) et l'approche ILM (intrusive).

Plusieurs travaux de recherches sur l'identification des appareils électriques domestiques (IAED) appliquent l'analyse harmonique pour l'extraction des paramètres STFS combiné avec des classificateurs tels que : HMM, KNN, ANN et SVM. Dans un travail publié récemment, les auteurs ont conçu un système IAED basé sur l'extraction des paramètres statistiques des coefficients harmonique STFS, associée au classificateur KNN combiné avec la stratégie de la règle de vote. Ce type de classificateur performant du point de vue précision exige plus de temps de calcul si le nombre de vecteurs de paramètres enregistrés durant la phase d'apprentissage est assez grand. Ainsi, dans notre travail, on s'intéresse à proposer l'utilisation des modèles GMM dans le travail précédent dans le but de valider l'utilisation des paramètres statistiques et également de réduire la complexité du système.

Dans le but d'atteindre nos objectifs, plusieurs expériences ont été effectuées pour répondre sur des questions posées durant la conception d'un tel système IAED. Les résultats obtenus nous ont montrés l'efficacité des modèles GMM en termes de complexité évaluée par mesure du temps d'exécution de l'étape de classification, ainsi que la pertinence des paramètres statistiques. Une étude comparative nous a montrée l'efficacité du classificateur KNN du point vue taux de classification et également l'efficacité du classificateur GMM du point de vue temps d'exécution sans détérioration considérable du taux de classification.

Finalement, plusieurs perspectives peuvent être envisagées pour améliorer les performances d'un tel système IAED :

- Utilisation d'une base de données plus grande dans le but de bien estimer les paramètres des modèles probabilistes GMM.
- Utilisation d'une autre analyse telle que l'analyse par ondelette dans le but d'extraire une représentation plus compacte des vecteurs statistiques.

## References

- [1] Fateh Ghazali, Abdenour Hacine-Gharbi, Philippe Ravier, Tayeb Mohamadi, Papier « Extraction and selection of statistical harmonics features for electrical appliances identification using k-NN classifier combined with voting rules method » Turkish Journal of Electrical Engineering & Computer Sciences, (2019) 27: 2980 – 2997.
- [2] Mohamed Nait Meziane. « Identification d'appareils électriques par analyse des courants de mise en marche ». Thèse, Spécialité : Traitement du signal et de l'image. Université d'Orléans, 9/12/2016. France.
- [3] Kim, K., Chalida bhongse, T.H., Harwood, D., Davis, L.: « Real-time foreground background segmentation using code book model. Real-time Imaging 11 », (2005).
- [4] Hala Najmeddine. « Méthode d'identification et de classification de la consommation d'énergie par usages en vue de l'intégration dans un compteur d'énergie électrique ». Thèse, Spécialité : Electromagnétisme. Université Blaise Pascal - Clermont-Ferrand II, 9/12/2009. France.
- [5] Nabil Amirach « Détection d'évènements simples à partir de mesure sur courant alternatif » thèse de doctorat, Spécialité : Traitement du Signal, Université de Toulon, France. 10 Juin 2015.
- [6] Mittelsdorf M, Hüwel A, Klingenberg T, Sonnenschein M. « Submeter based training of multi-class support vector machines for appliance recognition in home electricity consumption data ». In: Proceedings of Smart Greens; Aachen, Germany; 2013. pp. 151-158.
- [7] De Paiva Penha D, Castro ARG. «Convolutional neural network applied to the identification of residential equipment in nonintrusive load monitoring systems ». In: 3rd International Conference on Artificial Intelligence and Applications; Chennai, India; 2017. pp. 11–21.
- [8] Ridi A, Hennebert J. « Hidden Markov models for NILM appliance identification». Procedia Computer Science 2014; 32: 1010–1015.
- [9] Nait-Meziane, M., Hacine-Gharbi, A., Ravier, P., Lamarque, G., Le Bunetel, J-C. and Raingeaud, Y. « HMM-based transient and steady-state current signals modeling for electrical

appliances identification »', Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM), pp.670–677. (2016)

[10] Thomas. Picon, Mohamed Nait-Meziane, Philippe Ravier, Guy Lamarque, Clarisse Novello, Jean-Charles Le Bunetel and Yves Raingeau, « COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification », arXiv:1611.05803v1 [cs. OH] 17 Nov 201.

[11] Abdenour Hacine-Gharbi, Philippe Ravier and Mohamed Nait-Meziane, « Relevant harmonics selection based on mutual information for electrical appliances identification », Int. J. Journal of Computer Applications in Technology, Vol. 62, No. 2, 2020,

[12] Fateh Ghazali, Abdenour Hacine-Gharbi, Philippe Ravier « Statistical features extraction based on the discrete wavelet transform for electrical appliances identification » international conference on intelligent system and pattern recognition conference ISPR '20, October 16–18, 2020, Virtual Event, Tunisia

[13] Application des algorithmes d'apprentissage automatique pour la détection de défauts de roulements sur les machines tournantes dans le cadre de l'Industrie 4.0, WendBenedo Arnaud Bienvenue ZOUNGRANA, 2020.Québec, Canada.

[14] HASSAN CHOUAIB « Sélection de caractéristiques méthodes et applications » thèse de doctorat spécialité : informatique, université paris Descartes, 08/07/2011.

[15] Site web : [http://www.abdelhamid-djeffal.net/index.php?p=1\\_32\\_Caneva](http://www.abdelhamid-djeffal.net/index.php?p=1_32_Caneva) Abdelhamid Djeffal « classification-concepts de base ». Concepts de base. Informatique Décisionnelle & Multimédia. 2015-2016.

[17] Leeb S. « A conjoint pattern recognition approach to non-intrusive load monitoring ». PhD, Massachusetts Institute of Technology, Cambridge, MA, USA, 1993.

[18] Anderson, K., Ocneanu, A., Benitez, D., Carlson, D., Rowe, A., and Berges, « MBLUED: à fully labeled public dataset for Event-Based Non-Intrusive load monitoring research ». In Proceedings of the and KDD Workshop on Data Mining Applications in Sustainability (Sust KDD), Beijing, China. (2012a).

- [19] Gao J, Giri S, Kara EC, Berges M. « Plaid: A public dataset of high-resolution electrical appliance measurements for load identification research: demo abstract ». In: Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings; New York, NY, USA; 2014. pp. 198–199.
- [20] Gulati, M., Ram, S. S., and Singh, A. « An in depth study into using emi signatures for appliance identification ». In Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, pages 70–79. ACM. (2014).
- [21] Kelly, J. and Knottenbelt, W. The UK-DALE dataset, « domestic appliance-level electricity demand and whole-house demand from five UK homes ». Scientific Data, 2(150007). (2015).
- [22] Kahl, M., Haq, A. U., Kriechbaumer, T., and Jacobsen, H.-A. « In Workshop on Non-Intrusive Load Monitoring (NILM) », Whited-a world wide household and industry transient energy data set 2016 Proceedings of the 3rd International. (2016).
- [23] Hacine-Gharbi A, Ravier P. « Wavelet cepstral coefficients for electrical appliances identification using hidden Markov models». In: 7th International Conference on Pattern Recognition Applications and Methods; Funchal, Portugal; 2018. pp. 541-549.
- [24] Shimada, A., Taniguchi, R. « Hybrid background model using Spatial-Temporal LBP ». Advanced Video and Signal Based Surveillance (2009).
- [25] Cédric Clastres. Les réseaux intelligents « régulation, investissement et gestion de la demande électrique. Ffhalshs-00539818 » thèse de doctorat, HAL Id : halshs-00539818. 2010, 28 p.