

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTÈRE 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'électronique*

## ***Mémoire***

*Présenté pour obtenir*

**LE DIPLOME DE MASTER**

**FILIERE : ELECTRONIQUE**

**Spécialité : Électronique Industrielle (MCIL)**

Par

- **BOUREGHIDA ZAKARIA**
- **SOUAYAH ABDESSETAR**

*Intitulé*

*Identification des appareils électriques basée sur les modèles HMM*

*Soutenu le : .....*

***Devant le Jury composé de :***

<b><i>Nom &amp; Prénom</i></b>	<b><i>Grade</i></b>	<b><i>Qualité</i></b>	<b><i>Etablissement</i></b>
<b><i>Pr. Messali Zoubeida</i></b>	<b><i>Professeur</i></b>	<b><i>Président</i></b>	<b><i>Univ-BBA</i></b>
<b><i>Dr. Hacine Gharbi Abdenour</i></b>	<b><i>MCA</i></b>	<b><i>Encadreur</i></b>	<b><i>Univ-BBA</i></b>
<b><i>Mme. Laouamri Asma</i></b>	<b><i>MAA</i></b>	<b><i>Examineur</i></b>	<b><i>Univ-BBA</i></b>
<b><i>Mr. Ghezzali Fateh</i></b>	<b><i>Dr</i></b>	<b><i>Co-encadreur</i></b>	<b><i>Univ-BBA</i></b>

*Année Universitaire 2022/2023*

# Remerciement

Louange à Dieu Tout-Puissant, qui nous a guidés sur le chemin de la réalisation de ce modeste travail.

Nous remercions sincèrement notre encadreur, le Dr. Hacine Gharbi Abdennour pour son suivi attentif. Sa disponibilité, son aide pratique et sa bienveillance ont été d'une aide précieuse.

Nous exprimons également notre gratitude envers M. Ghazali Fateh, notre co-encadreur, pour son assistance précieuse et son intérêt manifeste pour notre travail.

Nos remerciements vont également aux membres du jury pour leur temps et expertise dans l'évaluation de notre travail.

Nous sommes reconnaissants envers nos familles pour leur soutien inconditionnel tout au long de notre parcours.

Enfin, Nous sommes extrêmement reconnaissants envers tous les enseignants et étudiants département de l'électronique de l'université Mohamed el Bachir el Ibrahim.

# *Dédicace*

*A ma très chère mère*

*Quoi que je fasse ou que je dise, je ne saurai point te remercier*

*Comme il se doit. Ton affection me couvre, ta bienveillance me*

*Guide et ta présence à mes côtés a toujours été ma source de force*

*Pour affronter les différents obstacles.*

*A mon très cher père*

*Tu as toujours été à mes côtés pour me soutenir et m'encourager.*

*Que ce travail traduit ma gratitude et mon affection.*

*A mon très chers frère et sœur*

*Puisse Dieu vous donne santé, bonheur, courage et surtout*

*Réussite.*

# *Dédicace*

*Je tiens C'est avec grande plaisir que je dédie ce modeste travail :*

*A ma mère Source inépuisable de tendresse,*

*de patience et de sacrifice Ta prière et ta Bénédiction m'ont été d'un grand*

*secours tout au long de ma vie. Puisse Dieu tout puissant te préserver et*

*t'accorder santé, longue vie et Bonheur. . A mes chers frères et leurs enfants,*

*source de joie et de bonheur A toute ma famille, source d'espoir et*

*de motivation À Le petit enfant Lokman*

*À tous mes amis de promotion de 5 eme année MCI.L électronique,*

*A vous cher lecteur*

**Abdesetar**

# Table des matières

<i>Introduction générale</i> .....	1
<b><i>Chapitre I : Généralités sur les systèmes d'identification des appareils électriques domestiques IAED</i></b>	
<i>I.1 Introduction</i> .....	2
<i>I.2 Objectifs D'un système IAED</i> .....	2
<i>I.3 Méthodes d'identification des appareils électriques</i> .....	3
I.3.1 L'approche ILM.....	3
I.3.2 L'approche NILM.....	4
<i>I.4 L'impact de l'utilisation de NILM</i> .....	4
<i>I.5 Principe de fonctionnement d'un système IAED</i> .....	5
I.5.1 Détection des événements : .....	6
I.5.2 Analyse du signal et extraction des caractéristiques : .....	6
I.5.3 Identification des appareils électriques : .....	7
<i>I.6 État de l'art sur les bases de données</i> .....	7
<i>I.7 État de l'art sur les classificateurs</i> .....	8
I.7.1 Les diffèrent méthode de classification : .....	8
I.7.2 Comparaison entre les algorithmes de classification :.....	9
<i>I.8 Conclusion</i> .....	10
<b><i>Chapitre II Système d'identification des appareils électriques basé sur les modèles HMM</i></b>	
<i>II.1 Introduction</i> .....	12
<i>II.2 Travaux liés (related work)</i> .....	12
<i>II.3 Architectures du système proposé</i> .....	14
II.3.1 Etapes de fonctionnement du système IAED .....	14
II.3.2 Base de données COOLL .....	15
II.3.3 Prétraitements de base de données .....	17
II.3.4 Fenêtrages.....	18
II.3.5 Analyse de Fourier à court terme : .....	18
II.3.6 Les modèles de Markov.....	19
II.3.7 Qu'est-ce qu'une chaîne de Markov ? .....	20
II.3.8 Hidden Markov models (HMM).....	20
II.3.9 Les modèles de mélange gaussien (GMM) : .....	22

II.3.10	Fonctionnalités d'un HMM .....	22
II.4	<i>Conclusion</i> .....	23
<b>Chapitre III Implémentation et résultats</b>		
III.1	<i>Introduction</i> .....	23
III.2	<i>Implémentation du système d'identification</i> .....	24
III.2.1	Préparation des données : .....	24
III.2.2	Etiquetages : .....	25
III.2.3	Apprentissage des modèles HMM.....	26
III.2.3.1	L'initialisation des modèles HMM .....	26
III.2.3.2	La réestimation des modèles HMM :.....	28
III.2.3.3	L'augmentation de nombre de gaussiennes : .....	32
III.2.3.4	Définition du dictionnaire et du Modèle du Langage : .....	33
III.2.4	Classification .....	34
III.3	<i>Critères d'évaluation</i> .....	37
III.4	<i>Expérience et résultats</i> .....	38
III.4.1	Expérience 1 .....	38
III.4.2	Expérience 2 .....	38
III.4.3	Expérience 3 .....	39
III.4.4	Expérience 4 .....	40
III.5	<i>Conclusion</i> .....	41
<b>Conclusion Générale</b> .....		43

# List des Figures

Figure I 1 : Les trois étapes du l'approches NILM [8] .....	4
Figure I 2 : Analyse de l'impact de l'utilisation de NILM [10] .....	5
Figure I 3 : Schéma synoptique général du processus d'identification IAED [11].....	6
Figure II 1 : Le schéma synoptique qui représentant le principe de fonctionnement d'un système d'identification.....	12
Figure II 2 : : Le schéma synoptique représentant le principe de fonctionnement du système d'identification implémenté évalué sous la base de données COOLL.....	17
Figure II 3 : Signal de courant avant le prétraitement .....	17
Figure II 4 : Signal de courant après le prétraitement .....	17
Figure II 5 : Topologie graphique d'un HMM [20].....	21
Figure III 1 : Schéma synoptique de notre système IAED .....	24
Figure III 2 : Exemple d'un fichier label.....	25
Figure III 3 : proto a cinq états et vecteur de 250 Paramètres .....	27
Figure III 4 : fichier de configuration `config' .....	27
Figure III 5 : fichier trainlist .....	27
Figure III 6 : les fichiers résultant après l'exécutaient de la command HCompV .....	28
Figure III 7: L'initialisation des modèles HMMs par l'outil HCompV .....	28
Figure III 8 : Illustration du fichier "liste_models" .....	30
Figure III 9 : : Illustration s du fichier trainmlf .....	30
Figure III 10 : Illustration du fichier macro.....	31
Figure III 11 : Illustration du fichier hmmdef .....	31
Figure III 12 : L'entraînement des modèles HMMs par l'outil HERest. ....	31
Figure III 13 : Illustration du fichier hed1 .....	32
Figure III 14 : Décodage par l'outil HVite. ....	33
Figure III 15 : Illustration du fichier gramme.....	33
Figure III 16 : compiler une grammaire dans une structure de réseau en utilisons Hparse.....	34
Figure III 17 : Illustration du fichier rec.mlf .....	35
Figure III 18 : Illustration du fichier dictionnaire.....	35
Figure III 19 : Décodage par l'outil HVite .....	36
Figure III 20 : Évaluation des résultats par l'outil HResults .....	37
Figure III 21 : Résultat de la première expérience .....	38
Figure III 22 : Taux d'identification en fonction du nombre du coefficients harmoniques .....	39
Figure III 23 : Taux d'identification en fonction du nombre de coefficients harmoniques du rang impair.....	40
Figure III 24 : Taux TI en fonction du nombre de paramètres harmoniques de rang pair .....	41

## **List des tableaux**

Tableau I - 1 : Comparaison entre les différentes bases de données [14] .....	8
Tableau I - 2 : Comparaison entre les différents algorithmes de classification [12] .....	9
Tableau II - 1 : Résumé de la base de données COOLL [18].....	16
Tableau III - 1 : Taux TI pour différentes valeurs de Nbr_e et Nbr_eg. ....	38

## Abstract

Growing demand for efficient energy management has required the implementation of smart meters to monitor and optimize energy consumption. In particular, real-time identification of electrical appliances is an indispensable feature. In our work, we have implemented an appliance identification system (IAED) based on harmonic analysis combined with the HMM (Hidden Markov Models) classifier. The performance of this system is evaluated on the COOLL database, using the classification rate of electrical current signals as a performance measure. The results show a maximum classification rate of **98.81%** obtained using 5-state HMM models, each one associated with a 4-Gaussian GMM model.

## Résumé

La demande croissante de la gestion efficace de l'énergie a nécessité la mise en place des compteurs intelligents permettant de surveiller et d'optimiser la consommation d'énergie. En particulier, l'identification en temps réel des appareils électriques est une caractéristique indispensable. Dans notre travail, nous avons implémenté un système d'identification des appareils électriques domestiques (IAED) basé sur l'analyse harmonique combinée avec le classificateur HMM (Hidden Markov Models). Les performances de ce système sont évaluées sur la base de données COOLL en utilisant le taux de classification des signaux du courant électrique comme mesure de performances. Les résultats nous ont montrés un taux de classification maximal de **98,81%** obtenu avec des Modèles HMM de 7 états associés chacun à un modèle GMM de 4 gaussiennes.

## ملخص

نظراً لتزايد الطلب على الإدارة الفعالة للطاقة، أصبح تركيب العدادات الذكية ضرورياً بحيث تلعب هذه العدادات دوراً رئيسياً في مراقبة وتحسين استهلاك الطاقة. على وجه الخصوص، يعد التعرف على الأجهزة الكهربائية في الوقت الفعلي ميزة لا غنى عنها. في عملنا، قمنا بتطبيق نظام تحديد الأجهزة (IAED) بناءً على التحليل التوافقي جنباً إلى جنب مع مصنف HMM (نماذج ماركوف المخفية). يتم تقييم أداء هذا النظام في قاعدة بيانات COOLL، باستخدام معدل تصنيف إشارات التيار الكهربائي كمقياس أداء. تظهر النتائج معدل تصنيف أقصى بنسبة **98.81%** تم الحصول عليه باستخدام نماذج HMM ذات 5 حالات، كل منها مرتبط بنموذج ج.م.م (GMM) ذو أربع مركبات غوصيان.

## Introduction Générale

L'augmentation de la consommation d'électricité dans le monde en raison de l'urbanisation et du coût élevé de l'énergie ainsi que du changement climatique suscite l'intérêt pour les économies d'énergie. Les chercheurs ont identifié une opportunité de réduction significative de la consommation d'énergie dans les bâtiments résidentiels et commerciaux, en fournissant des informations détaillées en temps réel sur l'utilisation de l'énergie [1]. Plusieurs travaux de recherche ont proposé l'installation des compteurs intelligents (« smart meter » en anglais).

Les compteurs intelligents sont des appareils de mesure puissants, permettant de remplir les fonctions des compteurs traditionnels, agir comme des capteurs dans le réseau, mesurer et enregistrer la consommation réelle d'électricité pendant une journée ou un intervalle de temps donné [2]. De plus, ils peuvent aussi identifier les appareils électriques présents sur le réseau résidentiel par l'utilisation des méthodes de surveillances des courbes de charges (courant et tension) des appareils domestiques. Ces méthodes se divisent en deux catégories, intrusive et non intrusive [3]. Plus particulièrement, les systèmes de surveillance de la charge non intrusive (NILM) offrent la possibilité de surveiller tous les appareils à partir des données du signal électrique d'un seul compteur [4].

Dans [5], les auteurs ont proposé d'utiliser les modèles de Markov cachés HMM combinés avec l'analyse harmonique pour la conception d'un système IAED. Ce système a été évalué sur la base de données PLAID. Notre objectif est de concevoir un système IAED performant basé sur les modèles HMM combinés avec l'analyse harmonique en utilisant la base de données COOL pour l'évaluation des performances du système. Plus particulièrement, nous allons chercher la bonne configuration du système permettant d'obtenir le meilleur taux de classification des signaux du courant électriques de la base de données COOLL contenant 840 signaux de courant et de tension échantillonnés à 100 KHz, appartenant à 12 classes d'appareils électriques.

Ce travail s'articule autour de trois chapitres. Le premier chapitre présente des généralités sur les systèmes IAED. Dans le deuxième chapitre, nous nous concentrons sur les détails du système proposé en décrivant les différentes phases impliquées dans la conception du système. Le troisième chapitre a pour objectif de mettre en œuvre le système conçu et d'analyser et discuter les résultats des expériences effectuées. On finit par une conclusion générale.

---

## **Chapitre I :**

*Généralités sur les systèmes d'identification  
des appareils électriques domestiques IAED*

---

## **I.1 Introduction**

Un système d'identification des appareils électriques (IAED) tels que les réfrigérateurs, les téléviseurs, les climatiseurs est un processus de reconnaissance des appareils électriques à base de leurs signatures ou modèles électriques extraits à partir des signaux électriques tels que le courant et la tension. Ce système implique une analyse des signaux électriques permettant l'extraction des caractéristiques discriminantes des différentes classes d'appareils. Les systèmes IAED peuvent être mis en œuvre à l'aide de diverses techniques, notamment la surveillance non intrusive de la charge (NILM), les algorithmes d'apprentissage automatique, le traitement des signaux et l'analyse des données. L'identification efficace des appareils électriques peut fournir des informations précieuses sur la consommation d'énergie, permettant aux utilisateurs de prendre des décisions éclairées en matière d'efficacité énergétique, de gestion de la charge et de mesures de conservation d'énergie [6].

## **I.2 Objectifs d'un système IAED**

L'identification des appareils électriques répond à une multitude d'objectifs ayant des implications importantes pour les consommateurs et les fournisseurs d'énergie.

D'un point de vue environnemental, la nécessité d'un équilibre entre la production et la consommation d'électricité est un défi permanent. L'identification des appareils électriques est essentielle pour atteindre cet équilibre, car elle fournit des informations détaillées sur la consommation d'énergie, ce qui permet une gestion efficace de l'électricité, une facturation basée sur l'utilisation et l'échange d'électricité avec d'autres fournisseurs. Pour les consommateurs d'énergie, l'objectif premier de l'identification des appareils est l'efficacité énergétique. En s'appuyant sur des systèmes tels que la surveillance non intrusive de la charge (NILM), qui désagrège la consommation d'énergie au niveau du compteur électrique principal, on peut acquérir une connaissance des appareils consommateurs d'énergie. Cette connaissance permet aux consommateurs d'optimiser leur consommation d'électricité, réduisant ainsi leurs coûts énergétiques [7].

Dans le contexte plus large du réseau intelligent, l'identification des appareils contribue à la gestion de la demande. En fournissant une vue détaillée de la consommation d'énergie pendant les heures de pointe, elle encourage les consommateurs à déplacer leur consommation vers les périodes creuses, ce qui permet d'obtenir un réseau énergétique plus équilibré et plus stable.

Associé aux progrès des technologies de réseaux intelligents et de la communication, ce système peut conduire à des améliorations substantielles de la réponse à la demande et de la connectivité du réseau.

En outre, l'identification des appareils peut être associée à la surveillance de l'état des signaux électriques et à la détection des défauts. Cette combinaison peut potentiellement permettre une identification précoce des défauts, prévenir les pannes inattendues et assurer un fonctionnement plus fiable des appareils.

Enfin, l'identification des appareils trouve également des applications dans la détection des activités à l'intérieur des maisons, contribuant ainsi à des secteurs tels que les soins de santé, les loisirs, la domotique et la surveillance de l'énergie.

Les objectifs de l'identification des appareils électriques sont multiples et jouent un rôle essentiel dans la promotion de l'efficacité énergétique, la maintenance prédictive et la contribution à une série d'applications dans divers secteurs.

### **I.3 Méthodes d'identification des appareils électriques**

L'identification des appareils électriques implique généralement la surveillance des courbes de charge afin de distinguer la consommation individuelle des appareils de la consommation globale d'énergie. Ce processus implique l'extraction de signatures qui caractérisent chaque courbe de charge en fonctionnement, ce qui peut être effectué de manière intrusive ou non intrusive.

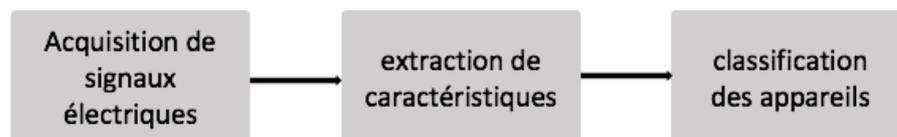
#### **I.3.1 L'approche ILM**

ILM signifie « Intrusive Load Monitoring » (surveillance intrusive de la charge). Il s'agit d'une méthode de surveillance de la consommation d'énergie qui consiste à fixer des compteurs de sous-comptage individuel à chaque appareil électrique. Ces compteurs collectent et transmettent des données sur la consommation d'énergie. Cette approche fournit des informations détaillées sur la consommation d'énergie de chaque appareil, ce qui peut être utile pour la gestion et la conservation de l'énergie. Même si l'installation des dispositifs dans une maison est simple, l'ILM présente certains inconvénients, notamment des problèmes de fiabilité dus à la présence de capteurs multiples, des défaillances potentielles du système causées par des dysfonctionnements d'un seul capteur, et un manque d'évolutivité, puisqu'il faut installer un

très grand nombre de capteurs dans tous les locaux des consommateurs. Il est clair que cela représente un coût prohibitif en matière de main -d'œuvre et de capital [8].

### **I.3.2 L'approche NILM**

Cependant, la surveillance non intrusive de la charge (NILM) ne nécessite que des conceptions énergétiques agrégées provenant d'un seul compteur fixé au point d'entrée du service. Ce compteur est utilisé pour estimer la composition des appareils en état de fonctionnement. Cette approche présente des avantages considérables pour les fournisseurs d'énergie, car elle ne nécessite aucun dispositif supplémentaire et n'impose aucun désagrément aux consommateurs. De plus, elle est relativement simple et peu coûteuse à installer. Le concept de NILM a été introduit pour la première fois par Hart dans les années 1980. Récemment, cette méthode a suscité un vif intérêt en raison de l'amélioration de la puissance de calcul des appareils embarqués et des préoccupations croissantes liées au changement climatique. Le système NILM commence par l'acquisition de signaux électriques, puis les caractéristiques intéressantes sont extraites des échantillons, et enfin les algorithmes de classification de la NILM sont appliqués pour classer les appareils [9].



*Figure I 1 : Les trois étapes du l'approches NILM [8]*

### **I.4 L'impact de l'utilisation de NILM**

Des recherches antérieures ont démontré l'impact positif de la mise en œuvre de la surveillance non intrusive de la charge (NILM) sur le comportement des ménages en matière de consommation d'énergie. En particulier, des études ont montré un potentiel d'économies d'énergie allant jusqu'à 15 % grâce au retour d'information sur la consommation des appareils électriques. Cependant, une analyse de 36 études menées sur 15 ans suggère que l'impact pratique de la surveillance non intrusive de la charge est plus modeste, avec une diminution potentielle de la consommation allant jusqu'à 12 % uniquement lorsque l'information en temps

réel est combinée à un retour d'information quotidien/hebdomadaire sur la consommation d'énergie au niveau des appareils [10], comme l'illustre la figure (I-2).

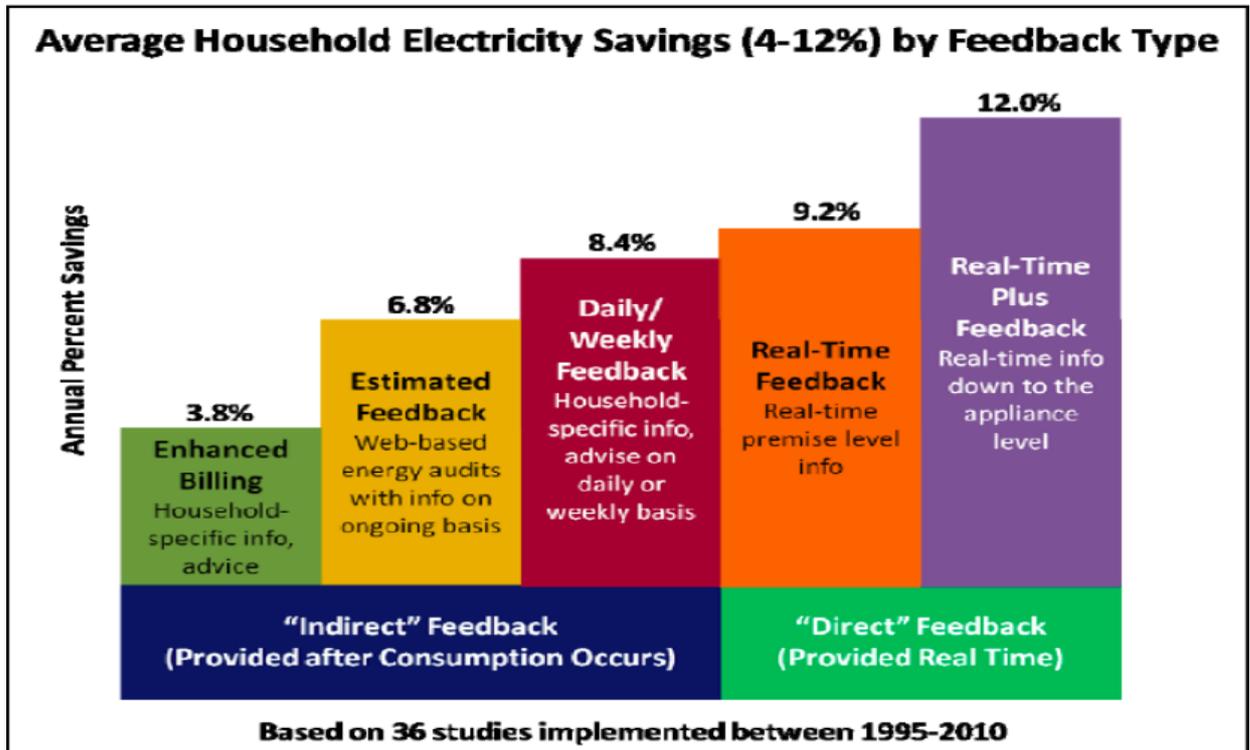


Figure I 2 : Analyse de l'impact de l'utilisation de NILM [10]

## I.5 Principe de fonctionnement d'un système IAED

Le système d'identification des appareils électriques (IAED) fonctionne sur le principe de l'analyse et de la classification des signaux émis par les appareils électriques. Ces signaux comprennent le courant électrique, la tension et la puissance consommée, qui sont ensuite analysés pour caractériser puis classer les différentes classes d'appareils. La procédure du système IAED se déroule en plusieurs étapes comme le montre le schéma synoptique dans (figure I-3).

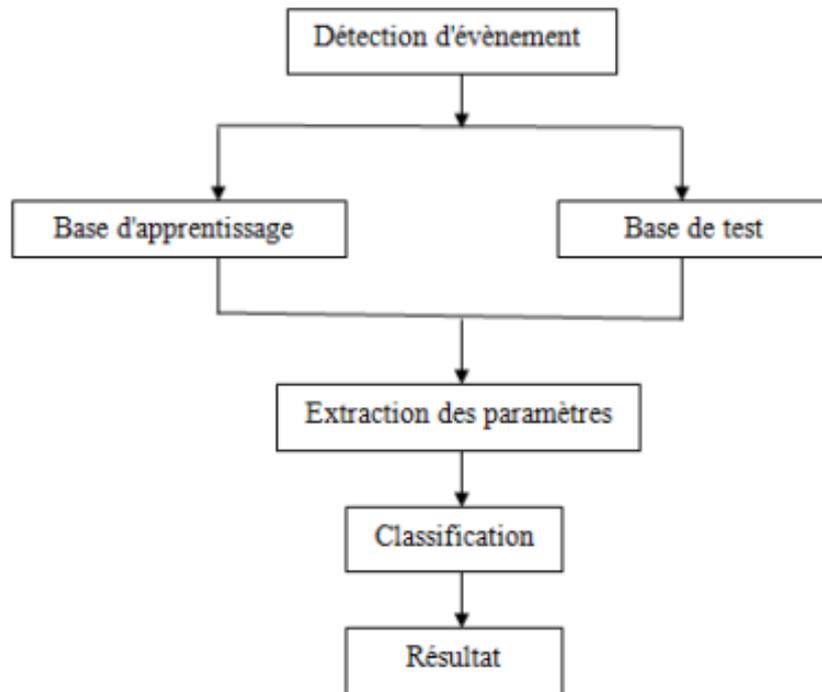


Figure I 3 : Schéma synoptique général du processus d'identification IAED [11]

### **I.5.1 Détection des événements :**

L'étape initiale du processus IAED est la détection des événements. Cette étape consiste à déterminer si un appareil est connecté ou déconnecté du système.

### **I.5.2 Analyse du signal et extraction des caractéristiques :**

Au cours de l'étape d'analyse des signaux et d'extraction des caractéristiques, les signaux émis par les appareils sont évalués et les caractéristiques uniques qui distinguent chaque appareil sont identifiées. Ce processus implique l'extraction de deux types de caractéristiques : les caractéristiques à l'état stable et les caractéristiques à l'état transitoire [11].

**Les caractéristiques de l'état stable :** ses caractéristiques correspondent aux conditions de fonctionnement stables des appareils électriques, observables lorsqu'ils ont atteint un niveau constant de consommation d'énergie. Les caractéristiques en régime permanent sont généralement extraites des composantes de puissance active (puissance réelle) et de puissance réactive (puissance imaginaire) des signaux électriques. Elles fournissent une caractérisation globale de la consommation d'énergie d'un appareil pendant son fonctionnement normal. La

puissance active (P), la puissance réactive (Q) et la puissance apparente (S) sont des exemples de caractéristiques en régime permanent [12].

**Les caractéristiques de l'état transitoire :** Elles sont associées aux moments initiaux où un appareil est mis en marche ou éteint, ou lorsqu'il passe d'un mode de fonctionnement à l'autre. Ses caractéristiques indiquent les variations à court terme des signaux électriques. Les caractéristiques transitoires sont souvent plus instructives que les caractéristiques d'état stable, car elles permettent d'identifier les appareils de manière unique sur la base de leur comportement au démarrage ou à l'arrêt. Cette caractéristique est particulièrement pertinente car les différents appareils ont souvent des signatures de mise en marche ou d'arrêt distinctes, ce qui les rend facilement identifiables au cours de ces périodes. Les signatures de mise en marche et d'arrêt et les harmoniques sont des exemples de caractéristiques transitoires

En résumé, les caractéristiques en régime permanent et en régime transitoire fournissent des informations essentielles pour l'identification des appareils électriques.

### **I.5.3 Identification des appareils électriques :**

Le signal électrique d'entrée d'un système IAED est associé à une classe d'appareils à l'aide de l'un des classificateurs standard tel que K-Nearest Neighbour (KNN), Hidden Markov Model (HMM)...etc. Cette étape aboutit à une vue désagrégée de la consommation d'énergie de chaque appareil.

La conception d'un système IAED repose en grande partie sur deux phases principales : la phase d'apprentissage et la phase d'identification et d'évaluation.

## **I.6 État de l'art sur les bases de données**

Les bases de données en NILM comprennent généralement des données sur la consommation d'énergie de divers appareils électriques domestiques, collectées sur une certaine période de temps. Cet ensemble de données peut contenir différents types de données tels que la puissance réelle, la puissance réactive, la tension, le courant et les harmoniques [14]. Ces bases de données peuvent être caractérisées par leurs tailles, la fréquence d'échantillonnage des signaux électriques, types de données (courant, tension, puissance, etc).

Le tableau suivant donne un état de l'art abrégé sur les différentes bases de données utilisées dans des travaux de recherche.

Tableau I - 1 : Comparaison entre les différentes bases de données [14]

<b>Base de données</b>	<b>Date de création</b>	<b>Fréquence d'échantillonnage</b>	<b>Nombre de classes d'appareils</b>
<b>Smart</b>	2012	1 Hz	25
<b>Tracebase</b>	2014	1 Hz	20
<b>PLAID</b>	2014	30 KHz	11
<b>COOLL</b>	2016	100 KHz	12
<b>BLOND</b>	2018	50-250 KHz	16
<b>WHIETED</b>	2016	44.1 KHz	47

## I.7 État de l'art sur les classificateurs

Il existe de nombreux algorithmes de classification qui sont utilisés pour identifier les appareils individuels sur la base de leurs modèles uniques de consommation d'énergie. Nous pouvons distinguer deux types d'algorithmes de classification, à savoir :

- **Algorithme Supervisé** : ces algorithmes nécessitent des données étiquetées pour entraîner le classificateur [12].
- **Algorithmes non supervisés** : contrairement aux méthodes supervisées, les méthodes non supervisées ne nécessitent pas de données étiquetées. Ces méthodes utilisent des techniques d'apprentissage non supervisé et visent à désagréger directement les mesures de charge agrégées sans effectuer aucune sorte de détection d'événement [12].

### I.7.1 Les différentes méthodes de classification :

**Méthode K-NN (les k voisins plus proches)** : il s'agit d'un algorithme d'apprentissage simple, basé sur des instances. Il classe les appareils en fonction de la similarité de leurs caractéristiques avec celles des instances enregistrées préalablement. Le "k" de k-NN fait référence au nombre de voisins les plus proches pris en compte lors de la prise de décision de classification. [11].

**Machines à vecteurs de support (SVM) :** les SVM sont des algorithmes d'apprentissage puissants et largement utilisés qui peuvent traiter des problèmes de classification linéaires et non linéaires. Il fonctionne en trouvant l'hyperplan qui sépare le mieux des classes dans l'espace des caractéristiques [15].

**Arbres de décision :** Un arbre de décision est un algorithme d'apprentissage supervisé non paramétrique, qui est utilisé pour les tâches de classification [16].

**Réseaux neuronaux artificiels (RNA) :** les réseaux neuronaux artificiels s'inspirent du cerveau humain, Il s'agit d'un algorithme supervisé utilisé pour la tâche de classification [17].

**Modèles de Markov cachés (HMM) :** les HMM sont des modèles statistiques qui supposent que le système modélisé est un processus de Markov avec des états cachés. Dans le contexte de la NILM, les états cachés pourraient représenter les états de puissance des appareils, et le HMM peuvent être utilisés pour estimer la séquence la plus probable des états des appareils compte tenu de la consommation électrique observée [18].

Chacune de ces méthodes a des avantages et des inconvénients et le choix de la méthode dépend souvent des caractéristiques des données et des exigences spécifiques de la tâche NILM.

### I.7.2 Comparaison entre les algorithmes de classification :

Tableau I - 2 : Comparaison entre les différents algorithmes de classification [12]

Algorithme d'apprentissage	Régimes permanent/transitoire	Précision %	Apprentissage supervisés/ non-supervisés)	Online/Offline
SVM	Les deux	75–98	<b>Supervisés</b>	Online
Bayes	État permanent	80–99	<b>Supervisés</b>	Les deux
HMM	État permanent	75–95	Les deux	Offline
Neural Networks	Les deux	80–97	<b>supervisés</b>	Online

<b>Algorithme d'apprentissage</b>	<b>Régimes permanent/transitoire</b>	<b>Précision %</b>	<b>Apprentissage supervisés/ non-supervisés)</b>	<b>Online/Offline</b>
KNN	Les deux	70–90	<b>supervisés</b>	Les deux
Optimization	État permanent	60–97	<b>supervisés</b>	Offline

## **I.8 Conclusion**

Dans ce chapitre, nous avons introduit les concepts généraux des systèmes d'identification des appareils électriques (IAED). Nous avons brièvement décrit l'objectif de cette tâche d'identification ainsi que son principe de fonctionnement qui se décompose en trois étapes principales : la détection des événements, l'extraction des caractéristiques et l'identification des appareils. De plus, nous avons présenté l'état de l'art sur les bases de données. Enfin, nous avons exposé un état de l'art concernant les méthodes de classification et les techniques d'extraction des paramètres.

---

## **Chapitre II**

*Système d'identification des appareils  
électriques basé sur les modèles HMM*

---

## II.1 Introduction

Dans ce chapitre, nous allons décrire notre système d'identification des appareils électriques en utilisant une méthode d'extraction de paramètres basée sur l'analyse harmonique, combinée avec l'algorithme de classification HMM.

## II.2 Travaux liés (related work)

Dans [5], les auteurs ont appliqué les modèles de Markov cachés HMM combinés avec l'extraction des paramètres harmoniques STFS pour l'identification des appareils électriques domestiques IAED. Le fonctionnement de ce système est illustré sur la figure (II-1).

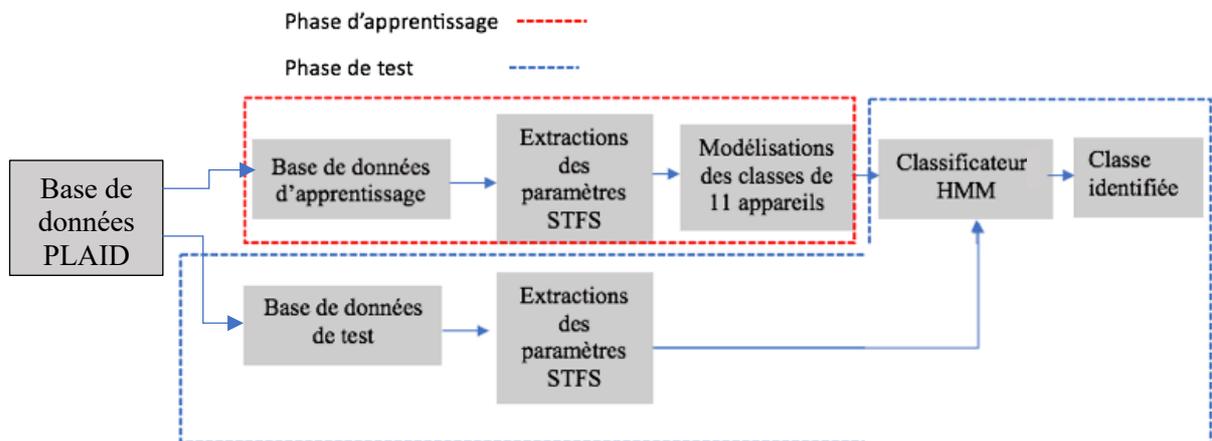


Figure II 1 : Le schéma synoptique qui représentant le principe de fonctionnement d'un système d'identification à base de données PLAID [5]

Ce système se déroule en une phase d'apprentissage permettant de modéliser les différentes classes d'appareils et une phase de test permettant d'identifier les classes de signaux électriques de test et d'évaluer ensuite les performances du système IAED sous une base de données. Chacune des phases exige une étape d'extraction de paramètres permettant de découper le signal électrique en une séquence de fenêtres d'analyse chevauchées et de convertir ensuite chacune des fenêtres en un vecteur de paramètres STFS. Ainsi, chaque signal électrique est converti en une séquence de vecteurs de paramètres STFS à utiliser à l'entrée du classificateur HMM. Les auteurs ont utilisé la base de données PLAID répartie en une base de données d'apprentissage et une base de données de test. Ils ont utilisé les signaux du courant électrique de cette base pour évaluer les performances du système IAED. Cette base de données est constituée de 1074 signaux de 11 appareils électriques enregistrés dans 56 maisons. Ce système

IAED est implémenté en utilisant la boîte à outils HTK sous l'environnement Matlab. Dans la phase d'apprentissage, les séquences de vecteurs de paramètres STFS correspondant aux signaux électriques d'entraînement sont utilisées pour estimer les paramètres des modèles HMM des différentes classes d'appareils en appliquant un algorithme d'apprentissage appelé « algorithme de Baum-Welch ». Dans la phase de test, les séquences de vecteurs de paramètres STFS correspondant aux signaux électriques de test sont utilisées pour identifier leurs classes d'appareils en utilisant un algorithme de classification appelé « algorithme de Viterbi ». Les différentes étapes d'implémentation du système sont :

- Préparation des données.
- Extraction des paramètres STFS
- Modélisation des différentes classes d'appareils par des modèles HMM.
- Identification des classes des signaux de test.
- Évaluation des performances en terme du taux d'identification (TI).

Dans [5], les auteurs ont étudié les points suivants :

- Comparaison des performances du régime transitoire et régime permanent.
- La pertinence des rangs pair et impair des paramètres harmoniques pour cette tâche d'identification.
- L'effet de la taille de la base d'apprentissage sur les performances du système.
- La réduction de la dimensionnalité des vecteurs de paramètres STFS.

Les auteurs ont conclu les points suivants :

- Le régime transitoire porte plus d'information sur les différentes classes d'appareils.
- Les paramètres harmoniques de rang impair sont plus pertinents par rapport à ceux du rang pair.
- Les bases d'apprentissage les plus larges donnent les meilleures performances.
- La réduction de la dimensionnalité peut améliorer les performances du système.

Dans le travail [7], les auteurs ont proposé l'extraction des paramètres statistiques des coefficients, combinée avec l'algorithme KNN et la règle de vote.

Ces systèmes d'identification sont évalués sous la base de données PLAID. Cette base de données est constituée de 1074 signaux échantillonnés à une fréquence de 30 KHz. [5]

Cependant, l'algorithme KNN peut être lent et exige beaucoup d'espace mémoire et temps de calcul, en particulier avec des grandes bases de données, car il doit calculer la distance entre un vecteur de test donné et tous les autres vecteurs d'apprentissage.

Dans notre travail, nous proposons d'implémenter un système IAED basé sur les modèles HMM et l'analyse harmonique en utilisant la base de données COOLL.

## **II.3 Architectures du système IAED implémenté**

Le système IAED à implémenter consiste à identifier 12 types d'appareils à partir des signaux du courant électrique. Le fonctionnement du système se déroule en une phase d'apprentissage et une phase de test. La phase d'apprentissage a pour objectif d'estimer les paramètres de chaque modèle HMM associé à chaque classe ou type d'appareils en utilisant les signaux d'entraînement. La phase de test consiste à identifier la classe d'appareil de chaque signal de test et d'évaluer ensuite les performances du système en termes du taux d'identification en utilisant la base de données COOLL. Chacune des phases exige une étape de prétraitement et d'extraction de paramètres caractérisant les différentes classes d'appareils.

### **II.3.1 Etapes de fonctionnement du système IAED implémenté**

Les différentes étapes de fonctionnement du système sont illustrées sur la figure (II.2).

La première étape consiste à prétraiter les données, en particulier les signaux électriques de la base de données. L'objectif est d'éliminer le bruit, en se concentrant uniquement sur les signaux de courant lorsque l'appareil est en état de fonctionnement.

Une fois l'étape de prétraitement des données est terminée, on divisera notre base de données en deux ensemble distincts: une base de données d'apprentissage et une base de données de test. La base de données d'apprentissage sera utilisée pour entraîner notre modèle, tandis que la base de données de test servira à évaluer les performances du système.

L'étape suivante est l'extraction des paramètres. Cette étape consiste à transformer les signaux de courant en une séquence des vecteurs de coefficients harmoniques. Chaque vecteur est extrait de l'application de l'analyse harmonique sur  $K$  fenêtres d'analyse chevauchées. Chaque fenêtre d'analyse est convertie en un vecteur de coefficients de séries de Fourier à temps court (STFS).

Une fois que les modèles de Markov caché (HMM) ont été entraînés en utilisant la base de données d'apprentissage, nous passons à la phase d'évaluation.

Nous utilisons la base de données de test pour effectuer cette évaluation. Cette procédure garantit une mesure des performances du modèle, ce qui nous permet d'évaluer la précision du système IAED en termes de taux de classification ou d'identification (TI) des signaux.

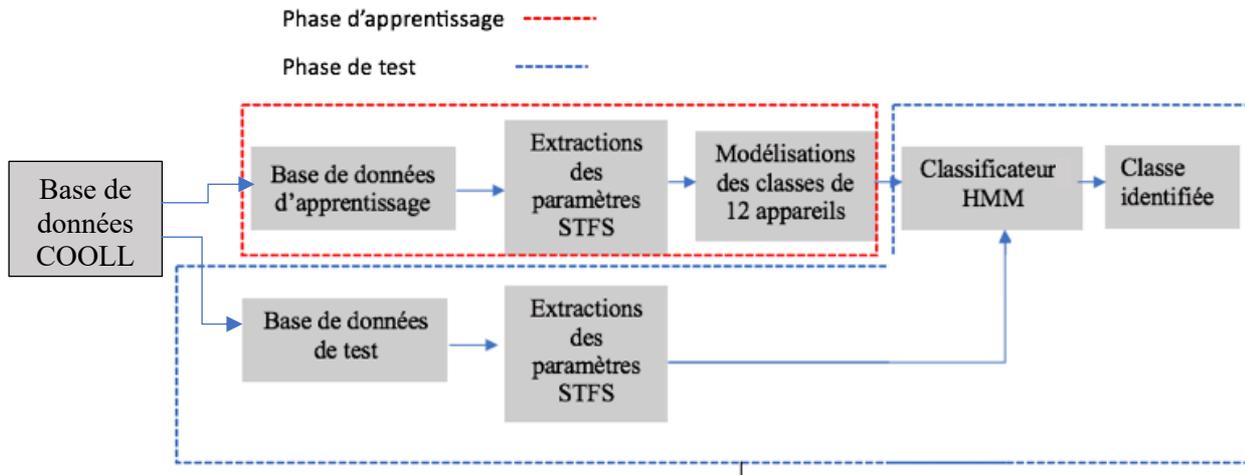


Figure II 2 : Le schéma synoptique représentant le principe de fonctionnement du système d'identification implémenté évalué sous la base de données COOLL

### II.3.2 Base de données COOLL

COOLL est une base de données échantillonné à une haute fréquence comprenant des mesures électriques au niveau des prises. Il contient le courant et la tension électriques d'un total de 42 appareils. Chaque appareil est utilisé pour enregistrer 20 signaux échantillonné à une fréquence ( $F_s$ ) de 100 KHz, ce qui donne une collection de 840 mesures de courant et de 840 mesures de tension. Les appareils comprennent 12 types distincts, avec un nombre variable d'instances pour chaque type [19].

Notamment, la base de données COOLL garantit que chaque appareil est mesuré de manière isolée. Cela signifie qu'il n'y a pas de cas dans la base donnée où plusieurs appareils fonctionnent simultanément. Le tableau (II.1) illustre des détails sur le type, nombre d'appareils de la base de données COOLL.

Tableau II - 1 : Résumé de la base de données COOLL [18]

<b>N °</b>	<b>Type d'appareil</b>	<b>Nombre d'appareil</b>	<b>Signaux de courant (20 par appareil)</b>
<b>1</b>	Drill	6	120
<b>2</b>	Fan	2	40
<b>3</b>	Grinder	2	40
<b>4</b>	Hair dryer	4	80
<b>5</b>	Hedge trimmer	3	60
<b>6</b>	Lamp	4	80
<b>7</b>	Paint stripper	1	20
<b>8</b>	Planer	1	20
<b>9</b>	Router	1	20
<b>10</b>	Sander	3	60
<b>11</b>	Saw	8	160
<b>12</b>	Vacuum cleaner	7	140
<b>TOTAL</b>		42	840

### II.3.3 Prétraitements de base de données

L'étape du prétraitement consiste à extraire que la partie utile du signal représentant l'état de fonctionnement de l'appareil considéré. La figure (II.3) illustre le signal du courant électrique avant le prétraitement, alors que la figure (II.4) représente le cas après le prétraitement. Ainsi, l'extraction des paramètres harmonique s'effectue sur la partie utile du signal.

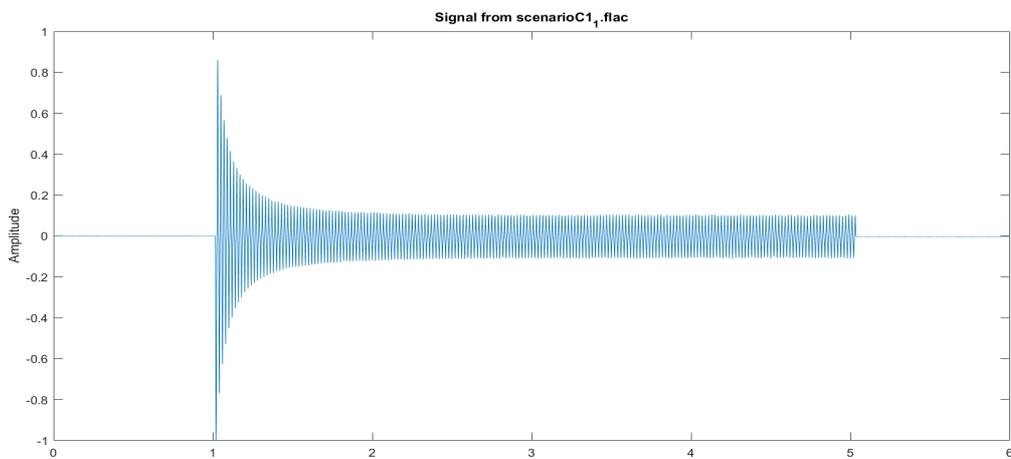


Figure II 3 : Signal du courant avant prétraitement

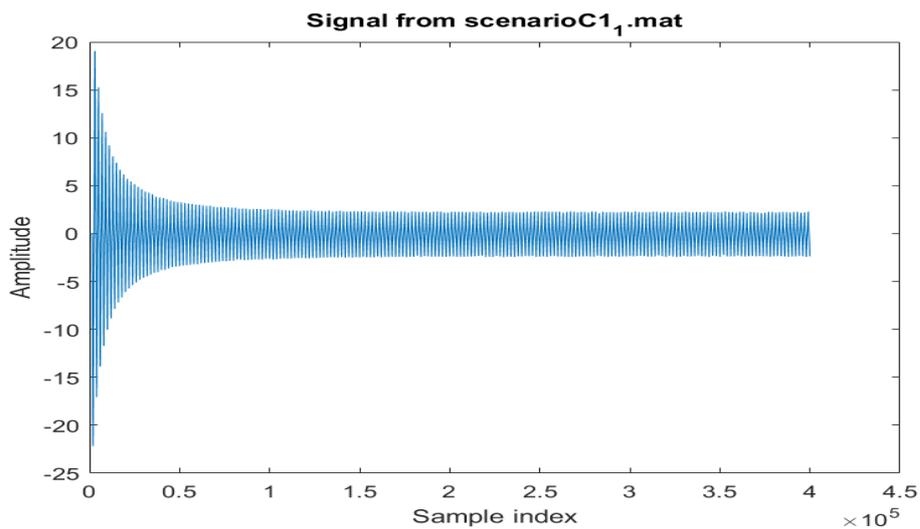


Figure II 4 : Signal de curant après le prétraitement

Dans cette étape, nous avons réduire le nombre d'échantillons des signaux en sous échantillonnant chaque à une fréquence d'échantillonnage réduite égale à 25 KHz, dans le but de réduire le temps de calcul et d'espace mémoire.

### II.3.4 Fenêtrages

Dans cette étape, nous avons divisé le signal en fenêtres d'analyse chevauché d'une durée de 20 ms (ce qui correspond à 1/50 Hz, où 50 Hz est la fréquence de réseau), puis nous avons converti chaque fenêtre d'analyse en un vecteur de  $N_f$  (Nombre de coefficients STFS) en appliquant l'analyse de Fourier.

La range des coefficients varie de 0 à  $N_f$

$$N_f = \frac{F_s}{2} \times \frac{1}{F_0} \quad (II.1)$$

$F_s$  : Fréquence d'échantillonnage qui égale à 25 KHz.

$F_0$  : Fréquence de réseau qui égale à 50 Hz.

### II.3.5 Analyse de Fourier à court terme :

Cette étape consiste à convertir chaque signal  $x(n)$  d'une fenêtre d'analyse en un vecteur de coefficients STFS en appliquant l'analyse de Fourier à court terme.

La représentation mathématique de cette décomposition est donnée par la formule suivante :

$$x(n) = \sum_{k=0}^{N-1} C_k e^{j2\pi kn/N} \quad (II.2)$$

Dans cette expression,  $x(n)$  est le signal périodique discret,  $C_k$  représente les coefficients de Fourier, calculés à l'aide de la formule :

$$C_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \quad (II.3)$$

$e^{j2\pi kn/N}$  est la série de fonctions exponentielles orthogonales. Ici,  $N$  représente le nombre total d'échantillons dans chaque fenêtre d'analyse.

La série de  $N$  coefficients  $NC_k$  avec  $k = 0, 1, \dots, N - 1$  forme le spectre de fréquence discret de la fenêtre d'analyse considérée. Ce spectre s'étend sur une plage de fréquences allant de 0 à  $F_s$ .

La bande de fréquence maximale du signal est  $[0, F_s/2]$  avec  $N/2$  points de fréquence représentant cette bande. Par conséquent, le spectre utile de la fenêtre d'analyse comprend les coefficients  $C_k$  allant de 0 à  $(N/2) - 1$ .

Chaque coefficient est porteur d'informations spécifiques :  $C_1$  correspond à la fréquence fondamentale, les coefficients avec  $k = 2, \dots, (N/2) - 1$  correspondent aux fréquences harmoniques, et le terme à  $k = 0$  correspond à la composante DC.

Dans ce travail, le module de chaque coefficient  $C_k$  pour  $k = 0, \dots, (N/2) - 1$  est extrait, fournissant un ensemble de paramètres qui représentent la fenêtre d'analyse. Par conséquent, chaque fenêtre d'analyse est caractérisée par un ensemble de  $(N/2)$  pour obtenir un total de 250 paramètres. Ces paramètres extraits, également connus sous le nom de coefficients de la série de Fourier à court terme (STFS), représentent le spectre d'amplitude de la fenêtre d'analyse.

### II.3.6 Les modèles de Markov

Les modèles de Markov portent le nom de nom à Andrey Markov, qui les a développés pour la première fois au début des années 1900. Les modèles de Markov sont un type de modèle probabiliste utilisé pour prédire l'état futur d'un système, sur la base de son état actuel. En d'autres termes, les modèles de Markov sont utilisés pour prédire l'état futur à base des états cachés ou observés actuels. Le modèle de Markov est une machine à états finis où chaque état a une probabilité associée d'être dans n'importe quel autre état après une transition. Les modèles de Markov peuvent être classés en modèles cachés et observables en fonction du type d'informations disponibles à utiliser pour faire des prédictions ou prendre des décisions. Les modèles de Markov cachés traitent de variables cachées qui ne peuvent pas être observées directement mais seulement déduites d'autres observations, alors que dans un modèle observable, également appelé chaîne de Markov, les variables cachées ne sont pas impliquées [18].

### II.3.7 Qu'est-ce qu'une chaîne de Markov ?

Une chaîne de Markov est un système mathématique qui connaît des transitions d'un état à un autre selon certaines règles probabilistes. La caractéristique principale d'une chaîne de Markov est que, quelle que soit la manière dont le processus est arrivé à son état actuel, les états futurs possibles sont fixes.

Une chaîne de Markov se compose des éléments suivants [18]:

- **Distribution de probabilité initiale** : Une distribution de probabilité initiale sur les états,  $\pi_i$  est la probabilité que la chaîne de Markov commence dans l'état  $i$ . Certains états  $j$  peuvent avoir  $\pi_j = 0$ , ce qui signifie qu'ils ne peuvent pas être des états initiaux. La distribution d'initialisation définit chaque variable cachée dans son état initial au moment  $t=0$  (l'état caché initial).
- **Un ou plusieurs états cachés**
- **Distribution de probabilité de transition** : Une matrice de probabilité de transition où chaque  $a_{ij}$  représente la probabilité de passer de l'état  $i$  à l'état  $j$ . La matrice de transition est utilisée pour montrer les probabilités de transition d'état caché à état caché.
- **Une séquence d'observations**
- **Probabilités d'émission** : Une séquence de vraisemblances d'observation, également appelées probabilités d'émission, exprimant chacune la probabilité qu'une observation  $o_i$  générée à partir d'un état  $I$ . La probabilité d'émission est utilisée pour définir la variable cachée en termes d'état caché suivant. Elle représente la distribution conditionnelle sur une sortie observable pour chaque état caché au temps  $t=0$ .

### II.3.8 Hidden Markov models (HMM)

Les modèles de Markov cachés (HMM) sont une autre évolution possible des chaînes de Markov. Ces nouveaux modèles se basent essentiellement sur deux processus stochastiques dépendants l'un de l'autre. Contrairement aux chaînes de Markov traditionnelles, l'état du système n'est pas directement observable, mais plutôt dissimulé par un processus d'observation [20].

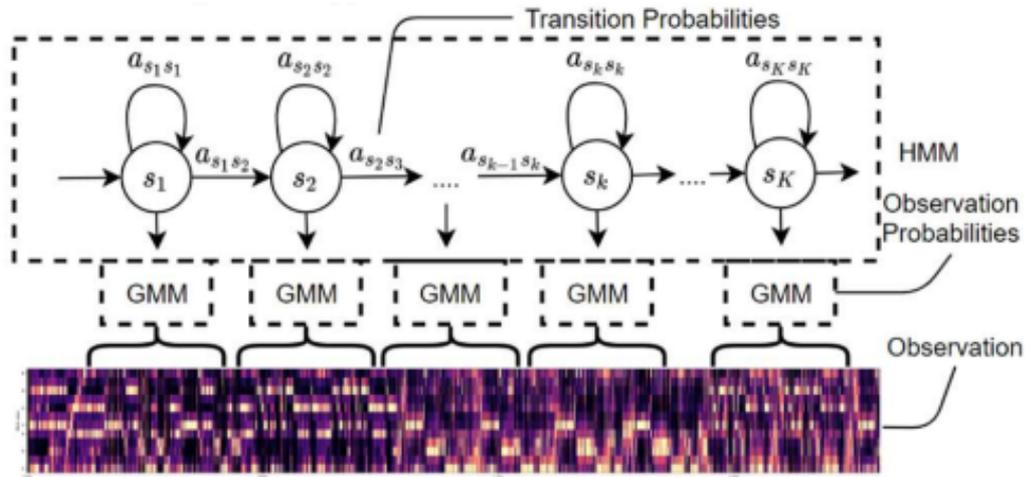


Figure II 3 : Topologie graphique d'un HMM [21]

Les HMM sont caractérisés par les paramètres suivants [22]:

- Le nombre N des états du modèle.
- Le nombre N des états du modèle.
- Le nombre M des symboles d'observations.
- La distribution A des probabilités des transitions des états.
- La distribution B des probabilités d'émission d'observations dans chaque état j.
- La distribution  $\pi$  des probabilités d'état initial ou de départ.

$$A = [\alpha_{ij}] \quad (II.4)$$

$$\alpha_{ij} = P\left(q_{k+1} = \frac{j}{q_k} = i\right) \quad (II.5)$$

$$B = [b_j(O_t)] \quad (II.6)$$

$$b_j(O_t) = P(q_{k+1} = (O_t/q_t = i)) \quad (II.7)$$

$$\pi = [\pi_1, \pi_2, \dots, \pi_N] \quad (II.8)$$

$$\pi_i = P(q_0 = i) \quad (II.9)$$

Ou  $i, j \in \{1, 2, \dots, N\}, t \in \{1, 2, \dots, t\} \quad (II.10)$

$O_t$  : Vecteur des observations

A : Matrice des probabilités de transition entre les états du modèle de Markov caché.

B : Matrice des probabilités d'émission d'observation.

$\pi$  : vecteur de distribution d'état initial.

### II.3.9 Les modèles de mélange gaussien (GMM) :

Les modèles de mélange gaussien (GMM) sont des modèles probabilistes utilisés pour représenter des bases de données du monde réel en les regroupant en plusieurs distributions gaussiennes. Le GMM est caractérisé par ses paramètres, notamment les poids, les moyennes et les covariances des composantes gaussiennes. L'estimation de ces paramètres est généralement réalisée à l'aide de l'algorithme de maximisation des attentes (EM). Le GMM est utilisé dans diverses applications telles que la reconnaissance d'images et de la parole [23].

### II.3.10 Fonctionnalités d'un HMM

Les modèles de Markov cachés (HMM) sont associés à trois problèmes fondamentaux qui nécessitent trois algorithmes correspondants pour être résolus [24][25]:

- **Problème d'évaluation de la vraisemblance** : étant donné les paramètres du modèle et les données observées, calculer la probabilité des données observées. Ce problème peut être résolu efficacement à l'aide de l'algorithme Forward-Backward.
- **Problème du décodage** : étant donné les paramètres du modèle et les données observées, trouver la séquence la plus probable d'états cachés. L'algorithme de Viterbi est utilisé pour résoudre ce problème. Il utilise la programmation dynamique pour trouver la séquence la plus probable d'états cachés, également connue sous le nom de chemin de Viterbi, qui aboutit à une séquence d'événements observés.
- **Problème d'apprentissage** : étant donné les données observées, estimer les paramètres du modèle. Ce problème est résolu par l'algorithme de Baum-Welch (un cas particulier de l'algorithme de maximisation des attentes). L'algorithme de Baum-Welch est utilisé pour entraîner les modèles HMM. Il estime les paramètres inconnus (probabilités de transition et d'émission) selon le principe du maximum de vraisemblance.

Ces trois algorithmes constituent la base de construction des les HMM. Ils nous permettent de calculer les probabilités des séquences observées, de déduire les séquences d'états cachés et d'apprendre les paramètres du modèle qui expliquent mieux les données observées. Dans notre travail, on a utilisé la boîte à outils HTK comprenant l'algorithme « Baum-Welch » utilisé pour l'estimation des modèles HMM à partir de la base de données d'apprentissage ainsi que l'algorithme de « Viterbi » appliqué pour l'étape de décodage et de l'identification des classes d'appareils des signaux électriques de test.

## **II.4 Conclusion**

Dans ce chapitre, nous avons abordé la conception d'un système d'identification des appareils électriques en utilisant les modèles de Markov cachés (HMM). Le système s'appuie sur des travaux antérieurs qui utilisaient l'extraction de paramètres STFS et la classification des signaux électrique basée sur les modèles HMM.

Nous avons détaillé l'utilisation de la base de données COOLL pour l'implémentation et l'évaluation du système IAED. Cela nous prépare à la phase suivante : la mise en œuvre pratique et l'analyse détaillée.

---

## **Chapitre III**

### *Implémentation et résultats*

---

### III.1 Introduction

L'identification des appareils électriques domestiques fait partie intégrante du fonctionnement des systèmes de comptage intelligents modernes, un élément crucial dans le cadre des efforts déployés pour économiser l'énergie. Cette tâche consiste principalement à distinguer le type d'appareil électrique sur la base de son signal électrique unique. Dans ce contexte, nous proposons d'appliquer les modèles de Markov cachés pour cette tâche.

Le système IAED que nous implémenté est conçu autour de l'extraction de paramètres discriminants à partir des coefficients de la série de Fourier à court terme (STFS) du signal du courant électrique de l'appareil. Ces paramètres sont ensuite introduits dans un classificateur HMM pour le processus d'identification. On utilise les signaux de la base de données COOLL pour effectuer cette tâche et évaluer les performances du système.

L'objectif principal de notre travail est d'évaluer l'efficacité du classificateur HMM et d'étudier la pertinence des paramètres STFS dans le contexte de l'identification des appareils électriques domestiques en utilisant la base de données COOLL constituée des signaux électriques de 12 classes d'appareils.

Les questions clés qui guident notre recherche sont les suivantes :

1. Quelle est la meilleure topologie des modèles HMM ?
2. Quel est le nombre optimal de paramètres STFS pour obtenir les meilleures performances ?
4. Quel est l'ordre optimal des coefficients STFS (pair/impair) ?

Pour répondre à ces questions, nous allons effectuer une série d'expériences portant sur chacun de ces aspects spécifiques.

L'implémentation de ce système est effectuée en utilisant la boîte à outils HTK sous l'environnement Matlab .

## III.2 Implémentation du système d'identification

La figure (III-1) illustre les différentes étapes nécessaires à la construction de notre système IAED.

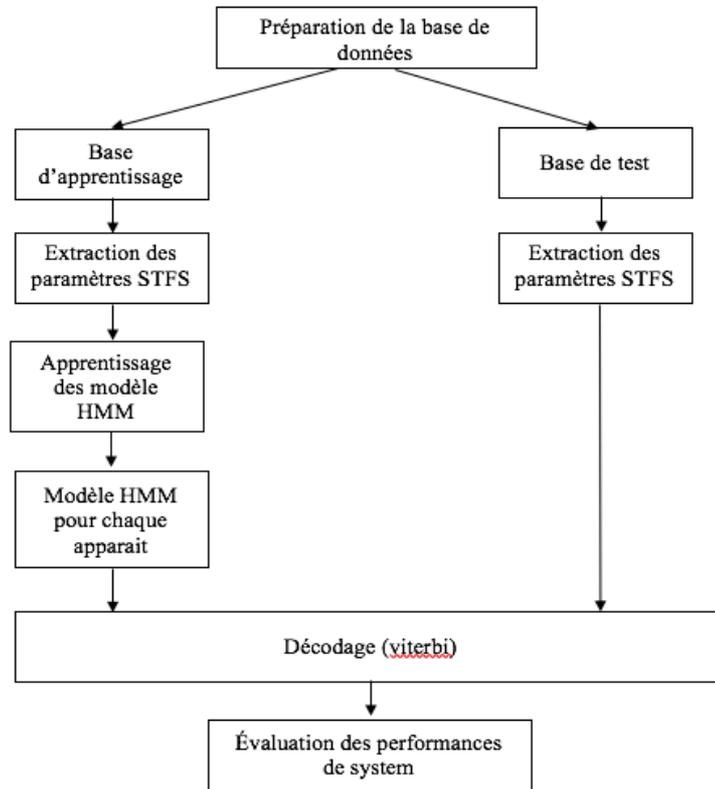


Figure III 1 : Schéma synoptique de notre système IAED

Ces étapes sont données comme suit :

1. Préparation de la base de données d'apprentissage et de la base de test.
2. Apprentissage des modèles HMM.
3. Classification des signaux électriques en classes d'appareils.
4. Évaluation des performances du système.

### III.2.1 Préparation des données :

La base de données est composée de 840 signaux répartis en un ensemble d'apprentissage et un ensemble de test. Chaque signal de la base données est nommé "scénarioC1\_X", où X va de 1 à 840. Les signaux portant des numéros pairs, par exemple "scénarioC1\_2", "scénarioC1\_4", et ainsi de suite, sont attribués à la base de données de test. Inversement, les signaux portant

des nombres impairs tels que "scénarioC1\_1", "scénarioC1\_3", etc. sont attribués à la base de données d'apprentissage.

Après avoir extrait les caractéristiques de la base de données COOLL à l'aide de l'analyse harmonique, les résultats obtenus sont stockés dans le format de fichier MFC à l'aide de la fonction suivante :

```
Writehtk (tlinen, z, fp1,9)
```

La fonction **writehtk** est une fonction MATLAB utilisée pour enregistrer des données au format HTK (Hidden Markov Model Toolkit).

- **tlinen** : le nom du fichier dans lequel les données seront sauvegardé
- **z** : les données à écrire.
- **fp1** : la durée de trame en seconde.
- **9** : le code de type représentant le type de données de Format user.

### III.2.2 Etiquetages :

Dans cette étape, nous générons un fichier d'étiquettes d'extension "Lab" prenant le même nom du fichier de format 'mfc. Par exemple, si le premier fichier de signaux de la base de données est nommé "scenarioC1\_1.MFC", nous créons un fichier d'étiquettes nommé "scenarioC1\_1.lab". Nous y inscrivons ensuite le nom de l'appareil.

La figure (III.2) présente le fichier label.

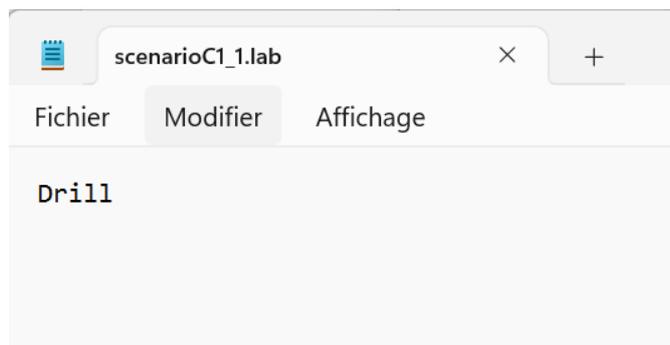


Figure III 2 : Exemple d'un fichier label

### III.2.3 Apprentissage des modèles HMM

La modélisation des modèles HMM s'effectue en trois étapes fondamentales suivantes :

- L'initialisation des modèles HMM.
- La réestimation de ces modèles.
- L'augmentation du nombre de gaussiennes.

#### III.2.3.1 L'initialisation des modèles HMM

L'étape d'initialisation consiste à initialiser tous les modèles HMM avec un modèle HMM initial. Ce modèle initial, appelé modèle prototype définit la topologie des modèles de classes.

Ce prototype spécifie la dimension des vecteurs de caractéristiques, le type de paramètres de caractéristiques, le nombre d'états, le type de topologie du modèle (par exemple, gauche-droite).

Dans notre travail, nous avons utilisé une configuration composée de cinq États (trois états actifs et deux états non actifs d'entrée et de sortie), associé chacun à un modèle GMM représenté par un vecteur de moyennes nuls et une matrice de covariance diagonale. La matrice de covariance diagonale suppose que les caractéristiques ou les observations sont indépendantes les unes des autres dans chaque État du HMM. Cette hypothèse simplificatrice permet une modélisation efficace et une complexité des calculs réduite.

```
HCompV -C config -f 0.01 -m -S trainlist.txt -M hmm0 proto`
```

- ``-C config`` : spécifie le fichier de configuration ``config`` qui contient les paramètres de configuration du HMM (voire la figure (III.4)).
- ``-S trainlist`` : spécifie le fichier qui contient une liste de fichiers de données d'apprentissages à utiliser pour l'estimation des paramètres (voire la figure (III.5)).
- ``-M hmm0`` : spécifie le répertoire de sortie ``hmm0`` où les modèles HMM initialisés seront stockés.
- ``proto`` : spécifie le fichier du modèle prototype qui définit la structure et les caractéristiques initiales des modèles HMM. (Voire la figure (III.3)).



En exécutant cette commande, l'outil **HCompV** de HTK effectue l'étape d'initialisation en estimant les paramètres initiaux du modèle HMM sur la base du modèle prototype fourni et Les fichiers de données d'apprentissage listés dans `trainlist` seront utilisés pour estimer les paramètres du HMM. Les modèles HMM initialisés qui en résultent seront stockés dans le répertoire `hmm0`. Le fichier `proto` sert de modèle prototype, définissant la structure et les caractéristiques initiales des HMM.

proto	15/05/2023 12:53	Fichier	19 Ko
vFloors	15/05/2023 12:53	Fichier	2 Ko

Figure III 6 : les fichiers résultant après l'exécution de la command HCompV

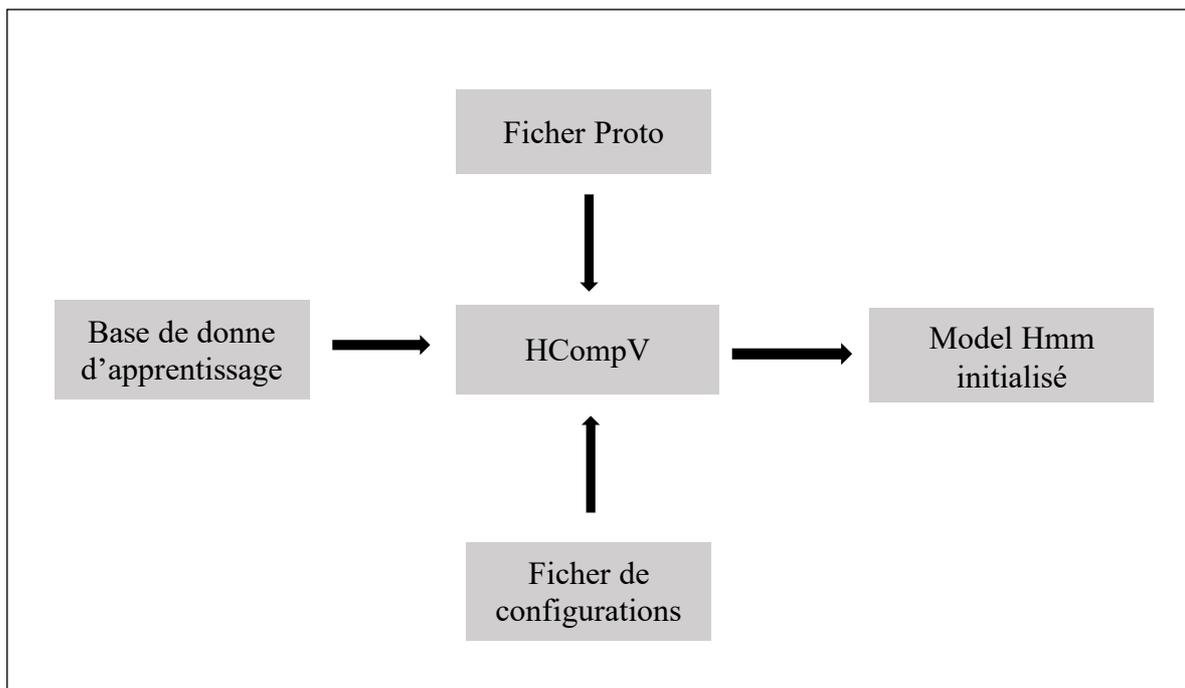


Figure III 7: L'initialisation des modèles HMMs par l'outil HCompV

### III.2.3.2 Réestimation des modèles HMM :

L'objectif de cette étape est d'effectuer une réestimation des paramètres pour les modèles de Markov cachés (HMM) associés aux différentes classes. Cette tâche est accomplie à l'aide de l'outil **HRest** fourni par Hidden Markov Model Toolkit (HTK). L'outil **HRest** utilise

l'algorithme de réestimation de Baum-Welch, pour mettre à jour les paramètres des modèles HMM.

```
HERest -C config -I trainmlf -L -t 250.0 150.0 1000.0 -S trainlist -H
hmm0\macros -H hmm0\hmmdef -M hmm1 liste_modeles
```

- **`-C config`** : Spécifie le fichier de configuration contenant les paramètres nécessaires pour la réestimation HMM.
- **`-I trainmlf`** : est un fichier qui contient les chemins d'accès à tous les fichiers d'étiquetage de la base de données d'apprentissage(voir la figure (III.9)).
- **`-S trainlist`** : spécifie le fichier **`trainlist`** qui liste les fichiers de données d'apprentissage à utiliser pour la réestimation.
- **`-M hmm1`** : spécifie le répertoire de sortie **`hmm1`** où les modèles HMM réestimés seront stockés.
- **`liste\_modeles`** : spécifie le fichier contenant la liste des noms des modèles HMM à réestimer (voir la figure (III.8)).
- **Hmmdef** : le fichier **`hmmdef`** est une collection de modèles HMM représentant différentes classes d'appareils(voir la figure (III.11)). Pour créer le fichier **hmmdef**, on peut suivre les étapes suivantes :
- 

1. Copiez les trois lignes du fichier **proto** :

```
"~o"
"<STREAMINFO>"
"<VECSIZE>"
```

2. Pour chaque classe d'appareil :

- On copie les lignes du fichier **proto** de **"~h proto"** à **"<ENDHMM>"**.
- remplacer la ligne **"~h proto"** par **"~h label"** de l'appareil **"**, ou **" label** de l'appareil **"** représente la classe de l'appareil électrique.
- On répète cette étape pour toutes les classes.

- **Macros** : (voire la figure (III.10)).

Le fichier macro est constitué du fichier **vFloors** en lui ajoutant la ligne suivante :

```
~o <VECSIZE> 250 <MFCC_E_D_A>
```

```
Drill
Fan
Grinder
Hairdryer
Hedgetrimmer
Lamp
Paintstripper
Planner
Router
Sander
Saw
Vaccumcleaner
```

Figure III 8 : illustration du fichier "liste\_models"

```
#!MLF!#
"/scenarioC1_1.lab"
Drill
.
"/scenarioC1_3.lab"
Drill
.
"/scenarioC1_5.lab"
Drill
.
"/scenarioC1_7.lab"
Drill
.
```

Figure III 9 : : illustration du fichier trainmlf

```

~o <VecSize> 250 <USER>
~v varFloor1
<Variance> 250
2.467909e-005 1.666053e-005 3.582377e-003 1.364480e-005 1.292672e-005 1.950900e-006 3.257861e-004 1.039581e-006 3.619383e-006
2.920722e-007 9.240366e-007 1.613381e-007 5.698692e-006 1.160831e-007 2.165293e-007 6.647424e-008 1.257823e-006 5.551767e-008
3.949466e-007 3.046319e-008 4.095709e-008 2.728571e-008 1.462537e-007 2.373844e-008 3.358188e-008 2.116726e-008 1.517998e-007
1.823254e-008 1.088270e-007 1.529235e-008 3.217871e-008 1.544929e-008 8.099013e-008 1.298034e-008 2.882230e-008 1.152117e-008
2.128765e-008 9.620687e-009 4.437280e-008 7.862476e-009 1.537439e-008 9.682298e-009 3.290808e-008 8.060490e-009 1.347443e-008
6.633800e-009 1.036927e-008 5.972854e-009 3.087707e-008 4.917830e-009 7.513507e-009 4.491592e-009 1.380725e-008 4.439663e-009
1.149669e-008 8.488906e-009 6.070181e-009 5.748046e-009 9.516484e-009 5.490008e-009 5.183872e-009 4.966844e-009 8.150359e-009
3.736932e-009 6.777647e-009 3.465488e-009 3.758772e-009 3.591373e-009 5.976626e-009 2.889528e-009 3.530176e-009 3.545598e-009
3.634463e-009 3.135731e-009 4.482344e-009 3.748326e-009 3.147868e-009 2.462667e-009 3.801941e-009 3.202573e-009 3.036024e-009
2.403642e-009 2.863447e-009 3.033733e-009 3.364240e-009 2.505958e-009 3.416215e-009 3.547256e-009 3.099088e-009 2.153773e-009
3.251525e-009 2.362563e-009 2.305497e-009 2.163109e-009 2.700009e-009 2.050693e-009 2.213888e-009 2.115711e-009 2.735735e-009
2.073065e-009 2.406648e-009 2.018003e-009 2.234995e-009

```

Figure III 10 : illustration du fichier macro

```

~o
<STREAMINFO> 1 250
<VECSIZE> 250<NULLD><USER><DIAGC>
~h "Drill"
<BEGINHMM>
<NUMSTATES> 5
<STATE> 2
<MEAN> 125
4.430586e-002 1.817954e-002 1.956903e+000 1.731328e-002 2.455438e-002 8.625033e-003 3.862752e-001 6.651890e-003 1.220312e-002 4.977957e-003 1.083082e-001
4.441328e-003 6.936706e-003 3.306855e-003 4.831773e-002 3.400797e-003 4.635404e-003 2.640079e-003 3.046481e-002 2.747881e-003 3.333334e-003 2.288420e-003
1.733395e-002 2.037284e-003 2.472467e-003 2.002780e-003 7.750053e-003 1.937090e-003 2.267430e-003 1.879067e-003 8.490074e-003 1.847953e-003 2.341021e-003
1.905161e-003 5.593611e-003 1.865602e-003 2.460899e-003 1.901216e-003 5.012873e-003 1.722558e-003 2.296135e-003 1.623999e-003 4.599334e-003 1.503780e-003
1.981515e-003 1.460392e-003 3.610060e-003 1.265027e-003 1.685399e-003 1.494530e-003 3.104323e-003 1.301896e-003 1.594158e-003 1.359000e-003 3.091017e-003
1.240073e-003 1.494177e-003 1.137348e-003 2.839599e-003 1.035395e-003 1.304214e-003 9.835702e-004 2.197070e-003 1.015227e-003 1.224079e-003 9.964098e-004
2.050526e-003 1.208909e-003 1.287960e-003 1.158301e-003 1.908456e-003 1.116350e-003 1.232151e-003 1.139703e-003 1.718633e-003 1.025628e-003 1.070086e-003
1.004899e-003 1.648263e-003 9.584294e-004 1.012548e-003 9.594053e-004 1.576216e-003 9.099382e-004 1.035556e-003 9.892567e-004 1.471004e-003 8.798701e-004
1.040711e-003 9.461537e-004 1.358833e-003 9.739249e-004 9.736451e-004 8.194465e-004 1.189789e-003 8.845854e-004 9.396202e-004 7.903337e-004 1.149268e-003
7.703717e-004 8.963043e-004 8.647837e-004 1.150697e-003 7.870540e-004 9.766354e-004 9.310180e-004 1.101522e-003 7.134061e-004 8.754844e-004 8.267291e-004
1.106379e-003 6.887471e-004 7.583851e-004 7.195456e-004 1.034570e-003 6.521042e-004 7.399877e-004 6.944724e-004 1.016744e-003 6.518994e-004 7.228018e-004
6.516277e-004 9.356680e-004 6.334705e-004 7.071255e-004
<VARIANCE> 125

```

Figure III 11 : illustration du fichier hmmdef

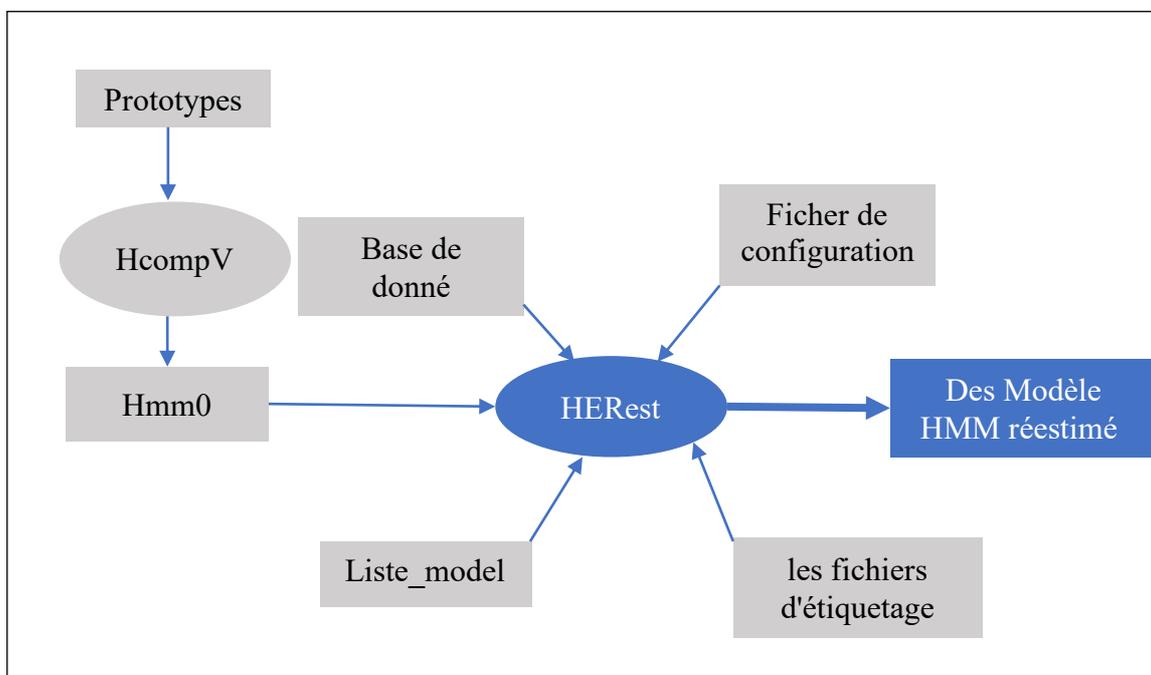


Figure III 12 : L'entraînement des modèles HMMs par l'outil HRest.

### III.2.3.3 L'augmentation de nombre de gaussiennes :

L'outil **HHed** applique des modifications spécifiées dans le fichier d'édition (`hed1`), aux modèles HMM réestimés précédemment. Le fichier de macro (`hmm3\macros`) et le fichier de définition des HMM (`hmm3\hmmdef`) fournissent les modèles réestimés. Les modèles HMM modifiés qui en résultent sont ensuite enregistrés dans le répertoire de sortie (`hmm4`) après ils seront réestimés suite à trois itérations en utilisant l'outil **HERest**.

Ce processus est répété en augmentant respectivement le nombre de gaussiennes à 2, 4, 8, 16, 32, 64.

```
HHed -H hmm3\macros -H hmm3\hmmdef -M hmm4 hed1 modelist
```

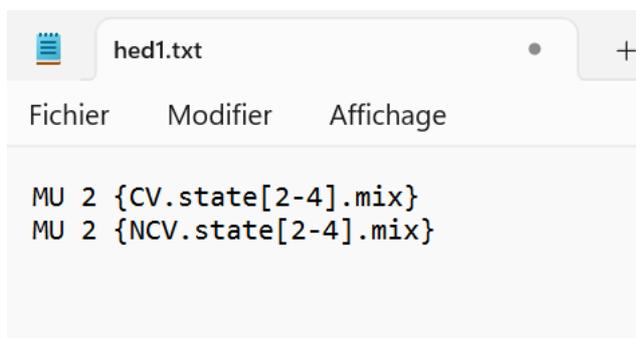
-**H** `hmm3\macros` : Spécifie le répertoire contenant les fichiers **macros**.

-**H** `hmm3\hmmdef` : Spécifie le répertoire contenant le fichier **hmmdef**

- **M** `hmm4` : spécifie le répertoire de sortie dans lequel les modèles HMM modifiés seront stockés.

- **Hed1** : est un fichier qui contient les instructions de l'augmentation du nombre

De Gaussiennes(voire la figure (III.13)).



```
hed1.txt
Fichier  Modifier  Affichage
MU 2 {CV.state[2-4].mix}
MU 2 {NCV.state[2-4].mix}
```

Figure III 13 : illustration du fichier `hed1`

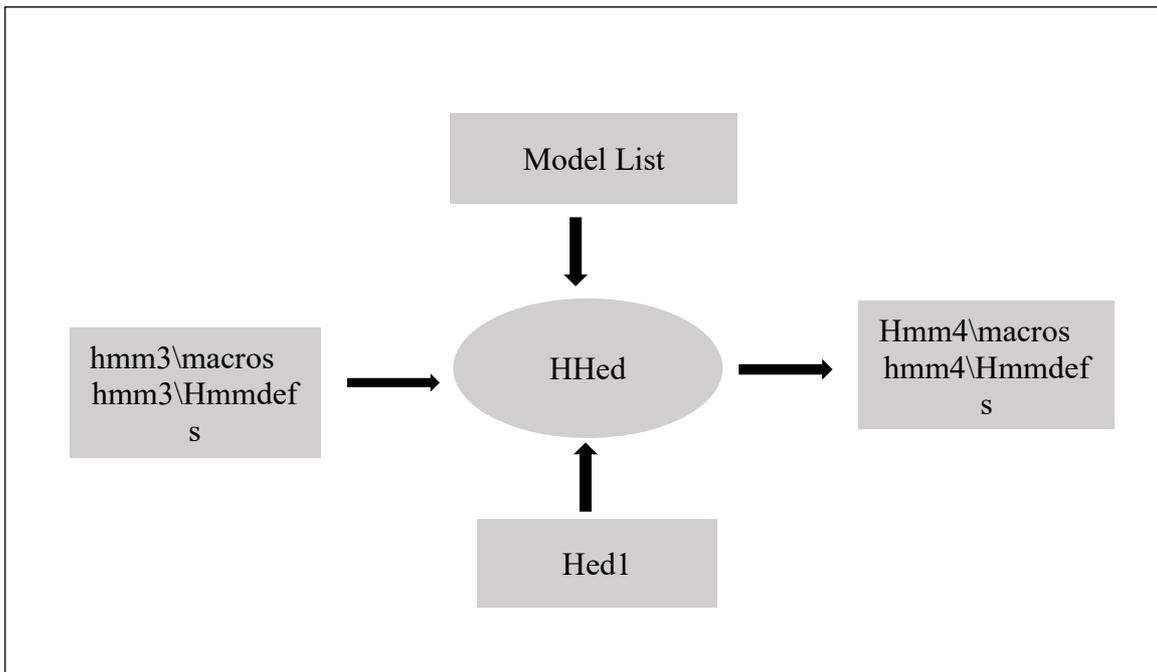


Figure III 14 : Décodage par l'outil HVite.

### III.2.3.4 Définition du dictionnaire et du modèle du Langage :

La commande **HParse** fait partie du Hidden Markov Model Toolkit (HTK) et est utilisée pour compiler une grammaire sans contexte dans une structure de réseau.

```
HParse -C gram.txt -T 1 net.slf
```

- **C gram.txt** : spécifie le fichier de configuration (gram.txt) contenant les règles de grammaire et les paramètres (voire la figure (III.15)).
- **net.slf** : spécifie le nom du fichier de sortie pour la structure de réseau compilée.



Figure III 15 : Illustration du fichier grammaire

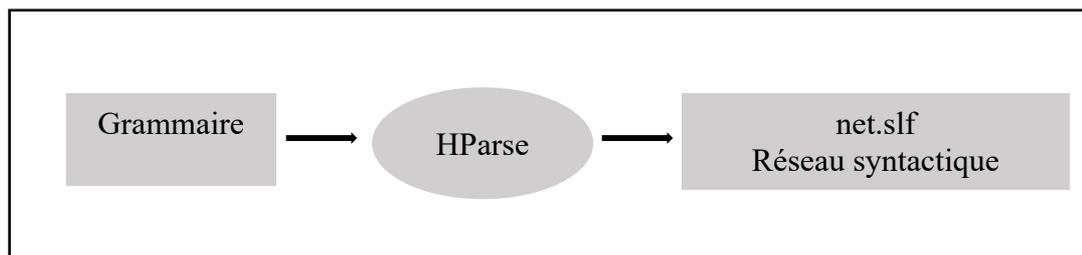


Figure III 16 : compilation d'une grammaire dans une structure de réseau en utilisant Hparse

En exécutant cette commande **HParse**, l'outil lit les règles de grammaire et les paramètres du fichier de configuration (gram.txt) et les compile dans une structure de réseau. La structure de réseau résultante est ensuite enregistrée dans le fichier de sortie spécifié **net.slf**.

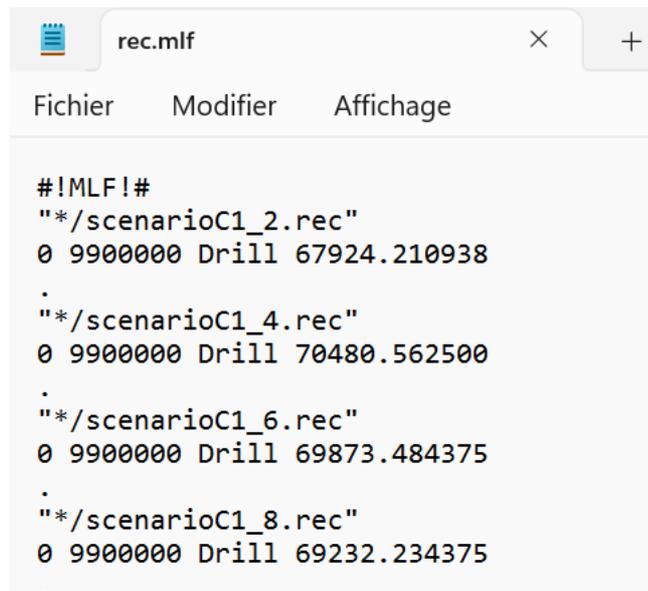
### III.2.4 Classification

La commande **Hvite** est utilisée pour effectuer l'étape de décodage, Elle utilise l'algorithme de Viterbi.

```

HVite -D -C ' conf_user , '-l ','*',' -S ' testlist,' -H ' mac_file, '-H ' mod_file,' -i
' file_mlf_test , '-w ' wnetf , ' dictionary_file, ' list_model
  
```

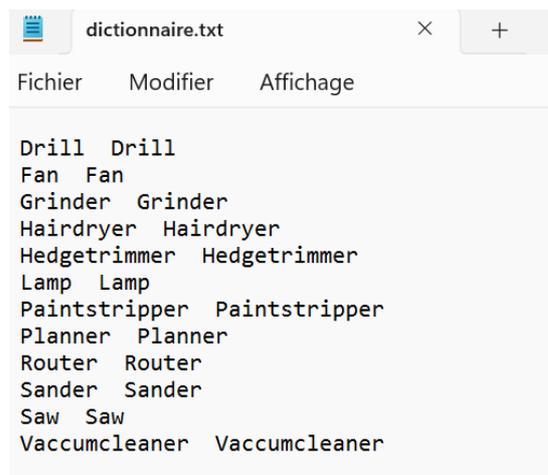
- - **testlist** : spécifie le fichier (testlist) contenant la liste des fichiers de base de données de test
- - **H mac\_file -H mod\_file** : Spécifie le fichier de macros (mac\_file) et le Hmmdefs (mod\_file).
- - **i fichier\_mlf\_test** : Spécifie le fichier de sortie (file\_mlf\_test) dans lequel les étiquettes reconnues ou la sortie décodée seront stockées.
- - **w wnetf** : Spécifie le fichier de réseau de mots (wnetf) représentant le modèle de langue.
- **dictionnaire\_fichier** : Spécifie le fichier (dictionnaire) contient le dictionnaire ou le vocabulaire(III.18).
- **liste\_modèle** : et le fichier (liste\_modèle) contenant la liste des noms de modèles ou des classes à utiliser pour la reconnaissance.



The screenshot shows a text editor window with a single tab titled 'rec.mlf'. The window has a menu bar with 'Fichier', 'Modifier', and 'Affichage'. The content of the file is as follows:

```
#!MLF!  
"/scenarioC1_2.rec"  
0 9900000 Drill 67924.210938  
.  
"/scenarioC1_4.rec"  
0 9900000 Drill 70480.562500  
.  
"/scenarioC1_6.rec"  
0 9900000 Drill 69873.484375  
.  
"/scenarioC1_8.rec"  
0 9900000 Drill 69232.234375  
.
```

Figure III 17 : Illustration du fichier rec.mlf



The screenshot shows a text editor window with a single tab titled 'dictionnaire.txt'. The window has a menu bar with 'Fichier', 'Modifier', and 'Affichage'. The content of the file is as follows:

```
Drill Drill  
Fan Fan  
Grinder Grinder  
Hairdryer Hairdryer  
Hedgetrimmer Hedgetrimmer  
Lamp Lamp  
Paintstripper Paintstripper  
Planner Planner  
Router Router  
Sander Sander  
Saw Saw  
Vaccumcleaner Vaccumcleaner
```

Figure III 18 : Illustration du fichier dictionnaire

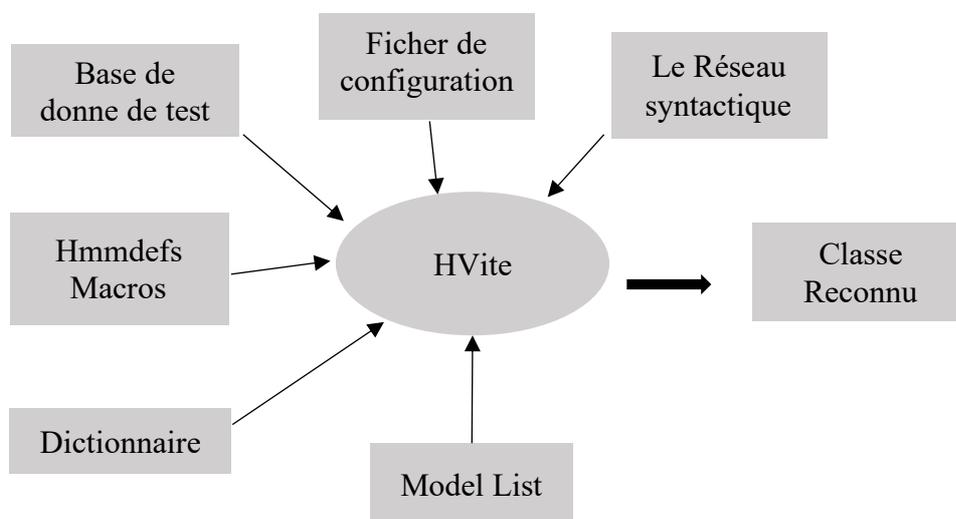


Figure III 19 : Décodage par l'outil HVite

En exécutant cette commande **HVite**, l'outil effectue le décodage à l'aide de l'algorithme de Viterbi sur la base des paramètres de configuration spécifiés, des modèles entraînés, du modèle de langage et des fichiers de caractéristiques d'entrée ou des échantillons. Les étiquettes reconnues ou la sortie décodée seront stockées dans le fichier de sortie (file\_mlf\_test). **HVite** utilise le dictionnaire, le réseau de mots et la liste de modèles pour effectuer la reconnaissance et comparer les signaux d'entrée aux modèles entraînés.

#### III.2.4.1 Évaluation du système de reconnaissance :

```
HResults -I ' file_mlf_testr, '-L ', cheminbaselab, ',list_model ', ' file_mlf_test,' >> ',
results]
```

La commande HResults est utilisée pour évaluer les performances d'un système de reconnaissance en comparant les étiquettes reconnues ou la sortie avec les étiquettes de référence.

- **-I fichier\_mlf\_testr** : spécifie le fichier d'étiquettes de référence (file\_mlf\_testr) contenant les étiquettes correctes pour l'évaluation.
- **-L cheminbaselab** : spécifie le répertoire (cheminbaselab) dans lequel se trouvent les fichiers d'étiquettes de référence.

- **list\_model** : spécifie le fichier contenant la liste des noms de modèles ou de classes utilisés pour la reconnaissance.
- **file\_mlf\_test** : spécifie le fichier d'étiquettes de sortie (file\_mlf\_test) obtenu par le système de reconnaissance.
- **>> résultats** : redirige la sortie de l'évaluation vers le fichier nommé results pour stocker les métriques d'évaluation.

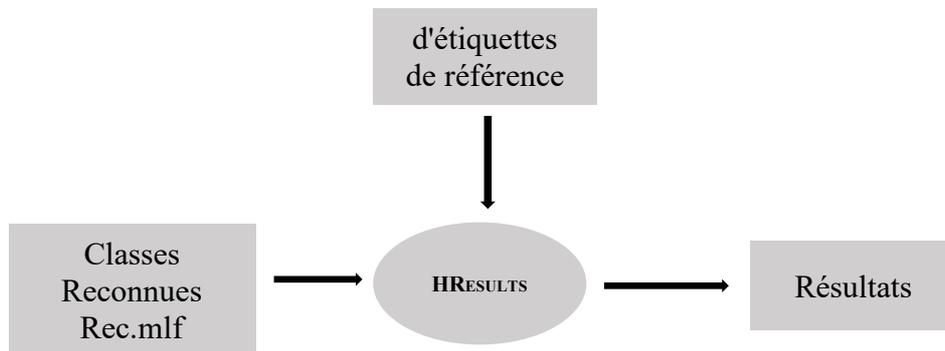


Figure III 20 : Évaluation des résultats par l'outil HResults

En exécutant la commande **HResults**, l'outil compare les étiquettes reconnues ou la sortie du système de reconnaissance avec les étiquettes de référence spécifiées dans le fichier **mlf\_testr**. Il calcule diverses mesures d'évaluation, telles que la précision, les taux d'erreur ou les matrices de confusion, sur la base de la comparaison.

### III.3 Critères d'évaluation

Le paramètre "correct" donne le taux d'identification (TI) des signaux reconnus correctement. Il est calculé par la formule suivante :

$$TI = \frac{H}{N} \times 100\%$$

- **H** : Nombre des signaux de test reconnus correctement.
- **N** : Nombre total des signaux de test.

### III.4 Expérience et résultats

#### III.4.1 Expérience 1

La première expérience de notre étude consiste à évaluer les performances de notre système standard basé sur des modèles de Markov cachés (HMM) de 5 états associés chacun à un modèle GMM de 4 gaussiennes entraînés sur des séquences de vecteurs de 250 coefficients STFS.

Les performances sont évaluées en termes de taux de reconnaissance (TI). Les résultats de cette expérience ont montré une bonne précision du système avec un TI égal à 96,67 % (voir la figure III 21).

```

===== HTK Results Analysis =====
Date: Sun May 14 20:56:09 2023
Ref : C:\HMM\Test.mlf
Rec : C:\HMM\rec.mlf
----- Overall Results -----
SENT: %Correct=96.67 [H=406, S=14, N=420]
WORD: %Corr=96.67, Acc=96.67 [H=406, D=0, S=14, I=0, N=420]
=====
    
```

Figure III 21 : Résultat de la première expérience

#### III.4.2 Expérience 2

L'objectif principal de cette expérience est d'optimiser la topologie des modèles HMM en cherchant le nombre optimal d'états (Nbr\_e) et le nombre optimal de gaussiennes (Nbr\_g) permettant d'améliorer les performances du système. Les performances du système sont évaluées en variant systématiquement le nombre d'états ainsi que le nombre de gaussiennes. Le tableau (III-1) illustre le taux TI pour différentes valeurs de Nbr\_e et Nbr\_eg.

Tableau III - 1 : Taux TI pour différentes valeurs de Nbr\_e et Nbr\_eg.

Nbr_e \ Nbr_g	4	5	6	7	8	9	10	11	12
2	77.62	91.67	95.48	95.71	96.90	96.43	97.62	96.43	95.71
4	82.62	96.67	95.95	<b>98.81</b>	98.81	98.33	98.57	97.14	97.38
8	81.90	96.67	95.95	98.81	98.33	98.10	98.57	97.14	98.10
16	81.67	96.90	95.95	98.81	98.33	98.10	98.33	97.38	98.33

<b>32</b>	81.67	96.90	95.95	98.57	98.33	98.10	98.33	97.38	98.33
<b>64</b>	81.67	96.90	95.95	98.57	98.10	98.10	98.33	97.38	98.33

Les résultats nous ont montrés une amélioration des performances avec un TI égal 98,81 % en choisissant un nombre d'états égal à 7 et un nombre de gaussiennes égal à 4. Ce choix de configuration sera pris en considération aux expériences suivantes.

### III.4.3 Expérience 3

Cette expérience a pour objectif d'évaluer l'impact de la dimensionnalité des vecteurs de coefficients STFS sur la précision du système IAED. Dans cette expérience, nous évaluons le taux TI en faisant varier le nombre de coefficients STFS de 1 à 250. La figure suivante montre le taux TI en fonction du nombre de coefficients.

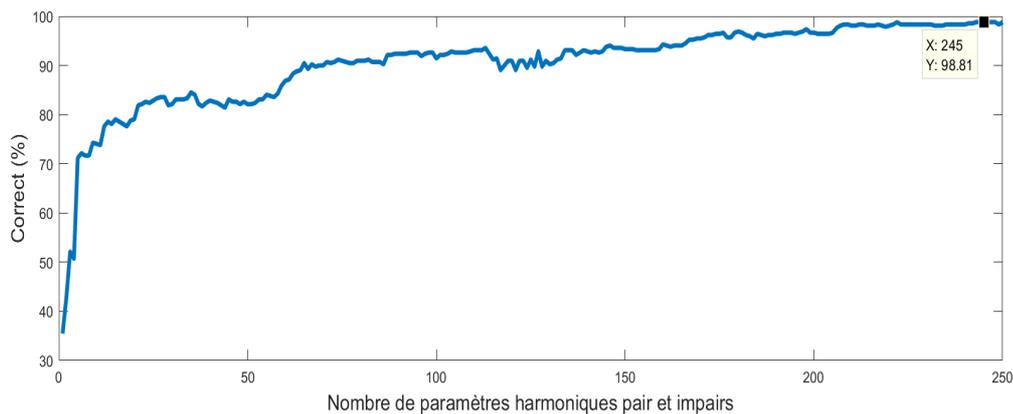


Figure III 22 : Taux d'identification en fonction du nombre des coefficients harmoniques

La figure (29) nous permet de remarquer les points suivants :

- Du coefficient 1 au coefficient 6, le TI augmente rapidement, passant de 35,48 % à 72,14 %. Cet ensemble de coefficients expliquent les différentes classes avec un pourcentage de 73.33% par rapport au taux TI correspondant aux 250 coefficients. Ce résultat montre la pertinence de sous ensemble de coefficients pour cette tâche d'identification.
- Lorsque le nombre de coefficients continue d'augmenter, la précision continue également de monter malgré quelques fluctuations. La précision atteint sa valeur

maximale au coefficient 243, puis elle se stabilise approximativement jusqu'à le nombre de coefficients atteint 250.

- Les résultats de cette expérience montrent qu'un sous ensemble réduit de coefficients STFS contribuent de manière significative à cette tâche d'identification. Toutefois, à partir d'un certain seuil de TI, des coefficients supplémentaires peuvent ne pas apporter d'amélioration significative sur les performances.

### III.4.4 Expérience 4

Dans la quatrième expérience, notre objectif est d'évaluer l'impact de l'utilisation de coefficients harmoniques pairs et impairs sur les performances du système. En incluant de manière sélective les coefficients pairs ou impairs tout en excluant les autres coefficients, nous pouvons mieux comprendre l'importance des coefficients pairs et impairs dans l'identification des appareils électriques.

La figure (III 23) illustre le taux TI en fonction du nombre de coefficients harmoniques de rang impair.

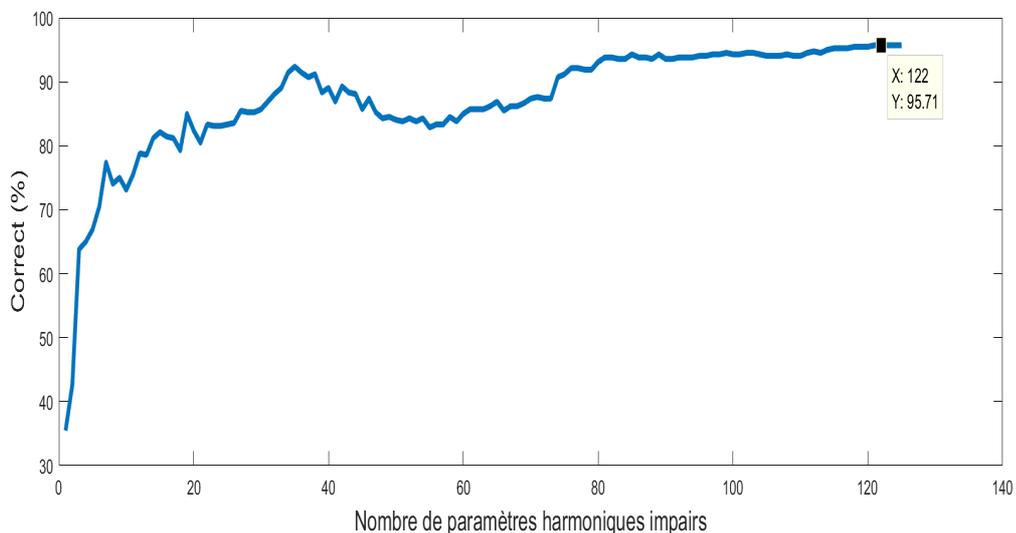


Figure III 23 : Taux d'identification en fonction du nombre de coefficients harmoniques du rang impair

Pour les sept premiers coefficients impairs, on observe une augmentation significative du taux TI atteignant 77,38 %. Ce taux atteint sa valeur maximale de 95,71 % au 121<sup>ème</sup> coefficient.

La figure (III 24) illustre le taux TI en fonction du nombre de coefficients harmoniques de rang pair. La courbe de TI augmente d'une manière lente par rapport à celle de la figure précédente en atteignant de 78,1 % au 32 coefficients. La précision maximale de **91,9 %** a été atteinte au 125<sup>ème</sup> coefficient. Ce résultat explique clairement la pertinence des coefficients de rang impair par rapport à ceux de rang pair. Ainsi, les coefficients de rang impair sont des caractéristiques discriminantes des différentes classes d'appareils électriques domestiques.

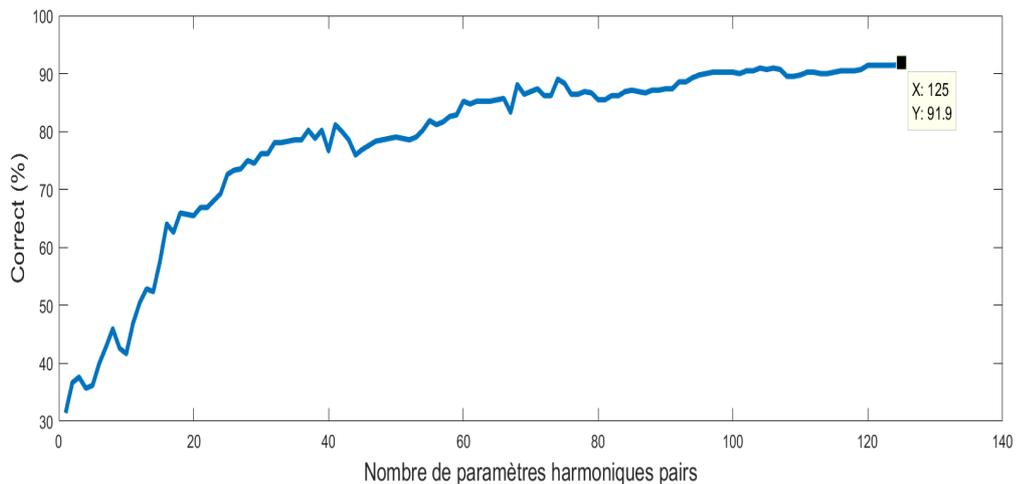


Figure III 24 : Taux TI en fonction du nombre de paramètres harmoniques de rang pair

### III.5 Conclusion

Dans ce chapitre, nous avons proposé un système d'identification des appareils électriques basé sur les modèles HMM combinés avec l'analyse harmonique. Ce système est implémenté en utilisant la boîte à outils HTK sous l'environnement Matlab 2017. Différentes expériences ont été effectuées pour répondre sur des questions posées préalablement sur la conception d'un système performant.

Les résultats nous ont montrés que la meilleure configuration du système IAED est obtenue en choisissant des modèles HMM de 7 états associé chacun à modèle GMM de 4 composantes gaussiennes. Cette configuration a permis d'obtenir un taux TI de 98.81%. De plus, les 6 premiers coefficients contribuent fortement à expliquer les différentes classes d'appareils.

Nous avons également étudié l'importance des coefficients pairs et impairs. Les résultats nous ont montré la pertinence des coefficients de rang impair par rapport à ceux du rang pair. Ces résultats confirment les résultats obtenus en [12].

---

### *Conclusion Générale*

---

Dans notre étude, nous avons proposé un système d'identification des appareils électriques domestiques en utilisant le classificateur HMM appliqué à des séquences de vecteurs de paramètres STFS. Ce système est implémenté en utilisant la boîte à outils HTK sous l'environnement Matlab. Les performances du système sont évaluées en termes de taux de classification, en utilisant la base de données COOLL.

La conception du système se déroule en une phase d'apprentissage des modèles HMM associés aux modèles GMM entraînés à l'aide des données étiquetées, et une phase de test permettant d'évaluer les performances du système IAED implémenté en utilisant la base de données COOLL. Une étape d'extraction de paramètres est effectuée pour chacune des phases en transformant chaque signal du courant électrique en une séquence de vecteurs de paramètres STFS.

Les résultats nous ont montrés l'efficacité du classificateur HMM avec un taux de classification maximal de 98,81% obtenu en utilisant 7 états associés chacun à un modèle GMM de 4 gaussiennes. Nous avons également analysé l'impact de la variation du nombre de coefficients STFS sur les performances du système et nous avons observé que l'augmentation du nombre de coefficients améliore la précision du système. Les résultats nous ont confirmés la pertinence des harmoniques impaires discutés dans plusieurs travaux de recherches [5][7][26].

En conclusion, l'optimisation de la configuration des modèles HMM, l'augmentation du nombre de coefficients STFS dans une marge raisonnable et la priorité donnée à l'utilisation des coefficients impairs permettent d'obtenir des performances supérieures aux celles certains des systèmes IAED existant.. Les résultats de notre étude valident et soutiennent des travaux de recherches précédents [5][7][26]. Enfin, on propose comme perspectives d'utiliser d'autres répartitions des bases de données d'apprentissage et de test extraites de la base COOLL, ainsi que d'utiliser d'autres descripteurs de paramètres et d'autres classificateurs.

## Références Bibliographiques

- [1] J. Ouzine, M. Marzouq, S. Bennani, L. Khadija, et H. el fadili, *Overview of Non-Intrusive Load Monitoring: Probabilistic and Artificial Intelligence approaches*, vol. 351. 2022.
- [2] G. R. Barai, S. Krishnan, et B. Venkatesh, « Smart metering and functionalities of smart meters in smart grid - a review », in *2015 IEEE Electrical Power and Energy Conference (EPEC)*, oct. 2015, p. 138-145.
- [3] M. N. Meziane, « Identification d'appareils électriques par analyse des courants de mise en marche », phd thesis, Université d'Orléans, 2016.
- [4] R. Lukaszewski, K. Liszewski, et W. Winiecki, *Methods of electrical appliances identification in systems monitoring electrical energy consumption*. 2013, p. 14.
- [5] M. Nait-Meziane, A. Hacine-Gharbi, P. Ravier, G. Lamarque, J.-C. Le Bunetel, et Y. Raingeaud, *HMM-based Transient and Steady-state Current Signals Modeling for Electrical Appliances Identification*. 2016, p. 677.
- [6] GHALEM Dalila , ZEGADI Yousra, « Identification des appareils électriques basée sur les modèles GMM », Master thesis, Faculté des Sciences et Technologies, 2021.
- [7] M. Ghezzali Fatah, « Extraction et sélection des paramètres pertinents appliquées à l'identification des appareils électriques », Phd thesis, faculté des sciences et de la technologie univ bba, 2022.
- [8] Y. F. Wong, Y. Ahmet Şekerciöglu, T. Drummond, et V. S. Wong, « Recent approaches to non-intrusive load monitoring techniques in residential settings », in *2013 IEEE Computational Intelligence Applications in Smart Grid (CIASG)*, avr. 2013, p. 73-79.
- [9] G. W. Hart, « Nonintrusive appliance load monitoring », *Proc. IEEE*, vol. 80, n° 12, p. 1870-1891, déc. 1992.
- [10] A. Pasquet, « NILM Algorithms General Comparison and Test of Adaptability », Master thesis, Universitat Politècnica de Catalunya, 2020.
- [11] B. Nour El Houda, Regaigui Abir « Identification des appareils électriques basée sur le classificateur KNN combiné avec la règle de vote », Master thesis, Faculté des Sciences et Technologies, 2020.
- [12] A. Zoha, A. Gluhak, M. A. Imran, et S. Rajasegarar, « Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey », *Sensors*, vol. 12, n° 12, Art. n° 12, déc. 2012.
- [13] R. Reddy, V. Garg, et V. Pudi, « A feature fusion technique for improved non-

intrusive load monitoring », *Energy Inform.*, vol. 3, n° 1, p. 9, sept. 2020.

[14] D. P. B. Renaux *et al.*, « A Dataset for Non-Intrusive Load Monitoring: Design and Implementation », *Energies*, vol. 13, n° 20, Art. n° 20, janv. 2020.

[15] « Machine à vecteurs de support (SVM) : définition et cas d'usage », 24 janvier 2022. <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501879-machine-a-vecteurs-de-support-svm-definition-et-cas-d-usage/> (consulté le 05 juin 2023).

[16] « What is a Decision Tree | IBM ». <https://www.ibm.com/topics/decision-trees> (consulté le 05 juin 2023).

[17] R. Bala et D. Kumar, « Classification Using ANN : A Review », 2017.

[18] A. Kumar, « Hidden Markov Models: Concepts, Examples », *Data Analytics*, 27 janvier 2023. <https://vitalflux.com/hidden-markov-models-concepts-explained-with-examples/> (consulté le 25 mai 2023).

[19] T. Picon *et al.*, « COOLL: Controlled On/Off Loads Library, a Public Dataset of High-Sampled Electrical Signals for Appliance Identification ». arXiv, 17 novembre 2016.

[20] M. Sedira et A. Felkaoui, « Application des Modèles de Markov Cachés Dans la Classification des Défauts de Machines Tournantes » 2011.

[21] T. Li, « Study on a CNN-HMM Approach for Audio-Based Musical Chord Recognition », *J. Phys. Conf. Ser.*, vol. 1802, p. 032033, mars 2021.

[22] Charif rima, Mohamadi faiza « Reconnaissance acoustique des émotions basée sur les modèles de markov cachés HMM » Master thesis, Faculté des Sciences et Technologies, 2018.

[23] A. Kumar, « Gaussian Mixture Models: What are they & when to use? », *Data Analytics*, 8 mai 2023. <https://vitalflux.com/gaussian-mixture-models-what-are-they-when-to-use/> (consulté le 05 juin 2023).

[24] O. C. Ibe, « 14 - Hidden Markov Models », in *Markov Processes for Stochastic Modeling (Second Edition)*, O. C. Ibe, Éd., Oxford: Elsevier, 2013, p. 417-451.

[25] Amina Makhoulf « Reconnaissance Automatique de la Parole par les HMM en Milieu Bruité : Contribution par paramétrisation acoustique robuste » 2016.

[26] A. Hacine-Gharbi, P. Ravier, et M. Nait-Meziane, « Relevant harmonics selection based on mutual information for electrical appliances identification », *Int. J. Comput. Appl. Technol.*, vol. 62, p. 102-115, janv. 2020.