

UNIVERSITÉ DE SHERBROOKE  
Faculté de génie  
Département de génie électrique et informatique

# Détection et identification des activités de la vie quotidienne à l'aide d'un unique microphone par pièce

Mémoire de maîtrise  
Spécialité : génie électrique

Thomas Romain CAMIER

Jury : Roch LEFEBVRE (directeur)  
Philippe MABILLEAU (codirecteur)  
Jean ROUAT  
Hélène PIGOT

Sherbrooke (Québec) Canada

Janvier 2011

IV-2/09



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 978-0-494-79749-5*  
*Our file* *Notre référence*  
*ISBN: 978-0-494-79749-5*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**



*À mes grands-parents*



# RÉSUMÉ

De nos jours, le maintien à domicile des personnes éprouvant des troubles cognitifs est un réel défi. En effet, les centres spécialisés pour ces personnes ne sont pas suffisants et manquent parfois de ressources. C'est là que l'assistance à domicile intervient pour permettre à ces personnes de conserver un certain niveau d'autonomie avec le moins d'intrusion possible dans leur vie quotidienne. Pour répondre à cette demande, le laboratoire *DOMUS* s'est équipé d'un appartement intelligent pilote muni des Nouvelles Technologies de l'Information et de Communication (*NTIC*). L'infrastructure du laboratoire est composée entre autres de capteurs infrarouges, de capteurs de pression, d'écrans tactiles et de microphones qui ont pour but de rendre compte des activités de la vie quotidienne (*AVQ*) réalisées par la personne au sein de l'habitat. La présente étude rapporte les résultats d'une partie du système d'assistance à domicile permettant l'identification des *AVQ* par une personne au sein d'un appartement intelligent, à l'aide d'un unique microphone par pièce. Dès lors, on s'intéresse à la détection et la reconnaissance des événements sonores générés par la réalisation des *AVQ* de l'habitant. Les événements sonores testés correspondent aux activités qui ont lieu dans une cuisine, telle que des bruits de cuisson, d'écoulement d'eau, d'ustensile, de casserole, de vaisselle et de claquements de porte.

**Mots-clés** : troubles cognitifs, assistance à domicile, appartement intelligent, identification des *AVQ*, microphone, signal sonore



# TABLE DES MATIÈRES

RÉSUMÉ.....	v
CHAPITRE 1. INTRODUCTION.....	1
1.1 Contexte.....	1
1.2 Problématique.....	3
1.3 Définition du projet de recherche.....	4
1.4 Objectif du projet de recherche.....	5
1.5 Configuration de l'environnement de recherche.....	6
1.6 Plan du mémoire.....	8
CHAPITRE 2. ÉTAT DE L'ART.....	9
2.1 Détection des événements sonores.....	9
2.1.1 Méthodes d'analyse temporelle.....	9
2.1.2 Décomposition de Fourier.....	10
2.1.3 Décomposition en ondelettes.....	11
2.1.4 Méthodes de détection des événements sonores.....	11
2.2 Reconnaissance des activités sonores.....	12
2.2.1 Méthodes de paramétrage temporel.....	12
2.2.2 Méthodes de paramétrage fréquentiel.....	13
2.2.3 Création des dictionnaires : l'algorithme des k-moyennes.....	14
2.2.4 Classification des événements sonores.....	14
CHAPITRE 3. CONCEPTION.....	17
3.1 Présentation des signaux étudiés.....	17
3.2 Segmentation du signal sonore.....	18
3.2.1 Fenêtrage du signal sonore.....	19
3.2.2 Chevauchement temporel.....	20
3.2.3 Définition d'une trame sonores.....	21
3.3 Méthodes utiles à la caractérisation des signaux sonores.....	21
3.3.1 L'énergie temporelle.....	21
3.3.2 Le taux de passage par zéro.....	22



3.3.3	La transformée de Fourier .....	24
3.3.4	La transformée de Fourier discrète.....	25
3.3.5	La densité spectrale de puissance.....	26
3.3.6	L'énergie spectrale .....	28
3.3.7	Le centre de masse spectrale .....	30
3.3.8	La largeur de bande spectrale.....	32
3.3.9	La fréquence de coupure spectrale .....	33
3.3.10	La densité spectrale d'énergie en bande critique .....	34
3.3.11	La transformée en ondelettes.....	36
3.3.12	Transformée en ondelettes continues .....	38
3.3.13	Transformée en ondelettes discrètes .....	39
3.3.14	Énergie globale des coefficients en ondelettes discrètes.....	41
3.4	Méthodes utiles à la détection des accroches sonores.....	43
3.4.1	Flux.....	43
3.4.2	L'opérateur « Teager-Kaiser » .....	45
3.4.3	Calcul du seuil de détection via un circuit intégrateur .....	47
3.5	Méthodes utiles à la reconnaissance des événements sonores .....	48
3.5.1	Création des références via l'algorithme k-moyennes .....	49
3.5.2	Classification des événements sonores .....	50
CHAPITRE 4. RÉALISATION .....		53
4.1	Système de détection des objets sonores.....	53
4.1.1	Fonction de seuillage absolu .....	53
4.1.2	Fonction de seuillage des accroches sonores .....	54
4.1.3	Algorithme de détection des objets sonores .....	56
4.2	Système de reconnaissance des objets sonores .....	57
4.2.1	Calcul des paramètres de classification.....	57
4.2.2	Création des dictionnaires de références .....	59
4.2.3	Classification des objets sonores.....	59
CHAPITRE 5. TEST ET ANALYSE.....		61
5.1	Test du système de détection des objets sonores.....	61
5.2	Résultats de la 1 <sup>ère</sup> simulation .....	65

5.3	Résultats de la 2 <sup>ème</sup> simulation .....	69
5.4	Analyse des résultats des simulations.....	71
5.5	Résultats du système de reconnaissance des objets sonores .....	73
CHAPITRE 6. CONCLUSION .....		77
ANNEXE A – RÉSULTATS D’EXPÉRIMENTATION .....		81
LISTE DES RÉFÉRENCES.....		83



# LISTE DES FIGURES

FIGURE 1-1 : PLAN DE L'APPARTEMENT INTELLIGENT DOMUS AVEC LA NATURE ET LA LOCALISATION DES DIFFÉRENTS CAPTEURS .....	2
FIGURE 1-2 : SCHÉMA DE PRINCIPE DE LA RECONNAISSANCE DES ACTIVITÉS DE LA VIE QUOTIDIENNE .....	3
FIGURE 1-3 : SCHÉMA DE PRINCIPE DE LA RECONNAISSANCE D'UNE AVQ À L'AIDE DU SON .....	5
FIGURE 1-4 : SCHÉMA DE PRINCIPE DE LA RECONNAISSANCE DES ACTIVITÉS DE LA VIE QUOTIDIENNE BASÉE SUR L'ANALYSE SONORE.....	5
FIGURE 1-5 : PHOTO DE L'ARCHITECTURE DE LA CUISINE .....	7
FIGURE 3-1 : ALLURE TEMPORELLE DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D).....	17
FIGURE 3-2 : ALLURE TEMPORELLE DE LA FENÊTRE DE « <i>HAMMING</i> » .....	19
FIGURE 3-3 : PRINCIPE DU CHEVAUCHEMENT TEMPOREL .....	20
FIGURE 3-4 : TRACÉ DE L'ÉNERGIE TEMPORELLE DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	22
FIGURE 3-5 : TRACÉ DU TAUX DE PASSAGE PAR ZÉRO DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D)23	23
FIGURE 3-6 : IMAGES DE LA DENSITÉ SPECTRALE DE PUISSANCE DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	27
FIGURE 3-7 : TRACÉ DE L'ÉNERGIE SPECTRALE DE LA <i>DSP</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D)29	29
FIGURE 3-8 : TRACÉ DU CENTRE DE MASSE DE LA <i>DSP</i> EN <i>HERTZ</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D).....	31
FIGURE 3-9 : TRACÉ DE LA LARGEUR DE BANDE DE LA <i>DSP</i> EN <i>HERTZ</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	32
FIGURE 3-10 : TRACÉ DE LA FRÉQUENCE DE COUPURE SPECTRALE DE LA <i>DSP</i> EN <i>HERTZ</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	34
FIGURE 3-11 : GRAPHIQUE DE LA LOI DE CORRESPONDANCE ENTRE <i>MEL</i> ET <i>HERTZ</i> .....	36
FIGURE 3-12 : IMAGES DE LA <i>DSP</i> EN 20 BANDES HERTZIENNES DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	37

FIGURE 3-13 : IMAGES DE LA <i>DSP</i> EN 20 BANDES CRITIQUES DE <i>MEL</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	37
FIGURE 3-14 : PRINCIPE DE LA DÉCOMPOSITION EN ONDELETTES PAR UNE BATTERIE DE FILTRES SELON L'ALGORITHME DE MALLAT. ....	40
FIGURE 3-15 : IMAGES DE LA DÉCOMPOSITION EN ONDELETTES DYADIQUES DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	41
FIGURE 3-16 : TRACÉ DE L'ÉNERGIE GLOBALE DES COEFFICIENTS EN ONDELETTES DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	42
FIGURE 3-17 : IMAGES DU FLUX DE LA <i>DSP</i> EN BANDES CRITIQUES DE <i>MEL</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	44
FIGURE 3-18 : IMAGE DU FLUX DE L'ÉNERGIE DES COEFFICIENTS EN ONDELETTES DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	45
FIGURE 3-19 : IMAGES DE LA <i>DSP</i> EN BANDES CRITIQUES DE <i>MEL</i> À LAQUELLE ON A APPLIQUÉ L'OPÉRATEUR <i>TG</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	46
FIGURE 3-20 : IMAGES DE L'ÉNERGIE DES COEFFICIENTS EN ONDELETTES À LAQUELLE ON A APPLIQUÉ L'OPÉRATEUR <i>TG</i> DES SIGNAUX DE MANIPULATION D'USTENSILE (A), DE CASSEROLE (B), DE CLAQUEMENT DE PORTE DE PLACARD (C) ET D'UN ÉCOULEMENT D'EAU DU ROBINET (D) .....	47
FIGURE 3-21 : SCHÉMA DE PRINCIPE DE LA CRÉATION DES DICTIONNAIRES PAR L'ALGORITHME DES <i>K-MOYENNES</i> .....	50
FIGURE 4-1 : SCHÉMA DE PRINCIPE DU CALCUL DES FONCTIONS DE SEUILLAGE DU SYSTÈME DE DÉTECTION DES OBJETS SONORES .....	55
FIGURE 4-2 : SCHÉMA DE PRINCIPE DE DÉTECTION DES ACCROCHES SONORES .....	56
FIGURE 4-3 : SCHÉMA DE PRINCIPE DE DÉTECTION DES CHUTES SONORES .....	57
FIGURE 4-4 : SCHÉMA DE PRINCIPE DU CALCUL DES PARAMÈTRES DE CLASSIFICATION .....	58
FIGURE 4-5 : SCHÉMA DE PRINCIPE DE LA CLASSIFICATION DES OBJETS SONORES .....	60
FIGURE 5-1 : ENREGISTREMENT TEST .....	62
FIGURE 5-2 : RÉSULTATS DE LA 1 <sup>ÈRE</sup> SIMULATION DE CAPTURE DES OBJETS SONORES SUR L'ENREGISTREMENT TEST .....	63
FIGURE 5-3 : RÉSULTATS DE LA 2 <sup>ÈME</sup> SIMULATION DE CAPTURE DES OBJETS SONORES SUR L'ENREGISTREMENT TEST .....	64
FIGURE 5-4 : CAPTURE D'UNE MANIPULATION D'ASSIETTE (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) POUR LA 1 <sup>ÈRE</sup> SIMULATION .....	66

FIGURE 5-5 : CAPTURE D'UN ÉCOULEMENT D'EAU (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) ET (D) POUR LA 1 <sup>ÈRE</sup> SIMULATION .....	67
FIGURE 5-6 : CAPTURE D'UN CLAQUEMENT DE PORTE (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) POUR LA 1 <sup>ÈRE</sup> SIMULATION .....	68
FIGURE 5-7 : CAPTURE D'UNE MANIPULATION D' ASSIETTE (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) POUR LA 2 <sup>ÈME</sup> SIMULATION .....	69
FIGURE 5-8 : CAPTURE D'UN ÉCOULEMENT D'EAU (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) ET (D) POUR LA 2 <sup>ÈME</sup> SIMULATION.....	70
FIGURE 5-9 : CAPTURE D'UN CLAQUEMENT DE PORTE (A) EN FONCTION DE $PARAM_{ABS}$ , $SEUIL_{ABS}$ (B) $PARAM_I$ , $SEUIL_I$ (C) POUR LA 2 <sup>ÈME</sup> SIMULATION.....	71
FIGURE A-1 : RÉSULTATS DE LA CAPTURE DES OBJETS SONORES SUR L'ENREGISTREMENT EXPÉRIMENTAL.....	82



# LISTE DES TABLEAUX

TABLEAU 5-1 : RÉSULTATS DU SYSTÈME DE RECONNAISSANCE DES OBJETS SONORES À L'AIDE DE L'ÉNERGIE DES COEFFICIENTS EN ONDELETTES .....	73
TABLEAU 5-2 : RÉSULTATS DU SYSTÈME DE RECONNAISSANCE DES OBJETS SONORES À L'AIDE DE LA <i>DSP</i> EN BANDES CRITIQUES .....	74
TABLEAU 5-3 : RÉSULTATS DU SYSTÈME DE RECONNAISSANCE DES OBJETS SONORES À L'AIDE DU CENTRE DE MASSE DE LA <i>DSP</i> .....	74
TABLEAU 5-4 : RÉSULTATS DU SYSTÈME DE RECONNAISSANCE DES OBJETS SONORES À L'AIDE DE LA LARGEUR DE BANDE DE LA <i>DSP</i> .....	75
TABLEAU 5-5 : RÉSULTATS DU SYSTÈME DE RECONNAISSANCE DES OBJETS SONORES À L'AIDE DE LA FRÉQUENCE DE COUPURE DE LA <i>DSP</i> .....	75



# CHAPITRE 1. INTRODUCTION

## 1.1 Contexte

Ce projet de recherche s'intéresse à l'assistance à domicile des personnes souffrant de troubles cognitifs. Le but de cette étude est de retarder le placement en institutions spécialisées des personnes atteintes. On définit alors la cognition comme la faculté du cerveau de penser, de traiter et d'emmagasiner de l'information afin de résoudre certains problèmes. En effet, les personnes concernées par ces systèmes de vigilance sont principalement les patients atteints de troubles cognitifs liés à l'âge, mais aussi à la démence de type Alzheimer, à la schizophrénie, à un traumatisme crânien ou à des déficiences intellectuelles. Ces patients éprouvent en général des problèmes de planification et de décision dans la réalisation des Activités de la Vie Quotidienne que nous désignerons sous l'acronyme *AVQ*. On comprend dans les activités de la vie quotidienne la mobilité au lit, le déplacement dans les pièces de l'habitat, l'habillement, l'alimentation, l'utilisation des toilettes, l'hygiène personnelle, le bain et la douche [18]. On désire alors compenser par l'environnement les troubles cognitifs impliquant une incompétence de la personne dans la réalisation de ces *AVQ*. Pour arriver au but décrit précédemment, on dispose d'un habitat intelligent, aussi appelé habitat communicant que l'on définit comme un habitat classique qui intègre les Nouvelles Techniques de l'Informatique et des Communications (*NTIC*). Ainsi, il est possible de modéliser l'environnement de l'appartement à l'aide de différents types de capteurs tels que des capteurs d'identification radiofréquences, des débits-mètres, des capteurs infrarouges, des capteurs de pression pour le sol ou encore des microphones [4, 10, 19, 27]. La Figure 1-1 présente un plan de l'habitat intelligent pilote, ainsi que la configuration spatiale des différents capteurs dont nous disposons au sein du laboratoire de DOMotique de l'Université de Sherbrooke (*DOMUS*) pour réaliser nos expérimentations.

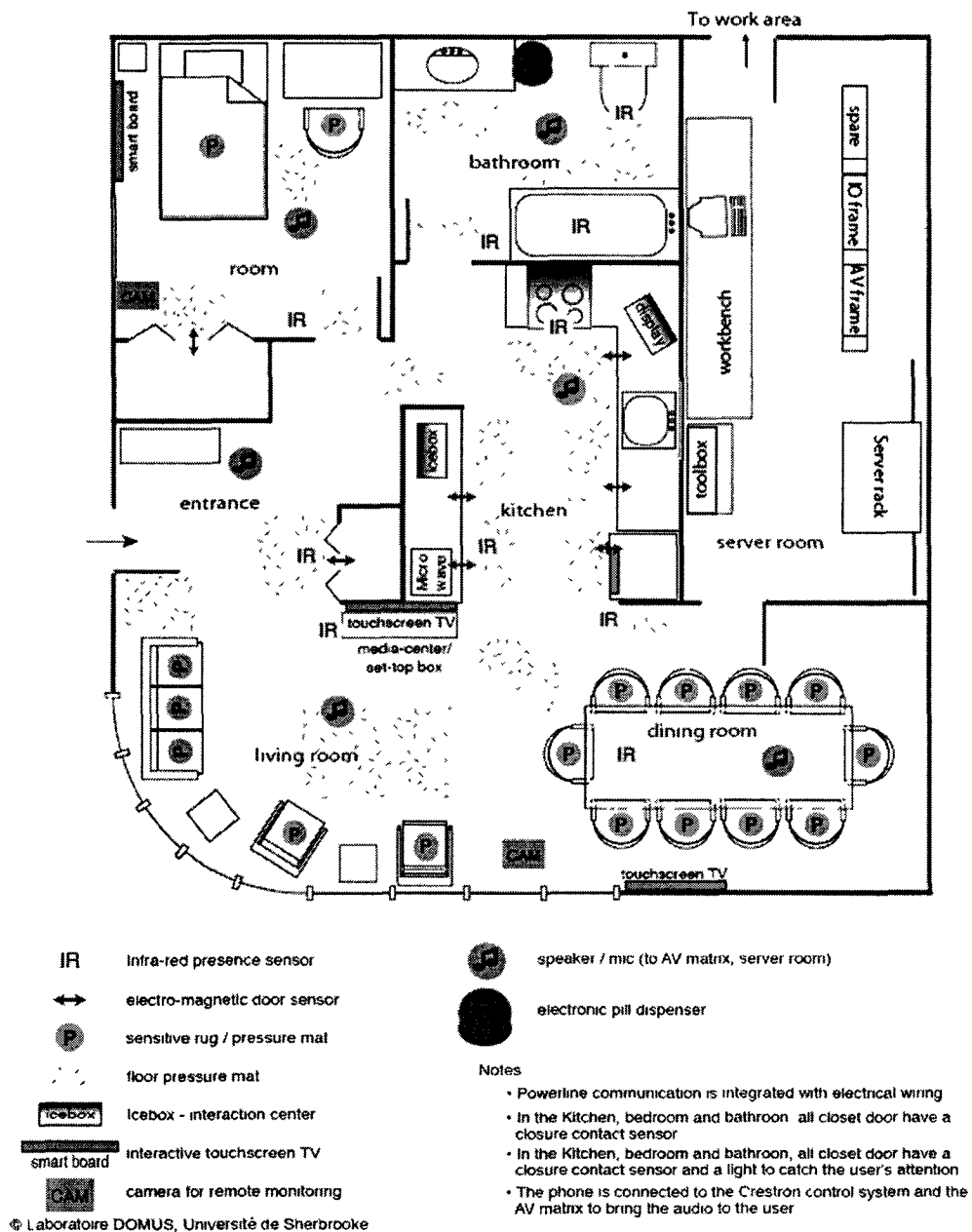


Figure 1-1 : Plan de l'appartement intelligent DOMUS avec la nature et la localisation des différents capteurs

L'infrastructure de l'appartement présenté à la Figure 1-1 permet de détecter et d'identifier les AVQ effectuées par une personne au sein de l'habitat. En jugeant de l'environnement de l'habitant, on peut détecter les situations à risque. Par exemple, le système devra être en mesure de prévenir la personne si elle oublie une casserole sur la gazinière ou bien de faire sa toilette. Il devra alors prévenir le cas échéant un intervenant extérieur si la personne est en état de crise. En identifiant ces situations-ci, il sera alors possible d'apporter une assistance

adaptée au contexte tout en prenant en compte le profil du patient. Ainsi, on pourra compenser les troubles cognitifs des personnes ciblées afin de leur permettre de conserver un certain niveau d'autonomie à domicile tout en étant le moins intrusif possible.

## 1.2 Problématique

Il existe déjà plusieurs systèmes de vigilance à domicile. Néanmoins, ces systèmes-là n'investiguent pas, ou de façon très succincte l'analyse continue des données sonores pour l'assistance des personnes à domicile [4, 10, 19, 27]. Dans le contexte de la surveillance à domicile, il paraît alors important de préciser que l'utilisation de capteurs sonores pourra s'avérer beaucoup moins dispendieuse et intrusive que de capteurs comme des accéléromètres ou encore des assistants personnels numériques qui doivent être portés en permanence par le patient. C'est pourquoi l'étude proposée se concentre exclusivement sur le traitement des données sonores afin de répondre à la problématique de la reconnaissance des *AVQ* au sein d'un habitat intelligent. Ainsi, les informations récupérées par les microphones viendront compléter celles des autres capteurs de l'appartement pour rendre plus robustes la reconnaissance d'*AVQ* et la localisation de la personne. La Figure 1-2 présente le schéma de principe de la reconnaissance des *AVQ* dans l'habitat intelligent du laboratoire *DOMUS*.

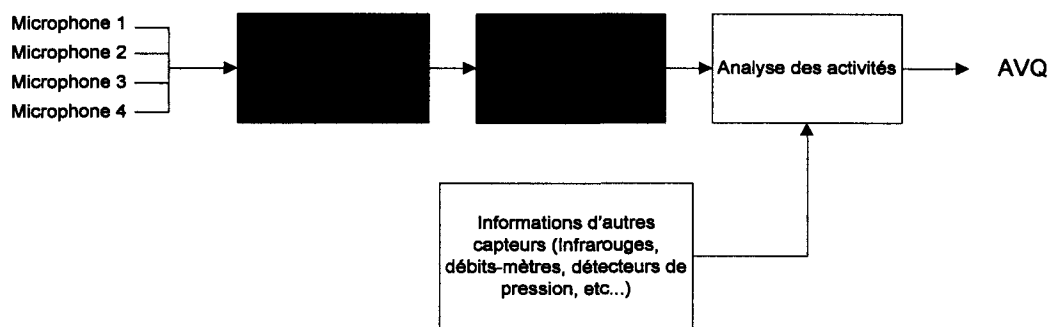


Figure 1-2 : Schéma de principe de la reconnaissance des activités de la vie quotidienne

### 1.3 Définition du projet de recherche

La partie du système à développer s'intéresse donc à la reconnaissance d'*AVQ* et se base sur l'analyse seule des signaux sonores. À l'aide des capteurs sonores placés à l'intérieur du laboratoire *DOMUS* à la Figure 1-1, on pourra identifier de quelle pièce provient tel ou tel événement sonore. Ceci permet de savoir dans quelle pièce se situe l'activité réalisée par le patient. Ceci apporte une première information qui peut s'avérer utile pour l'identification de l'action réalisée. À partir de là, est-il possible d'identifier l'action réalisée correspondant aux événements sonores détectés? Par exemple, asseyez-vous dans une cuisine et fermez les yeux. Seriez-vous capable en vous servant uniquement de votre ouïe, de reconnaître les actions réalisées par une personne dans cette pièce? Ainsi, le système à concevoir devra être capable de répondre à la question suivante : « Comment est-il possible de détecter et par la suite de reconnaître les activités de la vie quotidienne réalisées par une personne au sein d'un habitat intelligent à l'aide d'un simple microphone par pièce? »

On en profite pour introduire les notions d'action simple et d'*AVQ*. Ainsi, on sait qu'une activité de la vie quotidienne telle que faire du café se compose d'un enchaînement temporel d'actions simples. Pour la réalisation de l'*AVQ* de faire du café, on sait qu'il est nécessaire de faire couler de l'eau dans une casserole, de la faire bouillir, puis de sortir le café, une cuillère et une tasse des rangements. L'enchaînement de ces actions simples peut varier selon les scénarios possibles. Néanmoins, chacune de ces actions simples peut être identifiée à l'aide des événements sonores qu'elles génèrent. Puisqu'il paraît assez complexe d'identifier une activité de la vie quotidienne d'un seul coup ce qui impliquerait de créer des modèles pour chaque activité de la vie quotidienne, on décompose une activité de la vie quotidienne en plusieurs actions simples sur lesquelles on va se concentrer au cours de l'étude.

À une échelle macroscopique, on fait correspondre l'identification d'activité de la vie quotidienne à la réalisation de faire du café. Tandis qu'à une échelle microscopique, on définit la reconnaissance d'un événement sonore comme la détection et la reconnaissance d'une action simple réalisée comme faire couler de l'eau ou claquer une porte. Il sera alors nécessaire d'associer chacun de ces événements sonores à une action simple, afin de pouvoir par la suite faire correspondre un enchaînement d'actions simples avec une activité de la vie

quotidienne réalisée par l'habitant. La Figure 1-3 illustre ce principe en reprenant l'exemple de la réalisation d'une activité de la vie quotidienne comme faire du café.

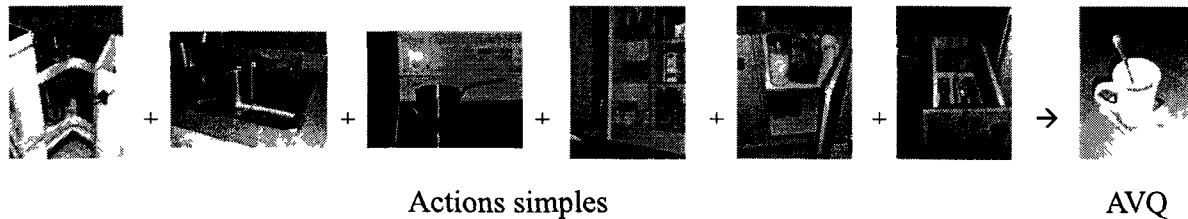


Figure 1-3 : Schéma de principe de la reconnaissance d'une AVQ à l'aide du son

## 1.4 Objectif du projet de recherche

Le système à réaliser devra permettre de détecter et capter les événements sonores, puis d'analyser ces signaux sonores en fonction de différents paramètres, et enfin de les classifier pour les faire correspondre à une action simple. La Figure 1-4 présente la méthodologie employée pour la reconnaissance des AVQ et plus particulièrement l'analyse des données sonores utilisée à cet effet.

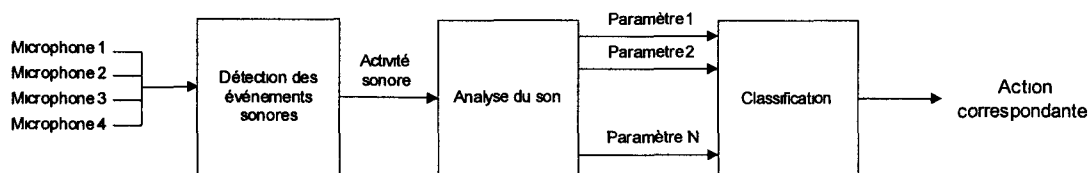


Figure 1-4 : Schéma de principe de la reconnaissance des activités de la vie quotidienne basée sur l'analyse sonore

On observe alors que le système de reconnaissance des AVQ se décompose en 3 étapes :

- 1- D'une part, il devra détecter et capturer les différents événements sonores qui ont lieu dans chaque pièce de l'habitat, en adaptant sa détection à l'environnement sonore de l'habitat.
- 2- D'autre part, il permettra de discriminer les actions simples réalisées à l'aide de l'analyse des événements sonores, aussi appelés objets sonores. C'est-à-dire que le système sera capable d'associer un objet sonore à une action simple faisant partie de sa base de données. C'est la première étape de classification.
- 3- Enfin, à partir de la reconnaissance et de l'enchaînement des différentes actions simples, ainsi que des informations provenant d'autres capteurs, le système conçu

pourra reconnaître les *AVQ* réalisées par l'habitant. C'est la seconde étape de la classification.

L'objectif principal du système à concevoir est de faire la preuve de concept qu'à l'aide d'un unique microphone par pièce, il est possible d'identifier les *AVQ* réalisées dans un habitat. Pour cela, une validation expérimentale devra valider la faisabilité d'une telle infrastructure, et ce, pour quelques événements sonores, principalement ciblés dans la cuisine. En plus de cet objectif, la détection des événements sonores engendrés par une activité de la personne permettra de déterminer dans quelle pièce la personne se trouve au sein de l'habitat.

## 1.5 Configuration de l'environnement de recherche

On se focalise dans notre étude sur les données sonores générées par des activités effectuées dans la cuisine. On retrouve dans la cuisine différents rangements : placards et tiroirs, ainsi qu'un four, un lave-vaisselle, un robinet, des ustensiles, des casseroles et des verres. Chacun de ces objets génère du bruit lorsque l'habitant interagit avec eux, et c'est sur ces informations sonores que l'on va travailler. On place un microphone « USB Dekstop Microphone<sup>1</sup> » au plafond de la cuisine à l'emplacement décrit à la Figure 1-5. Son utilisation vise à capter tous les bruits provenant de cette pièce. Il permettra l'enregistrement continu de l'environnement sonore de la cuisine. Cet enregistrement continu des données sonores est effectué à une fréquence d'échantillonnage  $F_s$  de 16 kHz. La fréquence de réponse du microphone est comprise entre 100 Hz et 16 kHz et sa sensibilité est de -47 dBV/Pa à  $\pm 4$  dB. Il est aussi possible d'appliquer un gain ou une atténuation du signal sonore reçu de façon logiciel. Ceci permettra d'utiliser au mieux la plage dynamique du microphone en fonction de l'éloignement de la source sonore. Par la suite, on étudie ces signaux à l'aide d'une analyse acoustique hors ligne à l'aide du logiciel *Matlab* qui simule le système à concevoir. La Figure 1-5 présente l'architecture de la cuisine.

---

<sup>1</sup> Spécifications du microphone sur cette page Web : <http://www.logitech.com/fr-fr/webcam-communications/microphones/devices/221>

microphone



Figure 1-5 : Photo de l'architecture de la cuisine

Il est à noter que toutes les études menées au sein du laboratoire *DOMUS* considèrent qu'une seule personne est présente dans l'appartement. De plus, les expérimentations menées pour ce projet considèrent un environnement peu bruyé et ne s'intéressent pas à la séparation de sources sonores. C'est-à-dire que les activités sonores sont considérées comme ayant lieu les unes après l'autre et non simultanément. Le système ne considèrera donc pas le cas d'une activité qui aurait lieu en même temps qu'une autre. Par exemple, un écoulement d'eau qui se produirait simultanément à un claquement de porte.

## 1.6 Plan du mémoire

Le présent document se décompose en 4 parties. Dans un premier temps, nous avons présenté le contexte et la problématique du projet de recherche proposé. Dans un second temps, nous étudierons la revue de la littérature des systèmes de détection et de reconnaissance des événements sonores existants. Puis nous présenterons les méthodes d'analyse de nos signaux sonores afin d'identifier leurs utilités dans le contexte recherche. Enfin, nous exposerons et discuterons des résultats fournis par notre système de détection et de reconnaissance des *AVQ*.



## CHAPITRE 2. ÉTAT DE L'ART

Plusieurs travaux ont déjà été effectués sur le traitement et l'analyse des données sonores pour des applications de surveillance. Ces travaux portent aussi bien sur la détection que sur la reconnaissance des événements sonores. La revue de littérature effectuée présente les outils d'analyse des données sonores et quelques méthodes de détection des événements sonores, ainsi que les différentes techniques de classification de ces mêmes événements.

### 2.1 Détection des événements sonores

On présente ici des méthodes qui permettent d'identifier la présence ou non d'un événement sonore. Celles-ci se basent sur une analyse du signal effectuée soit dans le domaine temporel, soit fréquentiel, soit dans les deux.

#### 2.1.1 Méthodes d'analyse temporelle

Certaines méthodes de détection d'activité sonore se basent sur le calcul de l'énergie temporelle à court terme [7]. C'est une mesure du niveau sonore en temps réel. Cependant, ce paramètre est très variable. Ceci est dû en partie à des conditions d'enregistrement différentes. En effet, une simple variation de la distance entre la source et le microphone suffit pour être un élément de perturbation du calcul de l'énergie. Pour s'affranchir de ce problème, une méthode d'adaptation de l'énergie à l'environnement sonore a été proposée. Des études proposent de calculer la moyenne de l'énergie sur plusieurs fenêtres d'étude de durée fixe, avec un recouvrement temporel. À partir du calcul de ces valeurs, un seuil a été fixé de façon empirique afin d'être comparé aux valeurs d'énergie calculées pour chaque trame courante. Si celles-ci sont supérieures au seuil, on considère alors qu'il y a eu une activité sonore, dans le cas contraire, on considère qu'il n'y en a pas eu [14, 33]. On peut remarquer que cette méthode s'adapte aux changements de l'environnement sonore puisque que la valeur seuil est mise à jour en temps réel. Néanmoins, ce principe de détection des événements sonores est très sensible aux nombres et à la durée des fenêtres d'analyse que l'on choisit pour le calcul du seuil. En effet, suivant ces paramètres, le seuil calculé, et par conséquent le nombre d'événements sonores détectés, seront très différents. Il est aussi possible de se baser sur le

taux de passage par zéro à court terme du signal pour la détection des événements sonores. Ce paramètre mesure le taux de changement de signe du signal sonore dans une fenêtre d'étude. Cela correspond au passage de l'amplitude du signal par zéro, d'une valeur positive à négative et vice-versa. L'avantage de cette méthode est qu'elle ne sera pas perturbée par une variation de distance entre la source sonore et le microphone. Il est aussi possible de moyennner la valeur de ce paramètre sur plusieurs fenêtres d'étude puis de calculer un seuil variable de la même façon que pour l'énergie temporelle [33].

Le tracé d'un signal sonore dans le domaine temporel présente les variations d'une forme d'onde au cours du temps. L'analyse fréquentielle permet de dire quelle proportion du signal appartient à telle ou telle bande de fréquence. On s'intéresse dès lors à l'étude de nos signaux, dans le domaine fréquentiel, dans le cadre de la détection d'événements sonores.

### 2.1.2 Décomposition de Fourier

L'idée de base de la transformée de Fourier est que toute forme d'onde peut être décomposée en une somme de sinusoides, éventuellement une infinité. Le résultat de cette transformée est appelé spectre en fréquences. On utilise une version échantillonnée de celle-ci pour l'analyse des signaux sonores numériques non périodiques, c'est-à-dire la Transformée de Fourier des Signaux Discrets, la *TFSD* [13, 24]. Cette analyse est effectuée sur une partie du signal segmenté à l'aide d'une fenêtre glissante avec un recouvrement temporel. Cette méthode est connue sous le nom de Transformée de Fourier discrète à fenêtre glissante ou à court terme. Elle permet d'obtenir un renseignement sur l'évolution fréquentielle du signal au cours du temps.

Dans la littérature, on rencontre souvent la Densité Spectrale de Puissance, la *DSP* qui est une représentation de l'énergie du spectre en fonction de la fréquence. Cette densité spectrale de puissance peut-être sous-échantillonnée en fonction d'un nombre prédéfini de bandes fréquentielles [9, 11, 34]. On assigne alors à chaque bande fréquentielle une valeur d'énergie. On peut alors définir l'échelle des fréquences de différentes façons, soit en *Hertz*, soit en *Mel*. En l'occurrence, les bandes critiques de *Mel* ont prouvé leur utilité pour l'extraction des paramètres sonores puisqu'elles permettent de se rapprocher au plus près de la perception humaine des sons [32, 36].

### 2.1.3 Décomposition en ondelettes

La décomposition en paquets d'ondelettes est aussi utilisée pour la détection des événements sonores [5, 17, 23, 24]. C'est une forme étendue de la transformée en ondelettes discrètes qui est elle-même une forme numérisée de la transformée en ondelettes. L'idée de base de la décomposition en ondelettes est que toute forme d'onde peut être décomposée suivant une famille d'ondelettes dilatées et translatées. Le système auditif ne réalise pas une analyse de Fourier, mais une analyse de type ondelette. En effet, la transformée de Fourier fait ressortir certaines caractéristiques spectrales qui ne sont pas importantes pour le système auditif. Le système auditif n'est pas un excellent estimateur spectral, il est un excellent estimateur des changements qui apparaissent dans l'estimation du spectre. Tout comme l'analyse spectrale, la transformée en ondelettes peut aussi extraire des caractéristiques non intéressantes. Cependant, contrairement à l'analyse de Fourier, on peut définir les bases d'ondelettes et les choisir de telles sortes que l'analyse soit plus proche de ce qui se passe dans l'audition. La décomposition en ondelettes pourra permettre une meilleure analyse des signaux présentant des discontinuités ou des phénomènes locaux. Ceci est un avantage pour notre système de détection des événements sonores qui s'intéresse à la mise en évidence des transitoires.

### 2.1.4 Méthodes de détection des événements sonores

À partir des informations apportées par l'analyse fréquentielle, des techniques d'analyse permettent de détecter les événements sonores qui se basent sur l'étude du spectre à court terme. Par exemple, l'indice de stationnarité mesure les variations d'énergie sur l'ensemble des composantes spectrales normalisées entre 2 fenêtres d'étude successives [24]. On remarque que cette méthode ne prend pas en compte l'environnement sonore sur une période plus longue que 2 trames sonores.

Pour s'adapter à l'environnement sonore de l'habitat, une solution possible est d'intégrer le spectre au cours du temps. On applique alors notre spectre en entrée d'un circuit intégrateur du premier ordre, et ce, de façon récursive pour chacune de nos fenêtres d'étude. Le coefficient de lissage est un paramètre qui permet de pondérer le spectre de la fenêtre d'étude courante par rapport au spectre intégré [13]. L'utilité d'une telle technique est d'adapter les variations de nos paramètres en fonction de l'environnement sonore en temps réel. Il est à noter qu'une

méthode semblable prenant d'autres types de paramètres en entrée du circuit intégrateur pourrait présenter des résultats intéressants pour la détection des événements sonores. À partir du spectre intégré et du spectre courant, la détection d'activité sonore se base sur l'estimation de différents critères. D'une part, on peut mesurer si l'environnement sonore considéré possède de fortes accroches ou coupures. Ce critère est obtenu grâce à un simple calcul large bande de la différence du spectre intégré pour 2 fenêtres d'analyse successives [13]. D'autre part, on peut s'intéresser à la détection d'événements sonores contenus dans de faibles bandes de fréquences [13]. Cette mesure compare simplement le pic maximum de la différence du spectre intégré à sa variance pour 2 fenêtres d'analyse successives. La détection des événements sonores peut aussi se baser sur l'énergie des coefficients du spectre regroupés en bandes fréquentielles. Un seuil est alors fixé, au-dessous duquel aucun événement sonore n'est détecté [12, 34].

## 2.2 Reconnaissance des activités sonores

On applique alors les événements sonores détectés en entrée de notre système de reconnaissance sonore. Les événements sonores détectés sont représentatifs des activités de la vie quotidienne réalisée par la personne au sein de l'habitat. On applique par la suite une analyse sur les événements sonores afin d'identifier l'action correspondante effectuée. Ceci permettra donc un gain de mémoire et de temps de calcul.

### 2.2.1 Méthodes de paramétrage temporel

Les études effectuées dans le domaine de la télésurveillance utilisent assez peu les paramètres temporels pour ce qui est de l'identification des activités sonores. Cependant, des descripteurs tels que le taux de voisement, le taux de passage par zéro et l'énergie temporelle à court terme sont parfois utilisés pour une classification grossière des événements sonores.

Par exemple, à partir de l'énergie temporelle à court terme, il est possible de définir deux seuils, l'un haut et l'autre bas, respectivement  $T_1$  et  $T_2$ . Si l'énergie calculée sur une durée prédéterminée est inférieure à  $T_1$ , on considère qu'il n'y a pas d'activité sonore, si l'énergie se situe entre les 2 seuils, on considère qu'une activité normale s'est produite. Enfin, si l'énergie est supérieure à  $T_2$ , on dit qu'un événement anormal s'est produit, par exemple, faire tomber

une chaise, un pot métallique ou encore pousser un cri. De la même manière, une absence d'activités sonores durant une longue période peut être révélatrice d'une situation problématique [4]. Une autre façon d'utiliser l'énergie temporelle à court terme est d'établir des modèles énergétiques à long terme pour chacune des activités considérées [13]. En fonction de ces modèles représentatifs de l'enveloppe temporelle de nos objets sonores, il sera alors possible de classer les sons environnementaux selon qu'ils soient simples, répétés, continus ou du bruit. Le taux de passage par zéro aide à la distinction d'événements répétitifs par rapport aux événements simples [3]. Par exemple, à l'aide de ce paramètre, il serait possible de discriminer des bruits de pas de marche et de course. De plus, un paramètre tel que le taux de voisement permet de séparer une activité sonore correspondante à de la parole, des autres activités. Le taux de voisement qui correspond au calcul de la corrélation croisée normalisée permet de déterminer la fréquence fondamentale, aussi appelé « pitch », pour un signal de parole [13, 34, 36]. On définit alors un seuil de voisement grâce à une valeur de corrélation normalisée. Au-dessous de ce seuil, les trames sont considérées non voisées, c'est-à-dire qu'elles ne contiennent pas d'information vis-à-vis de la fréquence glottale. Au-dessus de ce seuil, on considérera que le signal correspond à de la parole.

### 2.2.2 Méthodes de paramétrage fréquentiel

Dans le domaine spectral, on peut caractériser les événements sonores de différentes façons. Tout d'abord, il est possible de caractériser les événements sonores grâce à leur spectre ou leur densité spectrale de puissance en bandes critiques en fonction du temps [8, 9, 20, 21, 22, 25, 26, 32]. On s'intéresse aussi à paramétrer le spectre de chaque fenêtre d'étude à l'aide de différents descripteurs. On se réfère ici au centre de masse spectral, la largeur de bande spectrale, la fréquence de coupure spectrale. On rappelle que le centre de masse et la largeur de bande spectrale peuvent être vus respectivement comme étant les moments d'ordre 1 et 2 du spectre. Le centre de masse spectral permet de dire si le spectre est plutôt basse ou haute fréquence, tandis que la largeur de bande spectrale nous renseigne si le graphe en fréquence est étalé ou concentré autour du centre de masse [7, 13, 21, 22, 25]. La fréquence de coupure spectrale caractérise la décroissance de l'énergie spectrale dans les hautes fréquences. Ces paramètres permettent de caractériser le spectre en fonction de son contenu fréquentiel, basse ou haute fréquence, ainsi que la distribution du spectre. La transformée en ondelettes propose

une autre décomposition du signal sonore en fréquence qui permet de discriminer les objets sonores [8, 32]. Le domaine cepstral est souvent utilisé pour la classification des événements sonores. Il permet de séparer les composantes basses et hautes fréquences en appliquant une transformée de Fourier au logarithme du spectre [1, 3, 8, 22, 25, 26, 28, 32, 36].

### 2.2.3 Création des dictionnaires : l'algorithme des k-moyennes

La classification de nos événements sonore s'effectue à partir d'un dictionnaire de références que l'on va créer. On présente ici l'algorithme des *k-moyennes* qui est un algorithme de quantification vectorielle [12, 13, 26, 35]. Son principe est le suivant.

On dispose de vecteurs de dimension connue dans l'espace des observations que l'on souhaite rassembler en classes. À priori, on n'a pas de connaissance sur les propriétés de ces classes, seul leur nombre est connu. Logiquement, ce nombre de classes correspondra aux nombres d'activités sonores que l'on désire identifier. Pour cette technique, on dit que l'apprentissage est non supervisé, c'est-à-dire que cette méthode ne nécessite aucune information sur les données. À noter que plus le nombre d'enregistrements choisi pour l'élaboration du dictionnaire sera grand, et plus la classification des données d'observation sera robuste. Il sera aussi intéressant d'établir plusieurs modèles pour un même type d'événement sonore.

### 2.2.4 Classification des événements sonores

Une fois le dictionnaire des références créé, la classification des événements sonores consiste en une simple minimisation d'une distance choisie entre les vecteurs de l'espace d'observation et les vecteurs de référence qui composent nos dictionnaires. L'une des méthodes les plus utilisées est l'algorithme de recherche des *plus proches voisins*, «*Nearest Neighbor*», *NN*, en anglais [1, 13, 26, 32]. Cette méthode se base le plus souvent sur une minimisation de la distance euclidienne entre les vecteurs de l'espace d'observation et ceux du dictionnaire.

Une autre technique connue est l'algorithme de programmation dynamique, «*Dynamic Time Warping*», *DTW*, en anglais [8, 13, 35, 36]. Elle mesure la ressemblance entre 2 séquences de durée et de vitesse différentes. L'un de ces domaines d'application les plus répandus est la reconnaissance de la parole afin d'éliminer le problème des différents débits de parole, mais il

est aussi utilisé dans le cadre de la reconnaissance d'activités. Le processus de reconnaissance consiste à évaluer la distance d'une observation à chacune des références du dictionnaire. Toute la difficulté du décodage réside dans cette mesure d'un degré de similarité entre des formes acoustiques variables à la fois au niveau spectral (vitesse) et temporel (durée). L'algorithme de programmation dynamique réalise cet alignement en recherchant, parmi tous les alignements possibles, celui qui minimise une fonction de coût intégrant l'écart spectral des données alignées et un coût de distorsion temporelle. La distance retenue est celle correspondant à l'alignement de coûts minimaux. Par conséquent, le mot reconnu sera celui dont la référence est la plus proche de l'observation, soit le principe du plus proche voisin. Cette technique possède d'autres inconvénients importants qui limitent son champ d'application. D'une part, cette méthode est très sensible à la modélisation des références, néanmoins, ce problème peut être contourné en créant plusieurs références pour la même activité dans notre dictionnaire.

Les travaux déjà effectués et présentés dans la revue de la littérature ne s'appliquent pas au domaine de l'assistance à domicile. En effet, les actions identifiées ne correspondent pas à des activités domestiques. Dans notre étude, on se concentrera sur la détection des événements sonores relatifs aux activités domestiques au sein de l'habitat. Puis, à partir de l'enregistrement sonore de ces activités, on élaborera un système de classification capable de discriminer les différentes activités domestiques réalisées par la personne.





# CHAPITRE 3. CONCEPTION

## 3.1 Présentation des signaux étudiés

On présente à la Figure 3-1 l'allure temporelle de différents signaux sonores enregistrés dans la cuisine. Ces formes d'onde correspondent à des actions simples telles que la manipulation d'ustensiles, de casseroles, un claquement de porte de placard et un écoulement d'eau du robinet. L'amplitude du signal est exprimée en *dBV*.

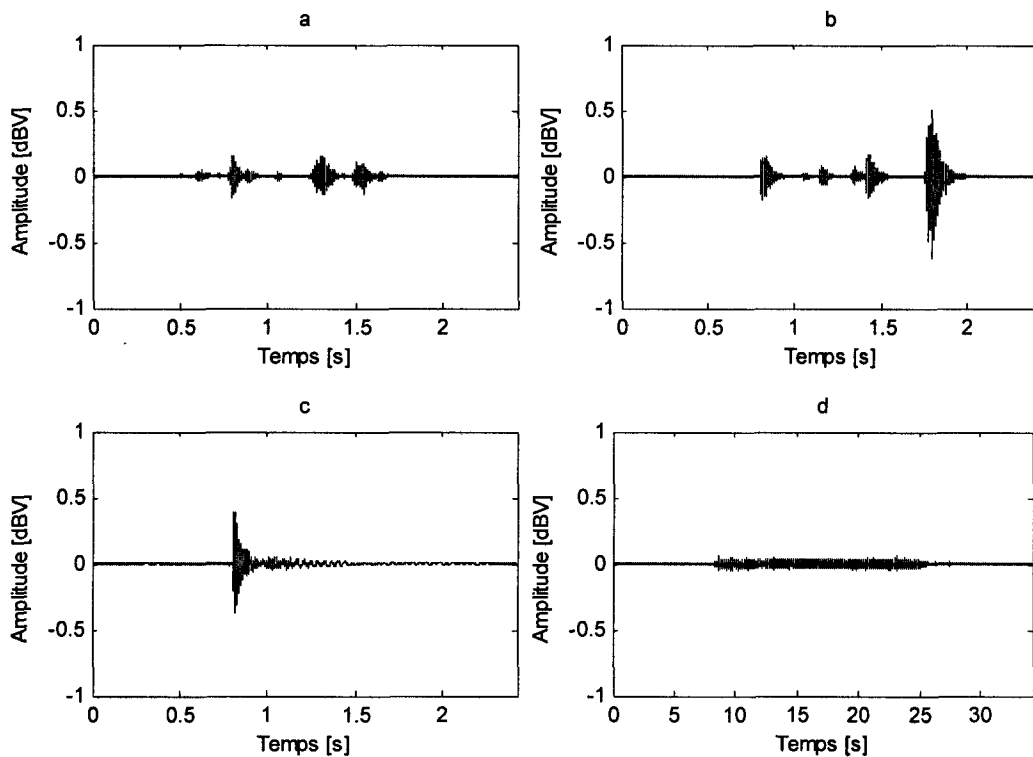


Figure 3-1 : Allure temporelle des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

On remarque que ces signaux présentent les formes d'onde différentes selon les actions effectuées. Pour ce qui est du son de l'écoulement d'eau du robinet, le signal est très stationnaire. Il présente une accroche et une chute sonore assez nettes. On observe aussi que l'amplitude de la forme d'onde est assez faible, entre  $\pm 0.06$  *dBV*. Le bruit de claquement de porte correspond à un signal impulsif de courte durée. L'accroche sonore est très marquée, elle possède une amplitude de  $\pm 0.4$  *dBV*. La chute sonore correspond à une décroissance

exponentielle. Les formes d'onde de manipulation d'ustensiles et de casserole possèdent de multiples accroches et chutes sonores. Elles sont constituées d'un enchaînement de bruits impulsifs. On définit ces signaux comme fluctuants. L'amplitude de ces signaux varie entre  $\pm 0.1$  et  $\pm 0.5$  *dBV*. Néanmoins, l'accroche et la chute sonore des événements sonores peuvent varier selon la physique même des objets avec lesquels la personne interagit. Par exemple, si la personne ouvre et ferme une porte de placard, le signal sonore résultant sera différent que si la personne ouvre et ferme une porte de chambre. Il en est de même pour les casseroles et les ustensiles. Enfin, l'amplitude de la forme d'onde de ces signaux dépend de l'éloignement du microphone vis-à-vis de l'action effectuée.

Ainsi, le système de détection et de reconnaissance des événements sonores à concevoir devra capturer avec précision ces différents types de signaux : stationnaires, impulsifs et fluctuants. De plus, il devra adapter sa détection aux signaux d'énergie différente et ainsi faire abstraction de l'éloignement de l'événement sonore vis-à-vis du microphone.

## 3.2 Segmentation du signal sonore

Dans notre étude, on utilise un fenêtrage de l'enregistrement sonore continu afin de sélectionner un signal de longueur finie. En effet, les calculs nécessaires à la détection des événements sonores de notre enregistrement sonore continu ne peuvent se faire que sur un nombre fini de points. Pour observer le signal sur une durée finie, on le multiplie alors par une fenêtre d'observation. La longueur de la fenêtre d'analyse  $N_t$  a été fixée égale à 512 points, ce qui correspond à une durée de 32 *ms* pour une fréquence d'échantillonnage du signal  $F_s$  de 16 *kHz*. Cette taille de fenêtre d'analyse de 32 *ms* permet d'assurer la quasi-stationnarité du signal sonore [7, 13]. Cette segmentation du signal est indispensable pour appliquer la transformée de Fourier. On définit alors 2 fenêtres d'analyse suivant le cas où l'on se place dans le domaine d'étude temporel, ou fréquentiel.

### 3.2.1 Fenêtrage du signal sonore

Dans le cas où l'on se place dans le domaine d'étude temporel des signaux, on choisit une fenêtre rectangulaire notée  $win_1$  :

$$win_1 = 1$$

Pour,  $1 \leq t \leq N_t$

Dans ce cas, la fenêtre d'analyse  $w_1$  permet uniquement de définir une partie du signal  $s$  de durée finie pour effectuer le calcul des paramètres temporels.

Dans le cas où l'on se place dans le domaine d'étude fréquentiel des signaux, on choisit une fenêtre de « *Hamming* » notée  $win_2$  [1, 8, 21, 36] :

$$win_2 = 0,54 * 0,46 * \cos(2\pi t / N_t)$$

Pour,  $1 \leq t \leq N_t$

La Figure 3-2 présente l'allure temporelle d'une fenêtre de « *Hamming* ».

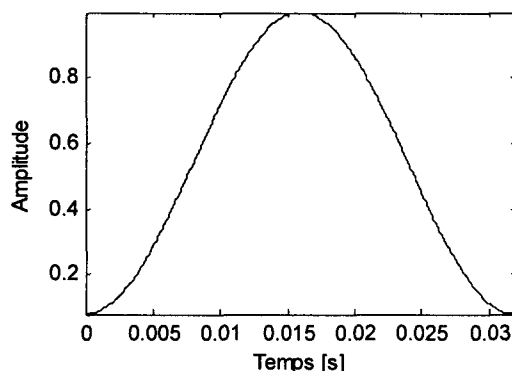


Figure 3-2 : Allure temporelle de la fenêtre de « *Hamming* »

On justifie alors l'utilisation d'une telle fenêtre dans le domaine d'étude fréquentiel par le fait qu'une multiplication dans le domaine temporel correspond à une convolution dans le domaine fréquentiel. Ainsi, pour ne pas biaiser le spectre du signal, il faudrait que l'allure de la fenêtre dans le domaine spectral corresponde à une fonction de Dirac, puisque son allure dans le domaine temporel est un signal constant infini. Malheureusement, ceci est impossible

en pratique. Pour s'en rapprocher le plus possible, on opte pour une fenêtre spectrale ayant un lobe principal qui soit le plus étroit possible, avec des lobes secondaires qui soient les plus atténués possible. Néanmoins, plus le lobe principal d'une fenêtre est étroit et plus les lobes secondaires seront importants. Il y a donc un compromis à faire entre ces 2 paramètres suivant les objectifs de l'analyse fréquentielle et c'est ce que fait en partie la fenêtre de « *Hamming* ».

### 3.2.2 Chevauchement temporel

On applique un chevauchement temporel de 50% au fenêtrage du signal sonore, « *temporal overlap* » en anglais [1, 7]. La valeur de chevauchement correspond à un nombre de points égal à 256 points, soit  $N/2=16\text{ ms}$ . Ainsi, au lieu de fenêtrer notre signal sonore tous les 512 points, celui-ci sera fenêtré de manière itérative tous les 256 points. La Figure 3-3 présente 3 fenêtres de « *Hamming* » auxquelles on a appliqué un chevauchement temporel de 50%.

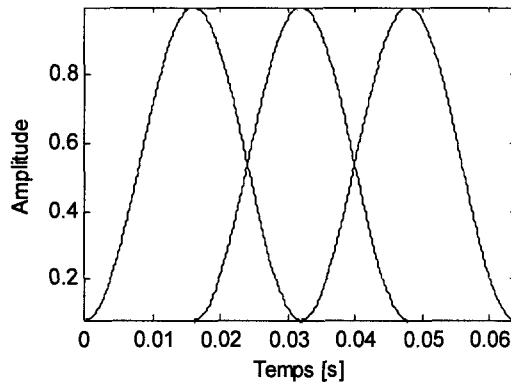


Figure 3-3 : Principe du chevauchement temporel

Cette durée de chevauchement de 16 ms correspond donc à la précision temporelle avec laquelle il nous sera possible de détecter l'accroche et la chute sonore d'un événement. Ainsi, un chevauchement temporel de 50% permet de diviser par 2 la résolution temporelle. Si l'on diminue la durée de chevauchement, alors on obtiendra une meilleure précision temporelle. Néanmoins, ceci augmenterait considérablement le nombre de fenêtres d'étude et par conséquent le temps de calcul des paramètres fréquentiels.

### 3.2.3 Définition d'une trame sonore

On définit deux trames sonores successives  $x_l$  et  $x_{l+1}$  avec  $n$  un instant donné et  $l$  un entier correspondant à l'indice du segment :

$$x_l[n] = s[ln:ln + N_t] * win_{1,2}[1:N_t]$$

$$x_{l+1}[n] = s[ln + N_t/2 : ln + N_t/2 + N_t] * win_{1,2}[1:N_t]$$

On utilisera l'appellation segment, trame ou fenêtre sonore tout au long du présent document pour désigner  $x_l$ .

## 3.3 Méthodes utiles à la caractérisation des signaux sonores

### 3.3.1 L'énergie temporelle

L'énergie temporelle est définie pour un segment  $l$  donné, comme étant la moyenne quadratique de l'amplitude des échantillons du segment sonore [7, 33]. C'est donc une mesure du niveau sonore. Cette énergie est alors normalisée par la longueur de la fenêtre d'étude. On applique un logarithme de base 10 au calcul de l'énergie afin de l'adapter à l'audition humaine puisque l'oreille est un récepteur logarithmique. L'énergie logarithmique est quantifiée en  $dB$ . La formule permettant le calcul de l'énergie temporelle logarithmique à court terme  $E_t(l)$  d'un segment sonore  $x_l$  est la suivante :

$$E_t(l) = 10 * \log_{10} \left( \frac{1}{N_t} \sum_{n=1}^{n=N_t} x_l[n]^2 \right)$$

La Figure 3-4 présente les variations de l'énergie temporelle des formes d'ondes des signaux présentés à la Figure 3-1.

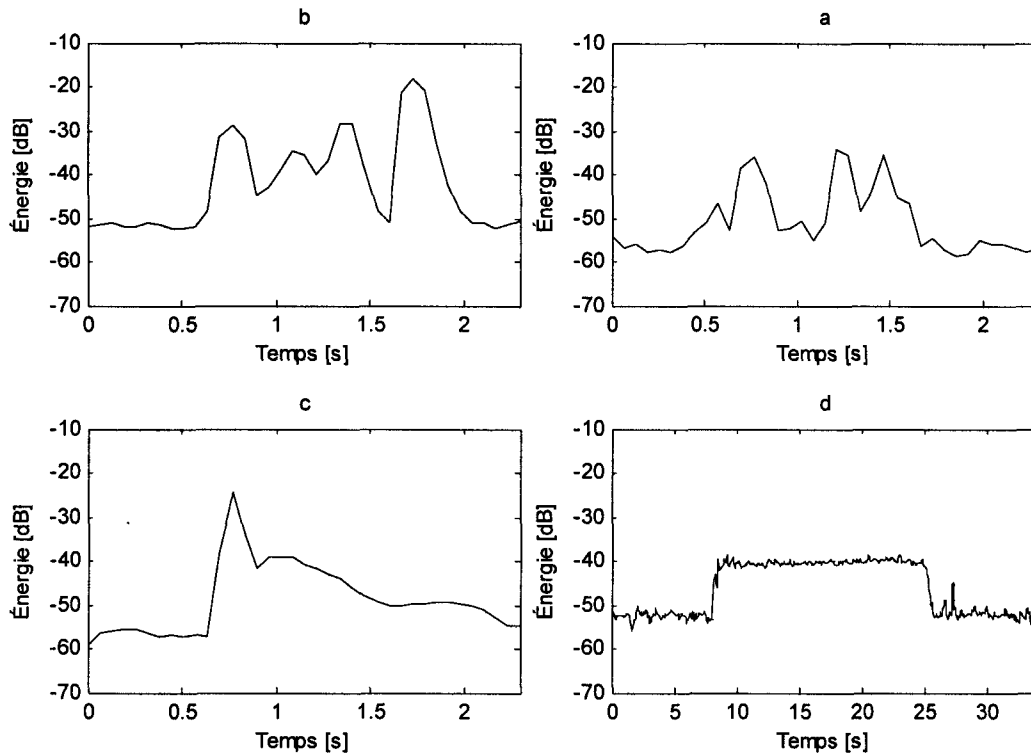


Figure 3-4 : Tracé de l'énergie temporelle des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

Les valeurs d'énergie calculées à la Figure 3-4 sont négatives puisque la valeur des échantillons de nos signaux est comprise entre  $\pm 1$  dBV, et que l'énergie est normalisée par rapport à la longueur de la fenêtre d'étude.

On observe que les variations d'énergie correspondent grossièrement aux changements d'enveloppe temporelle de nos signaux à la Figure 3-1. Pour un son correspondant à un écoulement d'eau, la valeur maximale se situe aux alentours de  $-40$  dB, alors qu'elle est de  $-20$  dB pour un claquement de porte. Cette valeur maximale oscille entre  $-30$  dB et  $-20$  dB pour des bruits de manipulation d'ustensile et de casserole. Ces valeurs d'énergie diffèrent selon les formes d'ondes. Ceci est dû aux contenus énergétiques différents selon les types d'événements, mais aussi à des conditions d'enregistrement différentes.

### 3.3.2 Le taux de passage par zéro

Le taux de passage par zéro, appelé en anglais « ZCR : Zero-crossing rate », mesure pour un segment sonore donné, le nombre de fois que l'amplitude du signal passe par zéro d'une valeur positive à négative et vice versa [3, 33]. Ce paramètre est aussi normalisé par rapport à

la longueur de la fenêtre d'analyse  $N_t$ . Le calcul du taux de passage par zéro  $ZCR(l)$  pour une trame sonore  $l$  est défini par la formule suivante :

$$ZCR(l) = \frac{1}{N_t} \sum_{n=1}^{N_t} |\text{sgn}\{x_l[n+1]\} - \text{sgn}\{x_l[n]\}|$$

Avec,  $\text{sgn}$  l'opérateur signe qui renvoie 1 si le scalaire est positif et 0 s'il est négatif. La Figure 3-5 présente les variations du taux de passage par zéro à court terme pour les sons de la Figure 3-1.

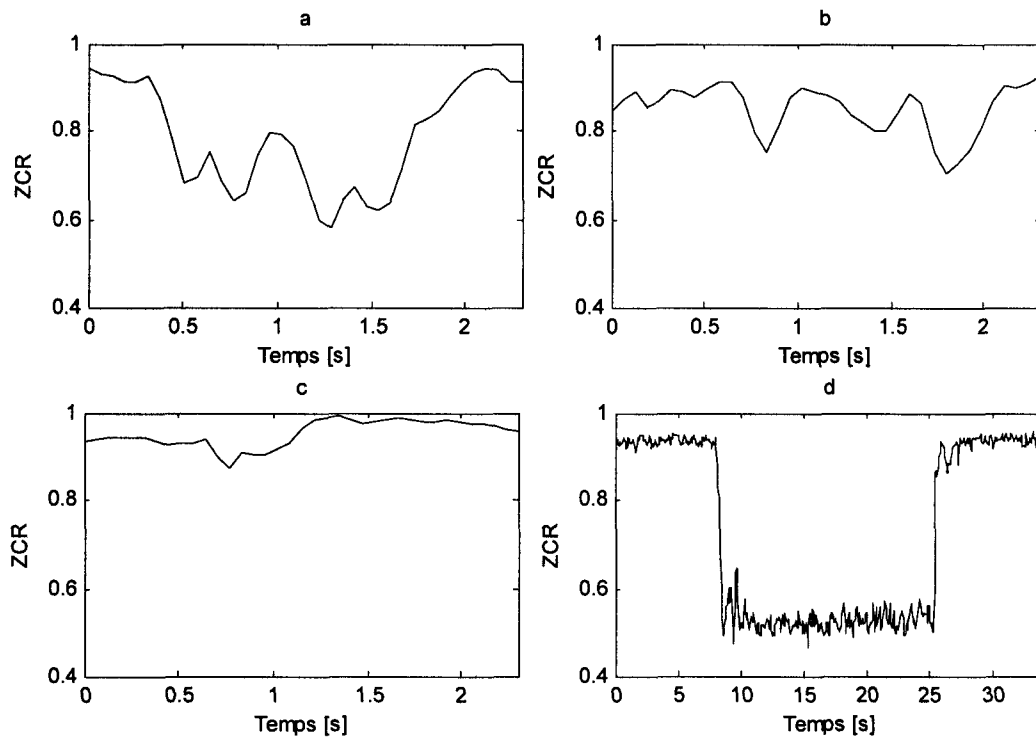


Figure 3-5 : Tracé du taux de passage par zéro des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

La Figure 3-5 montre que les variations du taux de passage par zéro correspondent à l'inverse des changements d'enveloppe temporelle de nos signaux. Lors d'une accroche sonore, sa valeur diminuera de façon significative et elle augmentera de nouveau lors d'une chute sonore. On voit aussi que les valeurs du taux de passage par zéro varient selon les événements sonores considérés. Ainsi, le taux de passage par zéro pour un claquement de porte atteindra une valeur minimale de 0.9, alors qu'il oscille autour de 0.5 pour un écoulement d'eau. Pour un

bruit de manipulation d'ustensile et de casserole, il atteint respectivement 0.6 et 0.7. On fait alors remarquer qu'un bruit d'écoulement d'eau est constitué d'une infinité de bruits de goutte d'eau, ce qui explique l'évolution de la valeur du taux de passage par zéro qui présente une valeur faible pour des bruits très répétitifs. De même, le bruit d'une manipulation d'ustensile est constitué d'un ensemble d'ustensiles qui s'entrechoquent, ainsi la valeur du taux de passage par zéro est plus faible que celle correspondant à une manipulation de casserole. Pour un bruit impulsif tel qu'un claquement de porte, la valeur du taux de passage par zéro reste très élevée. Ainsi, cet outil peut aider à la distinction d'événements excités par rapport aux événements normaux. On fait aussi remarquer que la valeur du taux de passage par zéro est indépendante des conditions d'enregistrement. Cependant, elle dépend de la nature du signal sonore, stationnaire, fluctuant ou impulsif.

Les paramètres temporels présentés ci-dessus ne permettent pas à eux seuls l'élaboration d'un système de détection d'événements sonores robuste et intégrant une large famille d'événements sonores en entrée. Néanmoins, l'énergie temporelle à court terme pourra être utilisée pour fixer un seuil absolu de détection en dessous duquel on considère qu'aucun événement sonore n'est révélateur d'une action de l'habitant. On s'intéresse alors aux méthodes de paramétrage fréquentiel de signaux.

### 3.3.3 La transformée de Fourier

La Transformée de Fourier noté  $TF$  est une application de la théorie des séries de Fourier à des fonctions non périodiques, ceci afin de leur associer un spectre. Le principe de cette transformée est d'obtenir l'expression d'une fonction  $f(t)$  comme une somme infinie, sous forme d'intégrale, de fonctions trigonométriques  $e^{-i2\pi ft}$  de toutes les fréquences  $f$  qui forment son spectre  $F(f)$ . On définit la transformée de Fourier continue  $F(f)$  du signal  $f(t)$  par :

$$F(f) = \int_{-\infty}^{+\infty} f(t) \cdot e^{-i2\pi ft} dt$$

Cependant, cette transformée s'applique aux signaux analogiques alors que dans notre étude nous travaillons exclusivement avec des signaux sonores numérisés. On va donc définir une version échantillonnée de cette transformée.



### 3.3.4 La transformée de Fourier discrète

On utilise alors une version échantillonnée de la *TF* appelée Transformée de Fourier Discrète noté *TFD*. On retrouve dans la littérature l'appellation *TFD* à court terme (en anglais « *Short Discret Fourier Transform* » connu sous l'acronyme *TFSD*) ou *TFD* à fenêtre glissante pour désigner les *TF* utilisées pour déterminer la fréquence sinusoïdale et la phase d'une section locale d'un signal numérisé, c'est-à-dire d'un segment sonore  $x$  [8, 12, 13, 20, 21, 25, 27]. On définit alors le spectre  $X_l[k]$  pour une valeur numérisée de fréquence  $k$ , d'une trame sonore  $x_l$  à l'aide de la formule suivante :

$$X_l[k] = \sum_{n=1}^{N_t} x_l[n]. e^{-2i\pi k \frac{n}{N_f}}$$

Pour :

$$0 \leq k < N_f$$

On en profite pour introduire le théorème de *Nyquist-Shannon*. Celui-ci énonce que la fréquence d'échantillonnage  $F_s$  du signal doit être égale ou supérieure au double de la fréquence maximale contenue dans le signal afin de le convertir d'une forme analogique à une forme numérique. Ainsi, le spectre  $X_l$  issu de la *TFSD* d'une trame  $x_l$  donnera un spectre dont les fréquences seront comprises entre  $k=-F_s/2$  et  $k=F_s/2$ . On introduit alors la fréquence de *Nyquist* notée  $F_n$  égale à  $F_s/2$  soit 8 kHz.

On remarque, d'après la formule de la *TFSD*, que le spectre  $X_l[k]$  sera de longueur  $N_f$ . Dès lors, on sait que la longueur de la *TFSD* a une influence directe sur la résolution fréquentielle du spectre. Ainsi, plus la longueur de la transformée de Fourier est grande et plus il y aura de points qui décriront son spectre. Dans notre cas, on a opté pour une taille de spectre  $N_f$  égale à 512 points, ce qui correspond donc à une résolution fréquentielle de  $F_s/512$  soit  $F_n/256$  égale à 31,25 Hz.

On introduit ici une propriété essentielle de la *TF* et de la *TFSD* par extension. Si  $x$  est une fonction réelle, alors :

$$X[-k] = X^*[k]$$

On en déduit que :

$$|X[-k]| = |X^*[k]| = |X[k]|$$

On peut alors déduire de ces 2 équations que la *TFSD* est symétrique par rapport à l'axe vertical de fréquence  $k=0$ . On peut donc conclure que seul le demi-spectre de  $X[k]$  compris entre  $k=0$  et  $k=N_f/2-1$  nous renseigne sur les fréquences sinusoïdales. Puisqu'on se désintéresse de l'information de phase, après troncature du demi-spectre des fréquences négatives, on se retrouve avec un spectre de taille  $N_f/2$  soit 256 points, qui contient l'information spectrale d'une trame sonore.

### 3.3.5 La densité spectrale de puissance

À partir de la *TFSD* énoncée précédemment, il est possible de calculer la Densité Spectrale de Puissance notée *DSP* [9, 22, 25, 26, 34]. Elle peut être calculée de différentes façons. Elle correspond soit au module au carré de la *TFSD*, soit au produit de celle-ci et de son conjugué. La *DSP* permet de représenter l'énergie des différentes composantes spectrales d'un segment sonore. On définit alors la  $DSP_l$  pour une trame  $x_l$  avec  $k$  les fréquences :

$$DSP_l(k) = \left| \sum_{n=1}^{N_f/2} x_l[n] \cdot e^{-2i\pi k \frac{n}{N_f}} \right|^2$$

$$DSP_l(k) = X_l[k] * X_l^*[k]$$

Pour :

$$0 \leq k < N_f/2$$

On expose à la Figure 3-6 les images des densités spectrales de puissance en fonction du temps pour les formes d'onde de la Figure 3-1. Le système de représentation graphique utilisé est un codage par niveaux de gris. Les composantes spectrales de plus forte énergie tendront vers le blanc, alors que les plus faibles tendront vers le noir. Les images spectrales présentées à la Figure 3-6 sont toutes normalisées par rapport aux pixels de plus haute et de plus faible valeur afin d'utiliser toute la plage dynamique du codage par niveaux de gris.

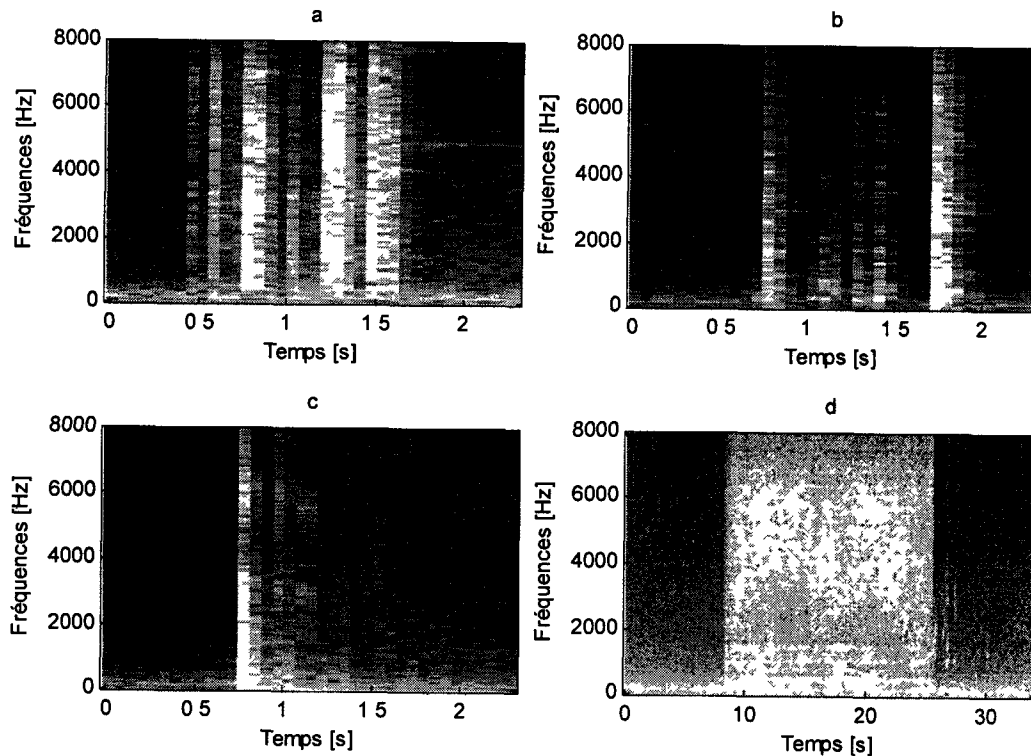


Figure 3-6 : Images de la densité spectrale de puissance des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

Les images spectrales mettent en évidence de façon significative les accroches et les chutes sonores des événements considérés. Pour les bruits de manipulation de casserole et d'ustensile, on observe alors des fluctuations de la *DSP*. L'image spectrale de l'écoulement d'eau du robinet montre une répartition de l'énergie de la *DSP* uniforme et constante au cours du temps sur toute la gamme des fréquences. Ceci est caractéristique d'un événement sonore stationnaire. De même pour le bruit de claquement de porte, on remarque des harmoniques fréquentielles qui font suite à l'accroche sonore, celles-ci sont représentatives de la résonance du placard. Ainsi, chacun des événements sonores considérés possède des accroches fréquentielles large bande, c'est-à-dire visibles sur toute la plage fréquentielle de la *DSP*. Par contre, la chute sonore n'est pas toujours représentée par une décroissance uniforme de l'énergie sur toutes les fréquences, elle peut être parfois en bande étroite.

Les images spectrales *DSP* prennent un espace mémoire important. Dans notre cas, cet espace de stockage doit contenir les 256 points de notre spectre multipliés par le nombre de trames sonore dont est constitué l'événement. Par conséquent, le temps de traitement de ces données

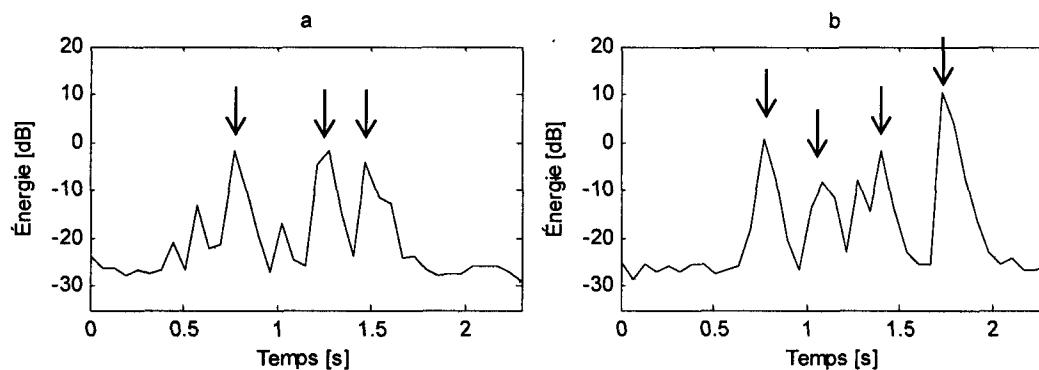
est important. Néanmoins, à la vue des images spectrales, une précision fréquentielle de 31,25 Hz n'est pas nécessaire à la détection des accroches et chutes sonores. C'est pour cela que l'on introduira au paragraphe 3.3.10 une version quantifiée de la DSP selon des bandes de fréquence. Pour l'instant, on se concentre sur les différentes méthodes de description de la DSP qui vont permettre de la caractériser.

### 3.3.6 L'énergie spectrale

L'énergie spectrale à court terme de la DSP correspond à la somme des échantillons spectraux de la DSP [22, 25]. De même que pour l'énergie temporelle, l'énergie spectrale est normalisée par la demi-longueur de la fenêtre d'étude, soit  $N_f/2$ , puisque l'on s'intéresse uniquement au demi-spectre. On lui applique un logarithme de base 10 et on l'exprime en dB. Son équation est la suivante :

$$E_{DSP} = 10 * \log_{10} \left( \frac{2}{N_f} \sum_{k=0}^{N_f/2-1} DSP(k) \right)$$

Ce paramètre rend compte de l'énergie fréquentielle contenue dans une trame sonore. Mais, il ne permet pas de dire si l'énergie se situe dans les basses ou hautes fréquences. On présente à la Figure 3-7, les variations d'énergie globale de la DSP des événements sonores considérés.



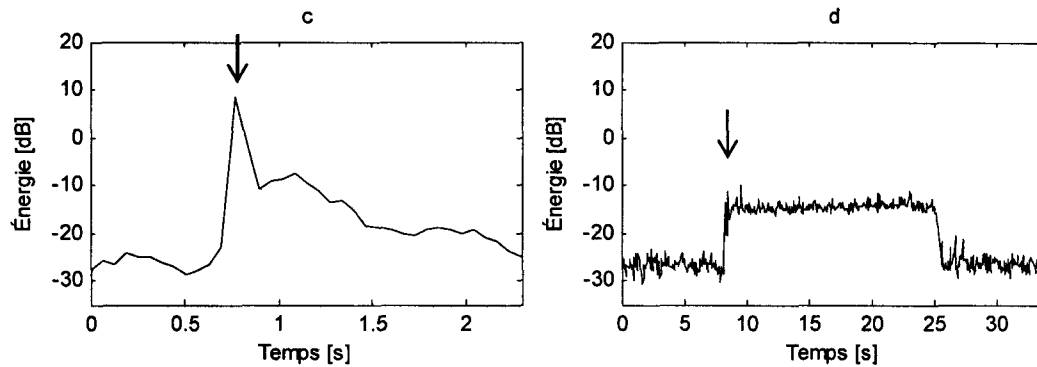


Figure 3-7 : Tracé de l'énergie spectrale de la *DSP* des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

Les changements d'énergie globale de la *DSP* sont en phase avec les variations temporelles des formes d'onde des événements sonores considérés. À la Figure 3-7, on identifie les principales accroches sonores des différents événements sonores à l'aide de flèches. Ces accroches sonores se traduisent par des changements de la valeur de l'énergie spectrale de la *DSP*. La valeur de l'énergie globale de la *DSP* atteint  $10\text{ dB}$  pour un claquement de porte et  $-15\text{ dB}$  pour un écoulement d'eau. Pour ce qui est de la manipulation d'ustensiles et de casseroles, les valeurs de l'énergie spectrale de la *DSP* des accroches sonores oscillent entre  $\pm 10\text{ dB}$ . L'énergie spectrale est donc dépendante des types d'événements considérés ainsi que des conditions d'enregistrement tout comme l'énergie temporelle. Ainsi, une variation de la distance entre la source sonore et le microphone perturbera le calcul de l'énergie spectrale. Cependant, les valeurs d'énergie de la Figure 3-4 diffèrent légèrement de celles de la Figure 3-7. En effet, les composantes spectrales de la *DSP* n'ont pas une amplitude strictement comprise entre  $\pm 1$  contrairement à l'amplitude des signaux sonores.

Il n'est donc pas envisageable de détecter les événements sonores à partir de ce seul paramètre. De la même façon que pour l'énergie temporelle, on pourra utiliser ce paramètre pour effectuer un seuillage absolu.

### 3.3.7 Le centre de masse spectrale

Le centre de masse spectral de la  $DSP$  noté  $C_{DSP}$  est aussi appelé centroïde spectral [7, 13, 21, 22, 25]. Il correspond au moment d'ordre 1 de la  $DSP$ . C'est la moyenne pondérée des amplitudes de la  $DSP$  normalisée par les fréquences. On obtient alors la formule suivante :

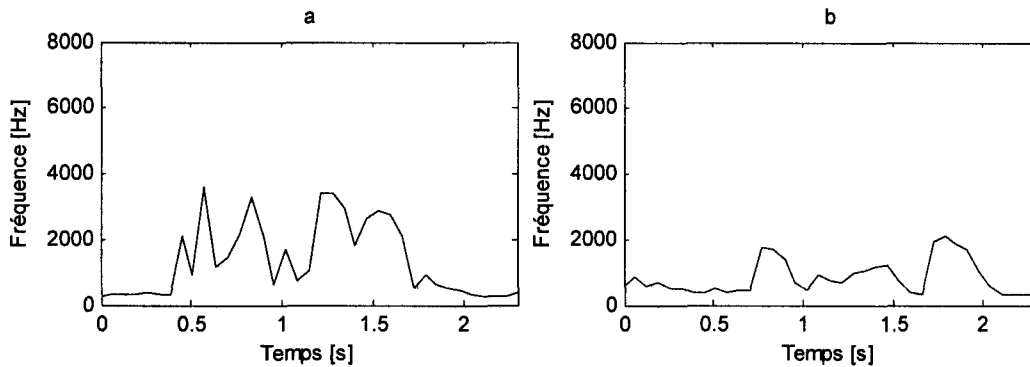
$$C_{DSP} = \frac{\sum_{k=0}^{N_f/2-1} DSP_{norm}(k) * k}{\sum_{k=0}^{N_f/2-1} DSP_{norm}(k)}$$

Avec,  $DSP_{norm} = DSP / \max_k(DSP)$

La valeur du centroïde spectral correspond à un indice. Afin d'obtenir des résultats exprimés en *Hertz*, on utilise la formule suivante :

$$C_{Hz} = C_{DSP} * (F_s/N_f)$$

Ce calcul sera appliqué pour d'autres paramètres spectraux tels que la largeur de bande et la fréquence de coupure spectrale présentés par la suite. Le centroïde spectral s'exprime donc en *Hertz* et est compris entre 0 et  $F_n=8$  kHz. On appelle parfois cet indicateur « *spectral brightness* » en anglais. Il permet de dire si la  $DSP$  présente des caractéristiques plutôt basses ou hautes fréquences. La Figure 3-8 présente l'évolution du centre de masse de la  $DSP$ .



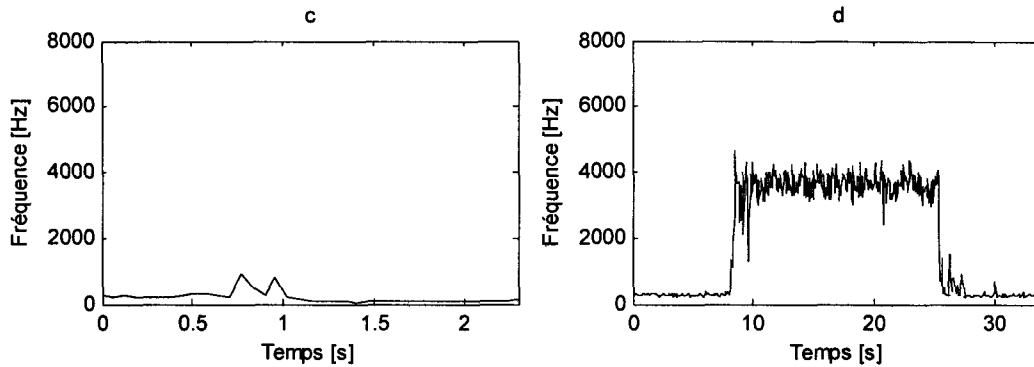


Figure 3-8 : Tracé du centre de masse de la DSP en Hertz des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

La Figure 3-8 montre que les changements des formes d'ondes des événements considérés correspondent aux variations du centre de masse de la DSP. Sa valeur maximale est proche de 3 kHz pour de bruits d'ustensile et de 2,2 kHz pour des bruits de casseroles qui sont des sons métalliques. Elle atteint 1 kHz pour un claquement de porte qui est un son plus lourd. Pour un écoulement d'eau, cette valeur est située autour de 4 kHz. Il est alors intéressant de faire remarquer que l'évolution temporelle de la valeur du centre de masse pour un écoulement d'eau rejoint nos commentaires faits au sujet de sa DSP. Étant donné la fréquence d'échantillonnage  $F_s$ , égale à 16 kHz, on sait que notre spectre va nous renseigner sur l'énergie de la DSP pour des fréquences entre 0 et  $F_n=8$  kHz. Puisque notre bruit d'écoulement d'eau se rapproche d'un bruit blanc, son énergie sera répartie suivant toute la gamme des fréquences de la DSP. Il est donc normal que le centre de masse de cet événement sonore se situe aux alentours de 4 kHz.

Le centre de masse spectral de la DSP présente des valeurs différentes selon les types d'événements considérés. On ne peut donc pas s'en servir à des fins de détection d'événements sonores. Mais ce paramètre aura son utilité pour ce qui est de la caractérisation de nos objets sonores puisqu'il est indépendant des conditions d'enregistrement tout comme la largeur de bande et la fréquence de coupure spectrale puisque la DSP est normalisée.

### 3.3.8 La largeur de bande spectrale

La largeur de bande spectrale de la  $DSP$  notée  $B_{DSP}$  peut être vue comme le moment d'ordre 2 de la  $DSP$  normalisée [7, 13, 21, 22, 25]. Sa formule mathématique est la suivante :

$$B_{DSP} = \sqrt{\frac{\sum_{k=0}^{N_f/2-1} DSP_{norm}(k) * (k - C_{DSP})^2}{\sum_{k=0}^{N_f/2-1} DSP_{norm}(k)}}$$

Avec,  $C_{DSP}$  le centroïde spectral et  $B_{DSP}$  la largeur de bande spectrale de la  $DSP$ . La valeur de ce descripteur du son se mesure en indice, mais elle peut être transformée en *Hertz*. Sa valeur est aussi comprise entre 0 et  $F_n=8$  kHz. Il permet de dire si la forme de la  $DSP$  est plutôt concentrée ou étalée autour de son centre de masse. La Figure 3-9 présente l'évolution de la largeur de bande spectrale de la  $DSP$  pour les signaux sonores considérés.

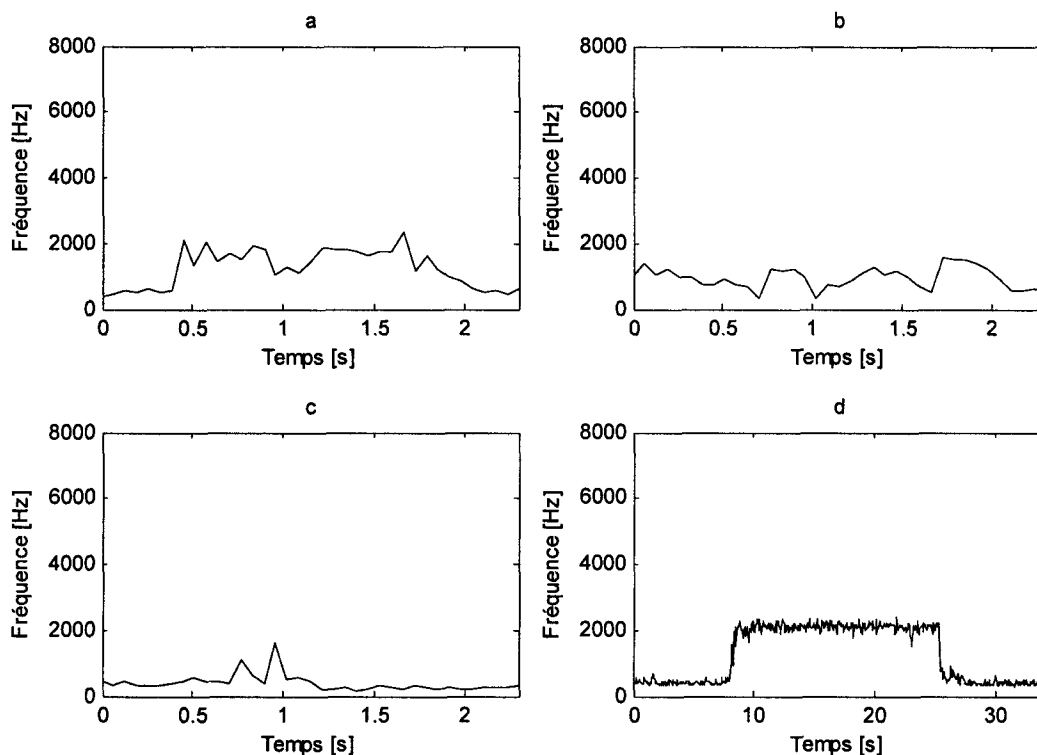


Figure 3-9 : Tracé de la largeur de bande de la  $DSP$  en *Hertz* des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

En observant la Figure 3-9, on voit que la largeur de bande des sons d'ustensile oscille autour de 2 kHz et que celle des sons de casserole avoisine les 1,5 kHz. Les bruits d'ustensile



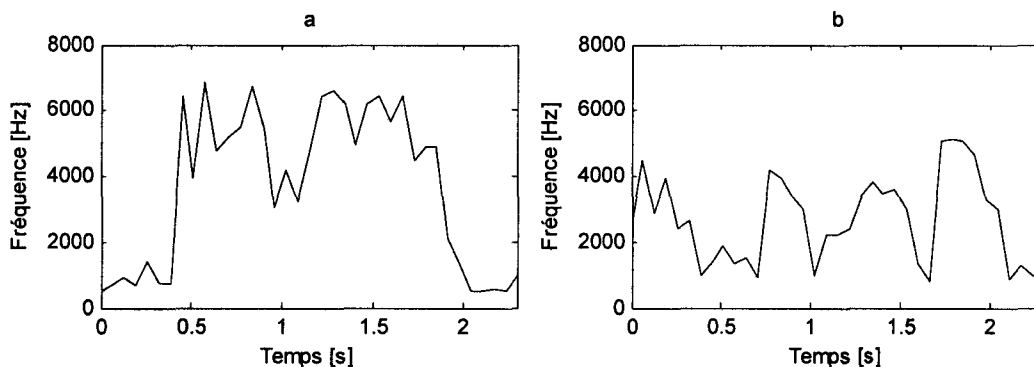
présentent donc un étalement des composantes spectrales de leurs *DSP* plus étalées autour du centre de masse que les bruits de casserole. Ceci peut se vérifier en se référant aux images spectrales présentées à la Figure 3-6. Pour ce qui est du bruit d'écoulement d'eau, la largeur de bande spectrale varie autour de 2 kHz. Il est aussi intéressant de faire remarquer que le bruit de claquement de porte engendre des harmoniques dus à la résonance du placard qui ont une largeur de bande spectrale égale à 1,6 kHz. Cette valeur est plus élevée que celle de l'accroche sonore qui est égale à 1 kHz. Ce paramètre sera utilisé dans le même but que le centre de masse spectral, c'est-à-dire pour caractériser nos événements sonores en vue de les classer.

### 3.3.9 La fréquence de coupure spectrale

La fréquence de coupure spectrale de la *DSP* notée  $F_{CDSP}$  est définie telle que 95% de l'énergie de la *DSP* normalisée soit concentrée en dessous de celle-ci [25] :

$$F_{CDSP} = \text{arg}_k \left\{ \sum_{k=0}^{k_c} DSP_{norm}(k) = 0,95 \sum_{k=0}^{N_f/2-1} DSP_{norm}(k) \right\}$$

La valeur de ce descripteur correspond à un indice que l'on transforme en *Hertz*. La fréquence de coupure spectrale est fonction de la décroissance de la *DSP* dans ces fréquences supérieures. Un son riche en hautes fréquences présentera une décroissance plus abrupte qu'un son concentré dans les zones fréquentielles plus basses. La fréquence de coupure spectrale sera donc plus élevée dans le premier cas que dans le second. La Figure 3-10 présente l'évolution temporelle de la fréquence de coupure.



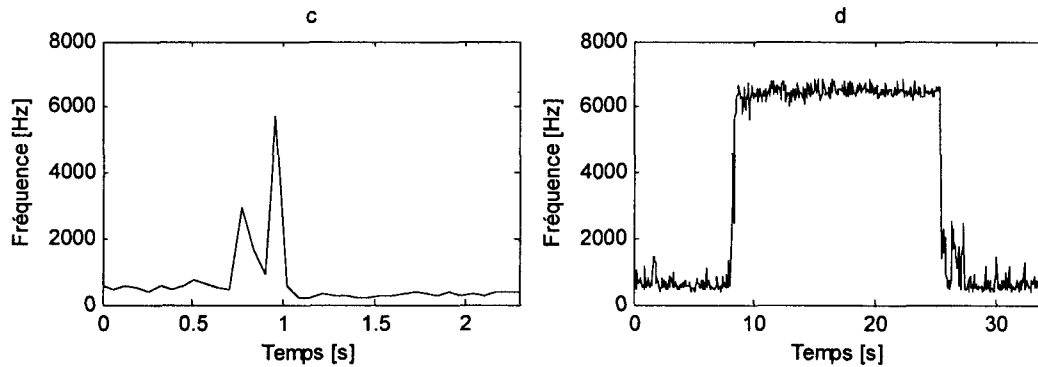


Figure 3-10 : Tracé de la fréquence de coupure spectrale de la *DSP* en *Hertz* des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

On peut faire les mêmes observations à la Figure 3-10 que pour la Figure 3-9. En effet, on voit que les formes d'ondes correspondant à des manipulations d'ustensile et de casserole possèdent une énergie fréquentielle plus importante dans les hautes fréquences qu'un bruit de claquement de porte. Pour un bruit de claquement de porte, la fréquence de coupure spectrale atteint  $3\text{ kHz}$  pour l'accroche sonore. Cependant, la résonance du claquement de porte a un contenu énergétique important dans les hautes fréquences, la fréquence de coupure spectrale avoisine alors  $5,8\text{ kHz}$ . La fréquence de coupure maximale d'un bruit d'ustensile se situe autour de  $6,8\text{ kHz}$ , ce qui est inférieur à celle d'un bruit de casserole qui avoisine  $5\text{ kHz}$ . L'écoulement d'eau du robinet présente une fréquence de coupure spectrale légèrement inférieure à  $7\text{ kHz}$ , car toutes les composantes fréquentielles sont présentes dans sa *DSP*. Ainsi, la fréquence de coupure spectrale servira tout comme le centre de masse et la largeur de bande spectrale à discriminer les événements sonores les uns des autres.

### 3.3.10 La densité spectrale d'énergie en bande critique

On dispose grâce à la *DSP* d'informations sur l'énergie de chaque composante spectrale de notre signal, chacune espacée de  $31,5\text{ Hz}$ . Au regard des *DSP* affichées à la Figure 3-6, on se rend compte qu'une précision fréquentielle de  $31,5\text{ Hz}$  n'est pas utile. On s'intéresse alors plus particulièrement à l'énergie assignée à différentes bandes fréquentielles comprises entre  $0$  et  $F/2$ , afin de visualiser avec plus de facilité les zones fréquentielles à fort contenu énergétique de celles à plus faible contenu [9, 12, 21, 22, 32, 34]. À cet effet, on introduit la *DSP* en bandes critiques, ainsi que l'échelle de *Mel*. On présente par la suite l'échelle de *Mel* qui permet d'adapter l'échelle des fréquences à l'audition humaine. Cette méthode vise à

quantifier l'énergie de la *DSP* en bandes fréquentielles appelées aussi bandes critiques. On formule alors la *DSP* décomposée en bandes critiques de la façon suivante :

$$DSP_l(b_i) = \sum_{k=k_i}^{k_{i+1}-1} DSP_l(k)$$

Avec,  $k_i$  l'indice correspondant à la première fréquence de la bande critique  $i$  et  $k_{i+1}$  l'indice correspondant à la première fréquence de la bande critique  $i+1$ . Le calcul de la *DSP* en bandes critiques a pour but d'assigner une seule valeur d'énergie par bande fréquentielle. Le but implicite étant de décimer l'information contenue sur les 256 points du spectre en seulement un certain nombre de valeurs d'énergie correspondant à la quantité d'énergie de la *DSP* comprise dans chaque bande fréquentielle. Ceci permettra de diminuer considérablement l'espace de stockage de nos données et le temps de calcul nécessaire à la capture de nos événements sonores. Il est aussi possible de jouer sur la résolution fréquentielle en augmentant ou en diminuant le nombre de bandes fréquentielles. Il y a donc un compromis à faire entre le temps de calcul, l'espace de stockage et la résolution fréquentielle de la *DSP* en bandes critiques.

L'échelle de *Mel* est une échelle de mesure des fréquences basée sur la perception humaine [8, 22, 26, 28, 32, 36]. Elle se mesure en *Mel*. Elle a été conçue de telle façon que 1000 Hz correspondent à 1000 *Mel*, et qu'une variation constante dans le domaine des *Mel* soit perçue comme une variation constante de la hauteur tonale par l'auditeur. En effet, au-delà de 500 Hz, l'oreille humaine ne perçoit plus le changement d'octave comme un doublement de la fréquence. Par contre, si l'on se place sur une échelle de *Mel*, un changement d'octave sera effectivement perçu comme un doublement de la fréquence. La formule qui effectue la conversion des *Hertz*  $f_{Hz}$  en *Mel*  $f_{Mel}$  est la suivante :

$$f_{Mel} = 2595 \log_{10}(1 + f_{Hz} / 700)$$

La Figure 3-11 présente le graphique de correspondance entre l'échelle de *Mel* et celle des *Hertz*. On a une fréquence maximale égale à  $F_n = 8 \text{ kHz}$ , ce qui correspond à 2840 *Mel*.

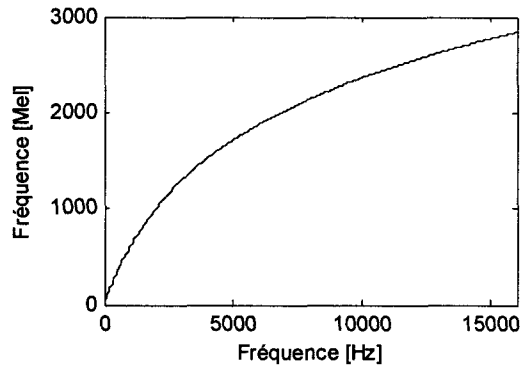
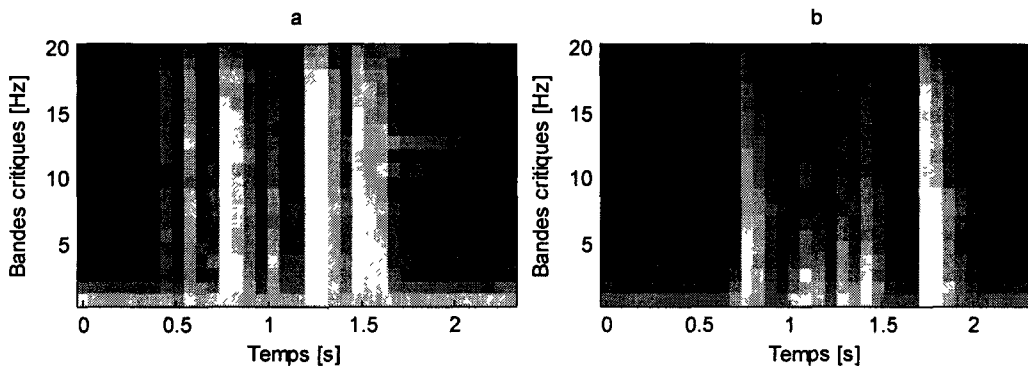


Figure 3-11 : Graphique de la loi de correspondance entre *Mel* et *Hertz*

On voit à la Figure 3-11 que la courbe de correspondance entre *Mel* et *Hertz*. C'est une courbe logarithmique qui peut être approximée par une droite en bas de 1000 Hz. Cette échelle a prouvé son utilité pour l'analyse fréquentielle des signaux sonores, puisqu'elle permet de se rapprocher au plus près de la perception humaine des sons. Dès lors, on peut décomposer l'échelle fréquentielle en bandes critiques suivant l'échelle de *Mel*. Dans notre étude on choisit de décomposer l'échelle fréquentielle en 20 bandes critique de *Mel* [32, 36]. À titre de comparaison, on décompose aussi l'échelle hertzienne en 20 bandes fréquentielles.

La Figure 3-12 et la Figure 3-13 présentent les images spectrales de la DSP découpées en bandes fréquentielles, que ce soit avec l'échelle des *Hertz* ou de *Mel*.



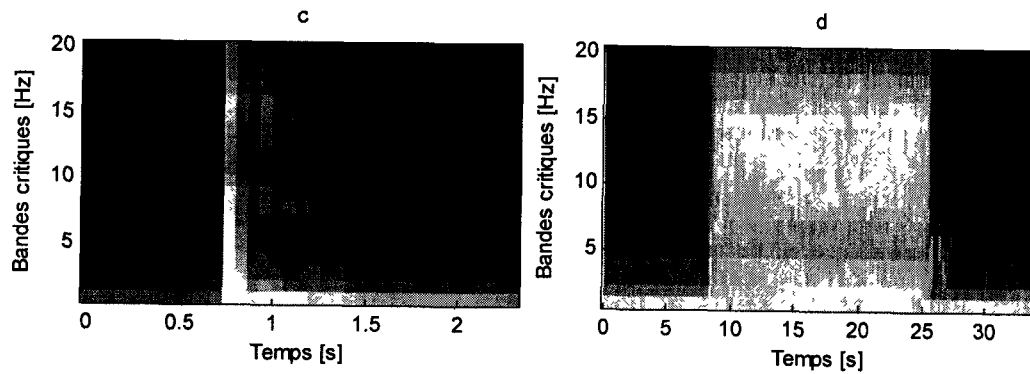


Figure 3-12 : Images de la *DSP* en 20 bandes hertziennes des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

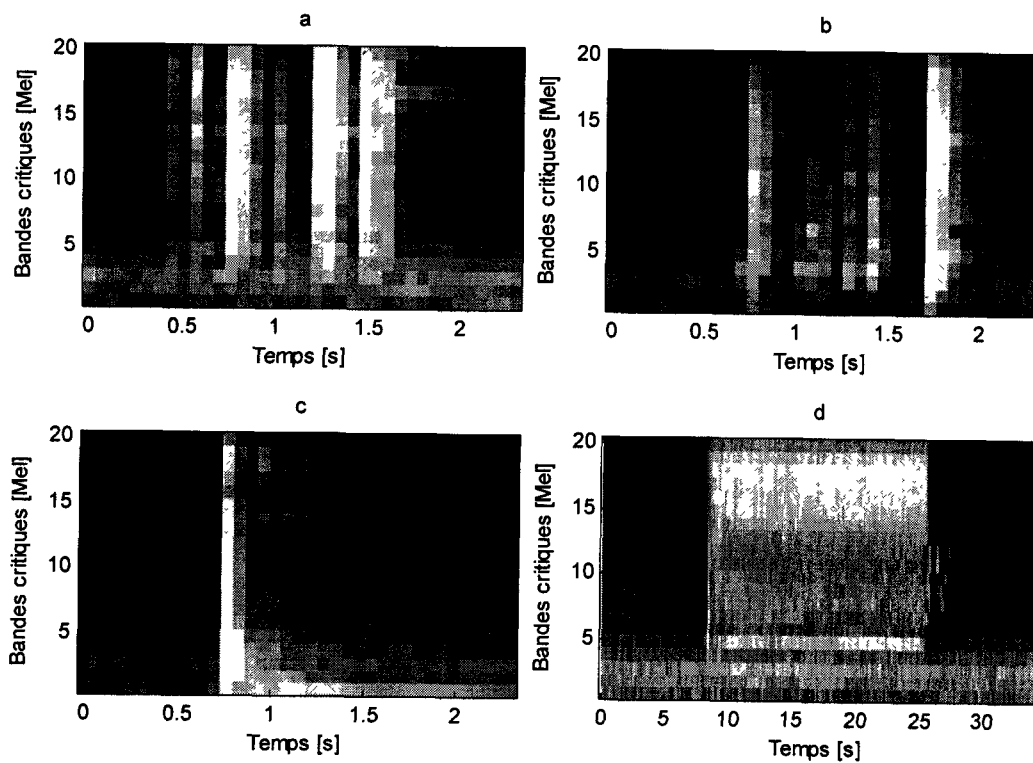


Figure 3-13 : Images de la *DSP* en 20 bandes critiques de *Mel* des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

La Figure 3-12 et la Figure 3-13 présentent une allure globale assez proche de la Figure 3-6. La discrétisation fréquentielle opérée par le découpage en bandes permet de visualiser plus facilement les changements d'énergie fréquentielle dans une bande de fréquence au cours du temps. Tout comme à la Figure 3-6, les images spectrales ont été normalisées.

À la Figure 3-13, on observe l'influence de l'utilisation de l'échelle de *Mel* pour le découpage des bandes critiques de la *DSP*. Ainsi, pour tous les événements sonores considérés, l'énergie de la *DSP* se déplace vers les bandes critiques de plus haute valeur compte tenu des propriétés de l'échelle de *Mel*. Par conséquent, les accroches sonores visualisables sur les images spectrales de la Figure 3-13 sont d'une plus grande largeur de bande fréquentielle que celles de la Figure 3-12. Il est alors possible à partir de la *DSP* découpée en bandes de discerner les accroches et chutes sonores de nos événements qu'ils soient de bande fréquentielle large ou étroite.

### 3.3.11 La transformée en ondelettes

La transformée en ondelettes permet une analyse temps-fréquence qui réalise un compromis entre les deux domaines d'étude temporel et fréquentiel. Ceci n'est pas permis par d'autres transformées par exemple la transformée de Fourier. Elle décompose le signal en fonctions sinusoïdales, périodiques et à support infini. Ainsi, la résolution fréquentielle est très précise, une impulsion de Dirac pour chaque fréquence. Par contre, cette représentation spectrale fait abstraction complète du temps, il n'est donc pas possible d'identifier les micros changements au sein d'un spectre au cours du temps. Néanmoins, la *TFSD* règle en partie le problème de stationnarité du signal sous la fenêtre d'analyse, ce qui a été expliqué au paragraphe 3.3.4.

La transformée en ondelettes décompose le signal suivant une famille d'ondelettes translatées et dilatées. Elle permet une bonne analyse des fonctions présentant des discontinuités ou des phénomènes locaux. En effet, les fonctions de base utilisées pour cette transformée sont à support fini et à largeur de bande parfois très grande ou périodiques lorsqu'elles sont exprimées au niveau spectral. Le choix de l'ondelette va déterminer les caractéristiques de la représentation de la transformée du signal, faisant ainsi ressortir certaines spécificités du signal. Ceci est un avantage puisque l'on s'intéresse tout particulièrement à l'analyse des transitoires dans le contexte de la détection des événements sonores.

### 3.3.12 Transformée en ondelettes continues

La transformée en ondelettes continues se présente comme la projection du signal sur une base d'ondelettes déterminée. Les coefficients en ondelettes de la fonction  $f(t)$  sont notés  $C$  et calculés de la manière suivante :

$$C(a, b) = \int_{-\infty}^{+\infty} f(t) \cdot \psi_{a,b}^*(t) dt$$

Où, \* désigne le conjugué,  $a$  le facteur de dilatation,  $b$  le facteur de translation et  $\psi_{a,b}(t)$  une base orthogonale obtenue par contraction ou dilatation d'une ondelette mère  $\Psi$  de la façon suivante :

$$\psi_{a,b}^*(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right)$$

Le paramètre  $b$  caractérise l'emplacement sur l'axe des temps de la fonction sur laquelle on projette le signal à analyser, c'est un décalage temporel. Le paramètre  $a$  caractérise l'échelle du temps. Le paramètre  $a$  est appelé facteur de dilatation et va servir à préciser le type de variation du signal que l'on veut observer. Une valeur de  $a < 1$  va permettre d'analyser des variations rapides sur une courte durée, alors qu'une valeur de  $a > 1$  va permettre d'analyser des phénomènes à variation lente sur une longue durée.

### 3.3.13 Transformée en ondelettes discrètes

Étant donné que l'on travaille sur des signaux numérisés, on utilise la transformée en ondelettes discrètes [5, 8, 17, 23, 24, 30, 32]. On choisit alors de fixer le facteur de dilatation  $a = a_0^j$  avec  $j$  le niveau de décomposition, et le facteur de translation  $b = kb_0$  avec  $k$  un entier relatif. Par la suite, on se place dans le cas d'une décomposition en ondelettes dyadique avec  $a_0 = 2$  et  $b_0 = 1$ . On définit alors l'énergie des coefficients de la transformée en ondelettes discrètes  $w_{j,l}$  de niveau  $j$  du segment sonore  $x_l$  tel que :

$$w_{j,l}(k) = 2^{-j/2} \sum_{n=1}^{N_l/j+1} |x_l(n) \cdot \psi(2^{-j}n - k)|^2$$

La décomposition en ondelettes dyadiques proposé par Mallat [29] peut être vue comme une batterie de filtres passe-bande en cascade qui découpe le signal en bandes de fréquences [23, 30]. Cette transformée correspond alors à un double filtrage passe-bas et passe-haut avec des bandes passantes complémentaires, puis un sous-échantillonnage par 2 du signal. Le signal filtré passe-bas va fournir une approximation du signal :  $a_j$ , alors que le filtrage passe-haut

donne le détail du signal :  $d_j$ ; avec  $j$  le niveau de décomposition. On applique ce traitement à l'approximation  $a$  du signal de façon itérative jusqu'au niveau de décomposition voulu. Le signal est donc filtré par une cascade de filtres passe-bas  $L$  et passe-haut  $H$ . La fréquence de coupure est donc divisée par 2 à chaque niveau de décomposition. On obtient donc une bonne résolution fréquentielle en basse fréquence et une bonne résolution temporelle en haute fréquence. Ce procédé est effectué pour chacune des fenêtres d'observation de notre signal. La Figure 3-14 illustre le principe de la décomposition en ondelettes discrètes du signal par un banc de filtres, et la répartition fréquentielle.

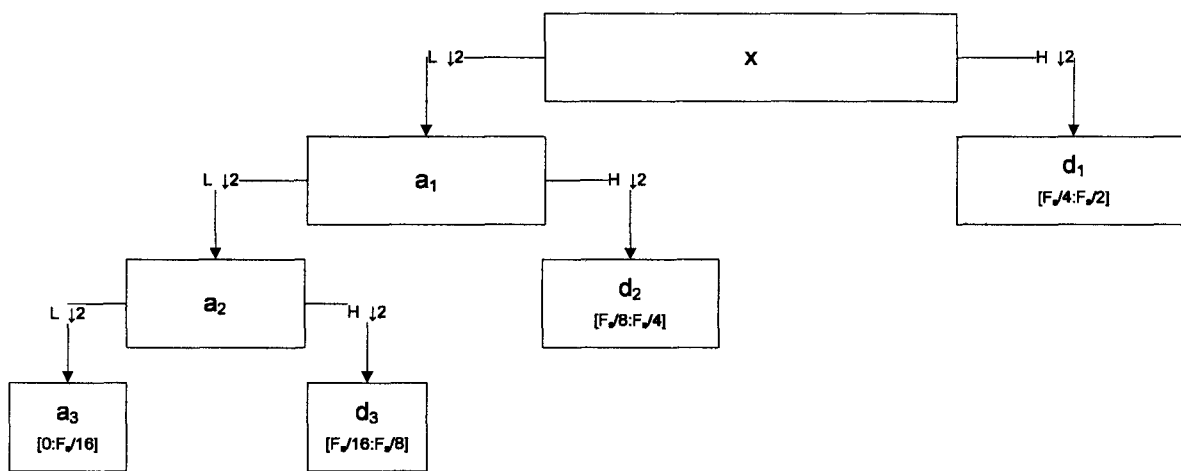


Figure 3-14 : Principe de la décomposition en ondelettes par une batterie de filtres selon l'algorithme de Mallat.

La Figure 3-15 présente sous forme d'image l'énergie des coefficients en ondelettes pour chaque niveau de décomposition en fonction du temps. On a choisi un niveau de décomposition  $N_w=3$ , et une ondelette mère de *Daubechies 3* [5, 8, 23]. On obtient alors les coefficients  $a_3$  à la bande 1,  $d_3$  à la bande 2,  $d_2$  à la bande 3 et  $d_1$  à la bande 4 en reprenant les notations de la Figure 3-14. Les images sont normalisées comme précédemment.



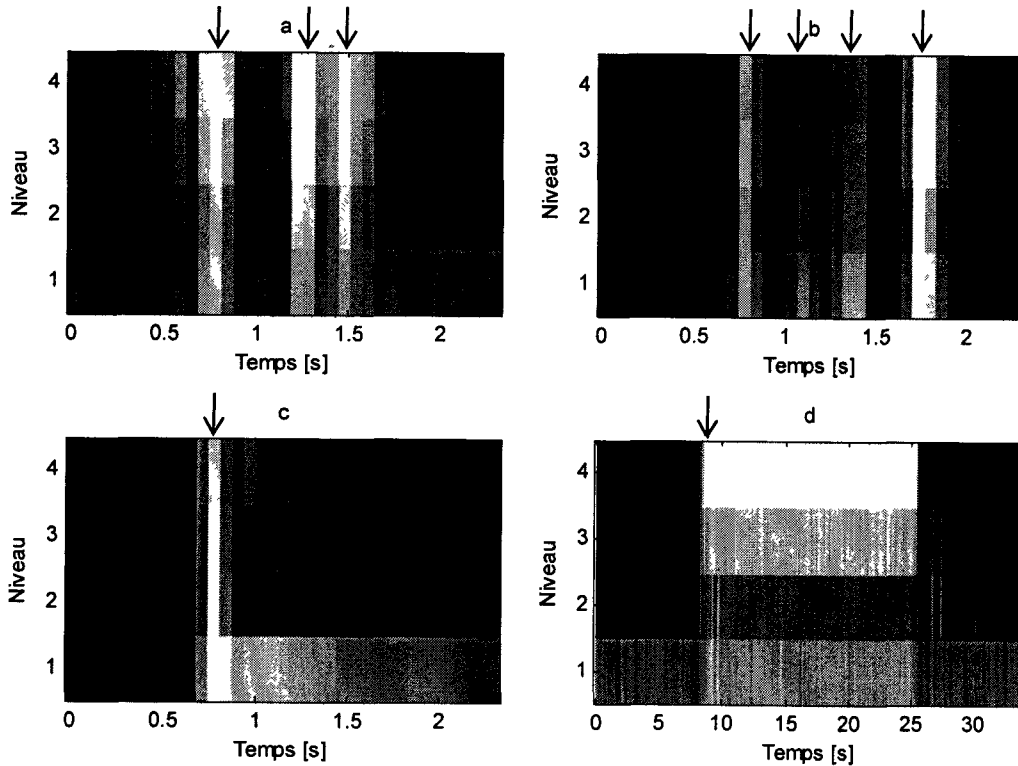


Figure 3-15 : Images de la décomposition en ondelettes dyadiques des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

On observe à la Figure 3-15 les images normalisées de l'énergie des coefficients issus de la décomposition en ondelettes. On observe, à l'aide des flèches, que les accroches et chutes sonores des différents événements sonores sont parfaitement identifiables. On distingue respectivement 3 et 4 accroches sonores pour les signaux de bruits d'ustensile et de casserole. De même, on observe l'accroche sonore des sons de claquement de porte et d'écoulement d'eau du robinet. La décomposition en ondelettes permet donc la mise en évidence des accroches sonores de façon significative. Ce paramètre paraît être d'une forte utilité pour l'analyse des transitoires dans le contexte de la détection des événements sonores. Par la suite, on prouvera aussi son efficacité pour ce qui est de la caractérisation des événements sonores.

### 3.3.14 Énergie globale des coefficients en ondelettes discrètes

Tout comme pour la *DSP*, on calcule l'énergie globale normalisée de la transformée en coefficients d'ondelettes. On lui applique un logarithme de base 10 et on l'exprime aussi en dB. Sa formule est la suivante :

$$E_w = 10 * \log_{10} \left( \sum_{j=1}^{N_w} \left( \frac{1}{N_t/j + 1} \sum_{k=1}^{N_t/j+1} w_{j,l}(k) \right) \right)$$

La Figure 3-16 présente l'énergie globale des coefficients en ondelettes pour les événements considérés à la Figure 3-1. Les paramètres utilisés pour la décomposition en ondelette sont les mêmes que précédemment.

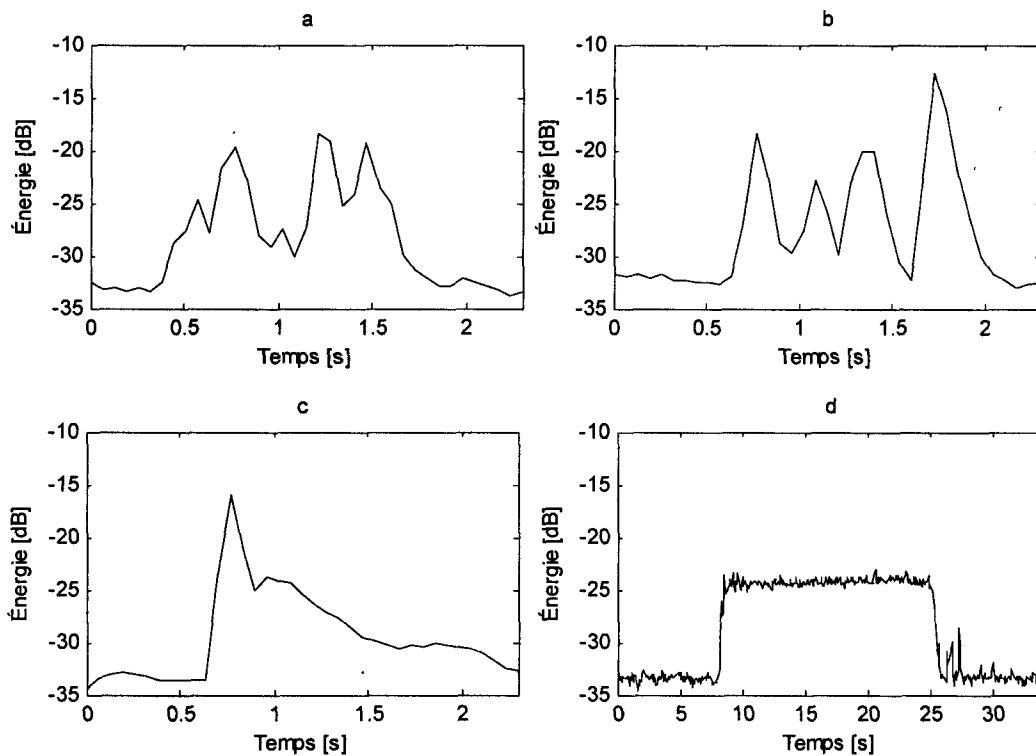


Figure 3-16 : Tracé de l'énergie globale des coefficients en ondelettes des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

En observant la Figure 3-16, on voit que l'évolution de l'énergie globale des coefficients en ondelettes est fortement corrélée avec les changements d'allure des formes d'ondes des signaux de la Figure 3-1. La valeur maximale de l'énergie globale des coefficients en ondelette se situe entre  $-13$  et  $-20$  dB pour des événements sonores comme la manipulation d'ustensiles et de casseroles. Pour un claquement de porte, cette valeur est égale à  $-15$  dB, alors qu'elle est de  $-24$  dB pour un écoulement d'eau. On s'aperçoit alors que les images de la Figure 3-16 sont similaires à celles de la Figure 3-7 et la Figure 3-4. De la même façon que les valeurs de l'énergie temporelle qui différaient de l'énergie globale de la DSP, la valeur de

l'énergie globale de la décomposition en ondelettes est différente aussi. En effet, les coefficients en ondelettes ont une précision fréquentielle plus élevée pour ce qui est des basses fréquences. La valeur de ce paramètre diffère selon les événements considérés ce qui limite son utilisation à un seuillage absolu de détection des événements sonores.

### 3.4 Méthodes utiles à la détection des accroches sonores

À partir de certaines méthodes de caractérisation de nos signaux comme la *DSP* et la décomposition en paquets d'ondelettes, il est possible d'extraire une information utile à la détection des accroches sonores des signaux. À cet effet, on utilise des méthodes permettant de mettre en évidence les transitoires de nos signaux sonores.

#### 3.4.1 Flux

Le flux permet de détecter les changements d'énergie entre 2 fenêtres d'analyse consécutives [13, 24, 31]. On applique cette méthode à chaque bande critique de *Mel* de la *DSP*, ou bien à chaque niveau de décomposition de l'énergie des coefficients en ondelettes. Cette méthode permet de rendre compte des changements d'énergie locaux qui ont lieu pour une bande critique ou pour les coefficients d'un niveau de la décomposition en ondelettes. La formule de calcul du flux est la suivante :

$$Flux_{param}(l) = \sqrt{(param_{l+1} - param_l)^2}$$

Avec *param* le paramètre d'étude auquel on applique le flux. La Figure 3-17 et la Figure 3-18 présentent respectivement les images du flux spectral résultant de l'étude de la *DSP* en bandes critiques de *Mel*, et l'énergie des coefficients de la décomposition en ondelettes pour les événements sonores de la Figure 3-1. Les paramètres choisis pour le calcul de la *DSP* et de l'énergie des coefficients de la décomposition en ondelettes sont les mêmes que précédemment. On fait aussi remarquer que les images ne sont plus normalisées dans ce cas. Il est alors possible de comparer la mise en évidence des accroches sonores pour les différents types d'événements sonores considérés.

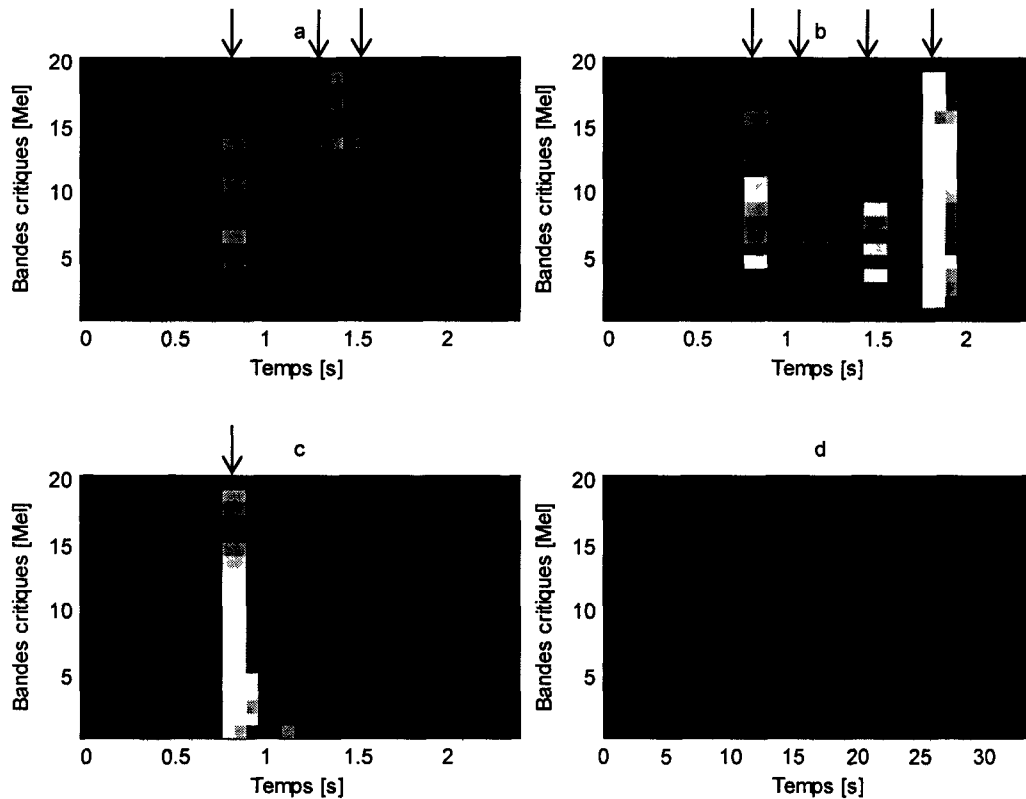


Figure 3-17 : Images du flux de la DSP en bandes critiques de Mel des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

Sur les images de la Figure 3-17, il est possible de discerner les accroches sonores des signaux considérés. Ainsi, pour les bruits d'ustensile et de casserole, on distingue, à l'aide de flèches, respectivement 3 et 4 accroches successives. Ces accroches sonores correspondent à des changements d'énergie sur une gamme de fréquences plus ou moins étalée. Par exemple, pour le bruit d'ustensile, ces accroches sont globalement visibles entre la 5<sup>ème</sup> et la 19<sup>ème</sup> bande critique. Tandis que pour un bruit de casserole, la 1<sup>ère</sup> la 3<sup>ème</sup> et la 4<sup>ème</sup> accroche sont globalement visibles entre la 3<sup>ème</sup> et la 19<sup>ème</sup> bande critique. Ce sont des accroches sonores large bande. Cependant, la 2<sup>ème</sup> accroche sonore est visible seulement entre la 6<sup>ème</sup> et la 8<sup>ème</sup> bande, c'est une accroche à bande étroite. Cela s'explique par le fait que les signaux sonores générés par une casserole qui s'entrechoque avec une autre casserole ou avec un rangement sont différents au niveau de leur contenu fréquentiel. Un bruit de claquement de porte possède une accroche sonore large bande située entre la 1<sup>ère</sup> et la 19<sup>ème</sup> bande critique. On remarque aussi qu'il n'est pas possible de discerner de façon visuelle les accroches et les chutes sonores d'un écoulement d'eau à la Figure 3-17

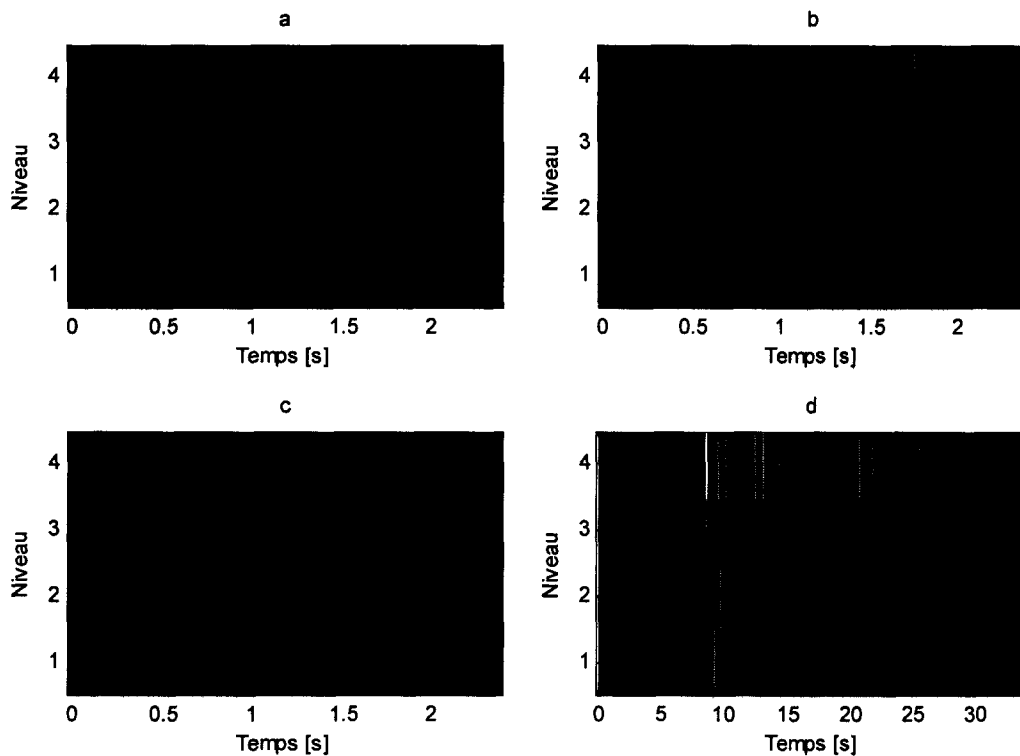


Figure 3-18 : Image du flux de l'énergie des coefficients en ondelettes des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

À l'œil nu, les images de la Figure 3-18 mettent en évidence les mêmes transitoires que la Figure 3-17. Cependant, on visualise aisément les accroches et chutes sonores d'un signal d'écoulement d'eau.

### 3.4.2 L'opérateur « Teager-Kaiser »

L'opérateur « *Teager-Kaiser* » noté  $tg$  est un opérateur mathématique non linéaire qui permet d'amplifier les changements d'allure d'un signal [30]. Cet opérateur est appliqué à la *DSP* en bandes critiques de *Mel* et à l'énergie des coefficients issus de la décomposition en ondelettes. Pour une valeur de paramètre  $param$  calculée pour une trame  $l$ , l'opérateur « *Teager-Kaiser* » noté  $tg_{param}(l)$  se calcule de la manière suivante :

$$tg_{param}(l) = param(l)^2 - param(l-1) * param(l)$$

On présente respectivement à la Figure 3-19 et à la Figure 3-20 l'influence de l'opérateur  $tg$  appliqué à la *DSP* en bandes critiques de *Mel*, et à l'énergie des coefficients en ondelettes.

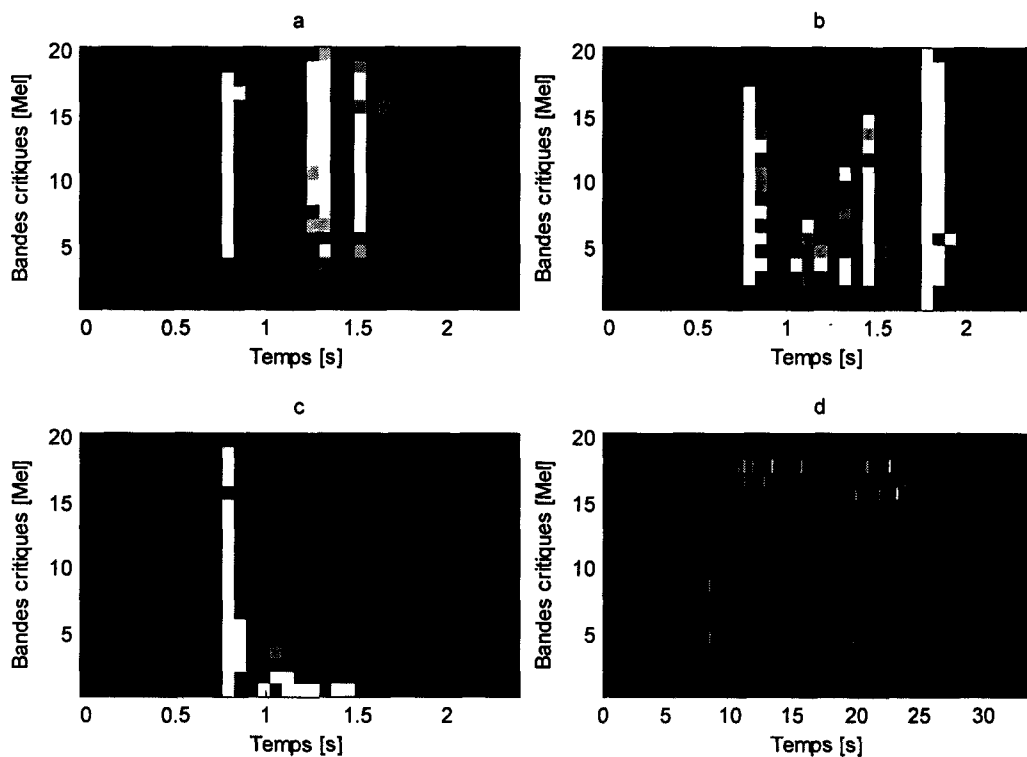
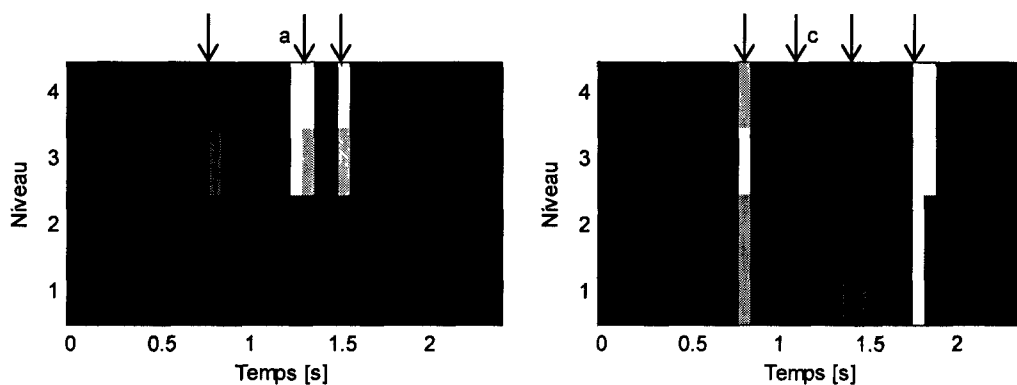


Figure 3-19 : Images de la DSP en bandes critiques de Mel à laquelle on a appliqué l'opérateur  $tg$  des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

À la Figure 3-19, on distingue que les accroches sonores des différents événements sonores se situent dans les mêmes bandes fréquentielles qu'à la Figure 3-17. Cependant, les accroches sonores sont mises en relief de façon plus significative avec l'opérateur « *Teager-Kaiser* ». Ainsi, il est possible de visualiser les multiples accroches sonores de l'écoulement d'eau.



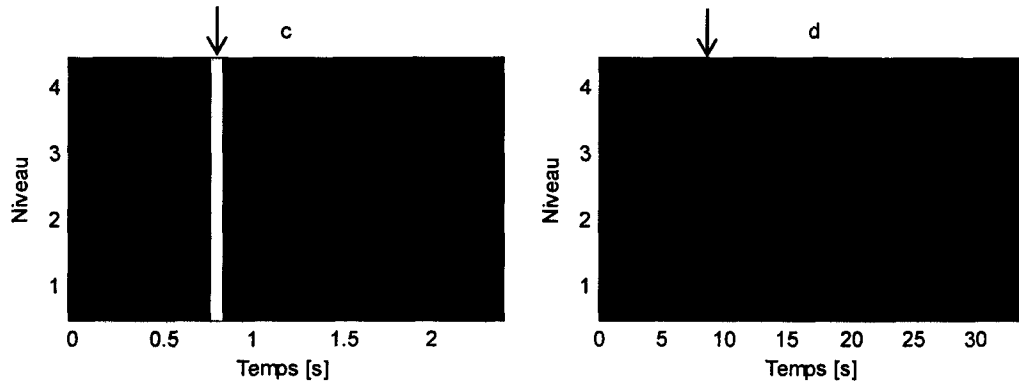


Figure 3-20 : Images de l'énergie des coefficients en ondelettes à laquelle on a appliqué l'opérateur  $tg$  des signaux de manipulation d'ustensile (a), de casserole (b), de claquement de porte de placard (c) et d'un écoulement d'eau du robinet (d)

À la Figure 3-20, les images présentées permettent la détection des accroches sonores de façon visuelle. À l'aide des flèches, on retrouve les 3 accroches sonores d'un bruit de manipulation d'ustensile, les 4 accroches de la manipulation de casserole et 1 accroche pour un claquement de porte. Pour ce qui est de l'écoulement d'eau, on distingue la 1<sup>ère</sup> accroche sonore à l'aide de l'énergie des coefficients de  $a_3$ .

L'identification visuelle des accroches sonores est donc plus aisée si l'on caractérise nos signaux à l'aide de l'énergie des coefficients de la décomposition en ondelettes plutôt qu'avec la *DSP* en bandes critiques de *Mel*. De plus, l'opérateur « *Teager-Kaiser* » paraît mieux adapté que le flux pour la mise en évidence des accroches sonores.

À partir des informations apportées par les méthodes de détection des transitoires, il faut maintenant être en mesure d'extraire, de l'enregistrement continu, les événements sonores correspondant à une action de l'habitant. On met donc en œuvre une technique de seuillage qui va permettre de discriminer les événements sonores du silence, tout en prenant en considération l'environnement sonore de l'habitat.

### 3.4.3 Calcul du seuil de détection via un circuit intégrateur

La technique de seuillage utilisée pour la détection des événements sonores est un circuit intégrateur du 1<sup>er</sup> ordre [13]. Cette technique prend en compte l'évolution temporelle de l'environnement sonore dans lequel on effectue notre enregistrement sonore. On applique en entrée de ce circuit intégrateur la valeur de paramètre de détection de nos transitoires. La formule mathématique appliquée est la suivante :

$$PARAM(l) = (1 - \gamma) \cdot |param(l)|^2 + \gamma \cdot PARAM(l - 1)$$

Avec *param* la valeur du paramètre de la trame courante, *PARAM* la valeur du paramètre intégré, et  $\gamma$  le coefficient de lissage temporel. Il est alors important de souligner que le paramètre  $\gamma$  permet de contrôler la pondération de la valeur du paramètre de la trame courante par rapport au paramètre intégré. Ceci permet donc de fixer un seuil de détection variable en fonction du différentiel d'amplitude des transitoires, et de l'environnement sonore. En ajustant le coefficient de lissage temporel  $\gamma$  et en choisissant un paramètre *param* adapté, on crée une fonction de seuillage pour la détection des accroches sonores. L'efficacité de cette méthode de seuillage sera testée et illustrée aux paragraphes 5.2 et 5.3. Il est important de faire remarquer que le calcul d'un seuil de détection via un circuit intégrateur nécessite un délai d'initialisation avant que celui-ci puisse calculer un seuil représentatif de notre environnement sonore.

À partir de là, il sera possible d'extraire les événements sonores de notre enregistrement sonore continu, puis de caractériser ces objets sonores à l'aide des paramètres spectraux et temporels pour enfin pouvoir les classer.

### 3.5 Méthodes utiles à la reconnaissance des événements sonores

La reconnaissance des événements sonores se base sur un ensemble de paramètres temporels et fréquentiels calculés à partir des objets sonores extraits. L'objectif de cette partie est de discriminer les différentes classes d'événements. Dans notre étude, on considère des événements sonores tels qu'un écoulement d'eau du robinet, une manipulation de casserole, d'ustensiles et d'assiette ainsi qu'un claquement de porte. C'est-à-dire que l'on souhaite créer 5 classes d'objets sonores différenciables les unes des autres.

Le système de reconnaissance des objets sonores se décompose alors en 2 étapes. Tout d'abord, on crée des dictionnaires de références pour chacun des paramètres, et ce pour chaque événement sonore considéré. À noter qu'un dictionnaire est constitué d'un nombre fixe de références. Ces dictionnaires de références sont conçus à partir d'une partie des enregistrements de notre base de données, c'est la base d'apprentissage de notre système de reconnaissance. Par la suite, l'étape de classification associe chaque objet sonore à une référence de l'un des dictionnaires. La classification des événements sonore se fait à partir



d'une partie des enregistrements de notre base de donnée différente de celle utilisée pour l'apprentissage de notre système de reconnaissance, on l'appelle la base d'évaluation.

### 3.5.1 Création des références via l'algorithme des *k-moyennes*

On s'intéresse dans un premier temps à la façon dont on crée le dictionnaire des références. Cette étape est répétée pour chaque paramètre d'étude et chaque classe d'événements sonores. L'un des algorithmes les plus utilisés à cet effet est l'algorithme des *k-moyennes* connu aussi sous le nom d'algorithme de Lloyd généralisé [11]. On explique alors les différentes étapes de cet algorithme :

- 1- On choisit aléatoirement  $k$  vecteurs de centre de masse initiaux  $c_1, \dots, c_k$  avec  $k$  le nombre de références.
- 2- On affecte chacun des vecteurs  $param$  au groupe  $i$  dont le vecteur de centre de masse  $c_k$  est le plus proche.  $c_k$  sera alors une approximation du vecteur  $param$ , on le note  $\widehat{param}$ .
- 3- On calcule ensuite la distorsion, notée  $d_n$ , entre tous les vecteurs  $param$  et leur vecteur de centre de masse associé  $\widehat{param}$  :

$$d_n = \frac{1}{L} \sum_{l=1}^L d(param_l, \widehat{param}_l) = \frac{1}{L} \sum_{l=1}^L (param_l - \widehat{param}_l)^2$$

Avec,  $d(,)$  l'opérateur qui représente l'erreur quadratique et  $n$  l'itération en cours.

- 4- Si la distorsion est telle que:  $(d_{n-1} - d_n)/d_n < \varepsilon$  , alors  $C = \{c_1, \dots, c_k\}$  est le dictionnaire souhaité et on stoppe.
- 5- Sinon, on calcule, pour tous les groupes  $i$ , les nouveaux centres  $c_i$  qui sont la moyenne des éléments du groupe  $i$ .
- 6- Revenir à l'étape 2.

La Figure 3-21 représente le schéma de principe de l'algorithme des *k-moyennes* et reprend l'explication précédente à l'aide des numéros des différentes étapes.

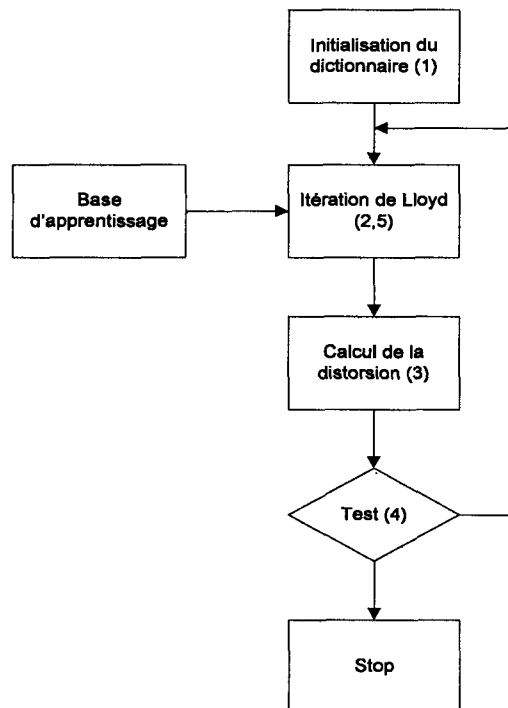


Figure 3-21 : Schéma de principe de la création des dictionnaires par l'algorithme des *k-moyennes*

Au regard de l'algorithme *k-moyennes* de création des dictionnaires, on se rend bien compte que les vecteurs de paramètre doivent être de taille égale. Cependant, la taille de ces vecteurs dépend directement de celle des objets sonores détectés. Or, on sait qu'un bruit d'écoulement d'eau aura une durée supérieure à un bruit de claquement de porte. Pour contourner ce problème, on choisit de ré-échantillonner nos vecteurs *param* en fonction d'un facteur de ré-échantillonnage pour qu'ils aient tous la même taille. Ainsi, selon les objets sonores, les vecteurs de paramètres seront soit sous-échantillonnés, soit sur-échantillonnés afin qu'ils aient tous la même taille, rendant ainsi possible la quantification vectorielle.

### 3.5.2 Classification des événements sonores

Une fois les dictionnaires de références créés, on passe à la phase d'évaluation de notre système de reconnaissance. On va alors tester les dictionnaires des références construits précédemment. Pour cette étape de classification, on utilise l'algorithme des plus proches voisins pour encoder nos vecteurs *param* de l'espace des observations. Cette méthode est connue sous le nom de « *Nearest Neighbor* », NN, en anglais [1, 13, 26, 32]. Le principe de cet algorithme se base sur la minimisation d'une distance, dans notre cas l'erreur quadratique

moyenne. Cette distance est calculée entre chaque paramètre de référence de nos dictionnaires et les vecteurs de paramètre d'entrée *param* des objets sonores extraits de la base de données d'évaluation. Ensuite, on remplace nos vecteurs d'entrées par la référence qui minimise la distance calculée, c'est-à-dire la référence la plus proche de nos vecteurs de paramètre. De cette manière, il est possible d'étiqueter chacun des objets sonores à l'aide des dictionnaires de références créés pour chaque type d'événement sonore.



## CHAPITRE 4. RÉALISATION

Tout d'abord, on précise que le système de détection et de reconnaissance des objets sonores a été implémenté en *Matlab* et on effectue une analyse hors ligne de notre enregistrement sonore continu. C'est-à-dire que l'on a dans un premier temps effectué les enregistrements sonores puis on a appliqué ces enregistrements en entrée du système de détection et de reconnaissance des objets sonores. On précise alors que les enregistrements utilisés pour l'apprentissage de notre système de reconnaissance des objets sonores sont différents de ceux utilisés pour la phase d'évaluation. Pour tester le système de détection des événements sonores, l'intégralité des enregistrements est utilisée puisque le système ne nécessite pas d'entraînement.

### 4.1 Système de détection des objets sonores

Le système de détection des objets sonores a été réalisé à l'aide de 2 fonctions de seuillage. D'une part, on calcule une fonction de seuillage pour la détection des accroches sonores. D'autre part, on calcule une fonction de seuillage absolu qui permet la détection des chutes sonores. Ce seuillage permet aussi d'éliminer les événements sonores de faible puissance qui ne sont pas indicateurs d'une action de l'habitant.

#### 4.1.1 Fonction de seuillage absolu

Pour effectuer un seuillage absolu des événements sonores, on a choisi des paramètres qui rendent compte de l'évolution de l'énergie des signaux sonores. On opte alors pour des paramètres tels que l'énergie globale de la *DSP* notée  $E_{DSP}$ , et l'énergie globale des coefficients de la décomposition en ondelettes notée  $E_w$ .

Ainsi, pour une trame sonore donnée, si la valeur de paramètre absolu noté  $param_{abs}$  est inférieure à la valeur seuil constante notée  $seuil_{abs}$ , alors aucun événement sonore ne pourra être détecté. Dans le cas contraire, un événement sonore sera détecté s'il respecte la condition de détection des accroches sonores. Les valeurs des seuils absolus de détection ont été fixées de façon empirique après visualisation des simulations effectuées. Ces valeurs sont les suivantes :

$$seuil_{abs} [E(w) ] = -52 \text{ dB}$$

$$seuil_{abs} [E(DSP) ] = -22 \text{ dB}$$

#### 4.1.2 Fonction de seuillage des accroches sonores

Les paramètres utilisés pour la détection de seuillage des accroches sonores sont la *DSP* en 20 bandes de *Mel*, et l'énergie des coefficients de la décomposition en ondelettes de niveau 3 avec une ondelette de *Daubechies 3*. Ces paramètres nous renseignent sur le contenu fréquentiel et l'évolution temporelle du signal sonore. Afin de faire ressortir davantage les changements brusques d'allure de nos paramètres, on leur applique l'opérateur « *Teager-Kaiser* ». On obtient alors un nouveau paramètre noté  $param_1$  qui servira à la détection des accroches sonores. On applique alors ce paramètre en entrée d'un circuit intégrateur du 1<sup>er</sup> ordre pour obtenir une fonction de seuillage des accroches sonores notée  $seuil_1$ . La valeur de coefficient de lissage temporel  $\gamma_1$  est fixée à 0.9.

On considère alors une valeur  $param_1$  pour un segment sonore donné. Si celle-ci est supérieure à la valeur  $seuil_1$  et que la condition de seuillage absolue est respectée, on considère que la trame contient une accroche sonore, et l'on commencera la capture de l'événement sonore.

On présente à la Figure 4-1 le schéma de principe du calcul des paramètres et des fonctions de seuillage de notre système de détection des transitoires. On a identifié sur le dessin les données que l'on va passer en entrée de notre algorithme de détection. On dessine les seuils de détection calculés en traits pointillés, et en traits pleins les paramètres de détection que l'on va comparer aux seuils.

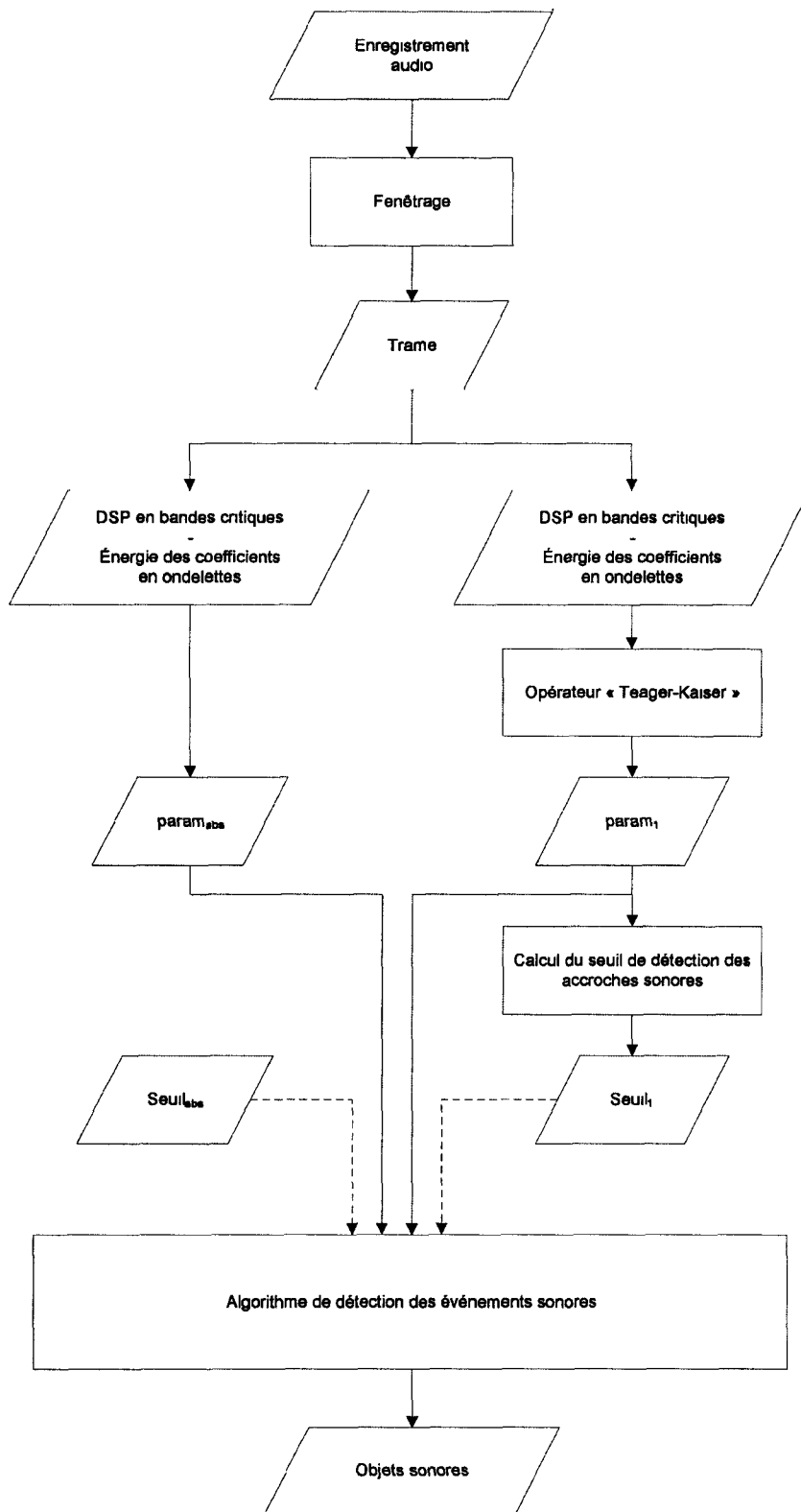


Figure 4-1 : Schéma de principe du calcul des fonctions de seuillage du système de détection des objets sonores

Par la suite, on explique plus en détail le procédé mis en place par l’algorithme de détection des objets sonores qui se base sur la fusion des paramètres et de seuils de détection  $param_{abs}$ ,  $seuil_{abs}$ ,  $param_1$ ,  $seuil_1$ .

### 4.1.3 Algorithme de détection des objets sonores

L’algorithme de détection des objets sonores se décompose en 2 parties. Dans un premier temps, il va détecter si une trame contient une accroche sonore ou non. La capture de l’objet sonore commence, si et seulement si, le seuil absolu est dépassé et si la condition de détection des accroches sonores est remplie. Dans le cas contraire, il va rechercher dans les trames suivantes si c’est le cas. Une fois la détection d’une accroche sonore effectuée, l’algorithme vérifie dans toutes les trames suivantes si une chute sonore est présente, c’est-à-dire si la valeur du paramètre absolu est inférieure au seuil de détection absolu. Dans ce cas, l’enregistrement de l’objet sonore est stoppé. Ce procédé est effectué continuellement sur l’ensemble des trames de l’enregistrement sonore continu.

La Figure 4-2 et la Figure 4-3 reprennent le principe de l’algorithme de détection des événements sonores, en illustrant respectivement la détection des accroches et chutes des objets sonores.

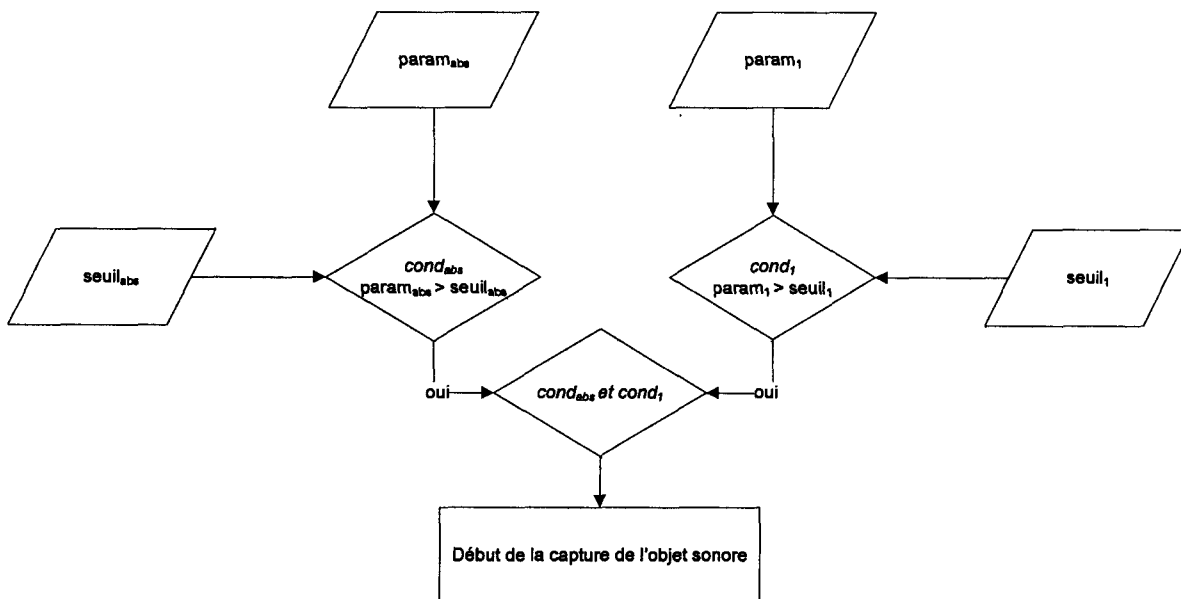


Figure 4-2 : Schéma de principe de détection des accroches sonores



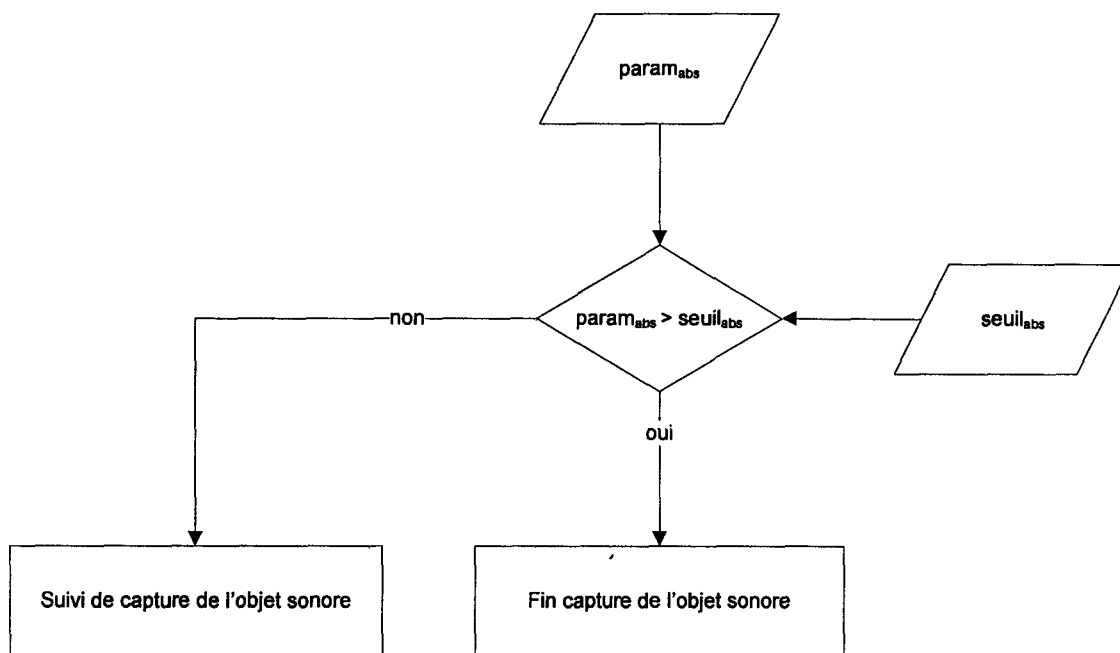


Figure 4-3 : Schéma de principe de détection des chutes sonores

## 4.2 Système de reconnaissance des objets sonores

Le système de reconnaissance des objets sonores se base sur les résultats du système de détection des événements sonores. Ainsi, on lui applique en entrée les objets sonores extraits de l'enregistrement continu de l'environnement sonore de l'habitat. L'élaboration du système de reconnaissance des objets sonores se décompose en 2 parties. D'une part l'apprentissage, et d'autre part l'évaluation du système.

### 4.2.1 Calcul des paramètres de classification

Les paramètres testés pour la classification des objets sonores sont : l'énergie des coefficients en ondelettes pour une décomposition de niveau 3 avec une ondelette de *Daubechies 3*, la *DSP* en 20 bandes critiques de *Mel*, ainsi que le centre de masse, la largeur de bande et la fréquence de coupure spectrale de la *DSP*. Ces paramètres sont calculés à partir d'un ensemble d'objets sonores constituant la base d'apprentissage de notre système. La Figure 4-4 présente le schéma de principe du calcul des paramètres utiles à la classification des objets sonores. On note l'ensemble de ces paramètres *param<sub>class</sub>*.

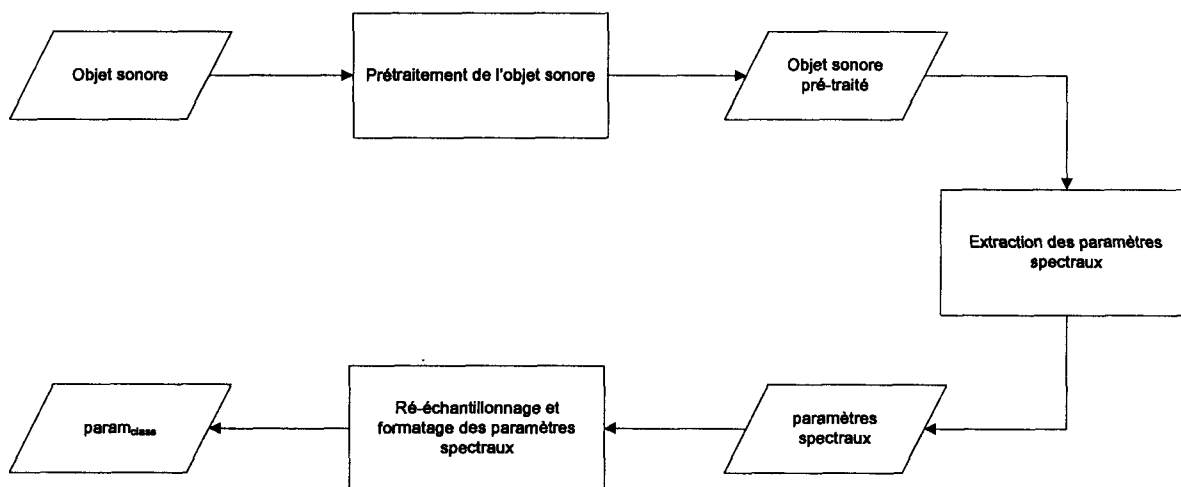


Figure 4-4 : Schéma de principe du calcul des paramètres de classification

Il est indiqué à la Figure 4-4 que l'on prétraite nos objets sonores en entrée du système de reconnaissance. Chacune des formes d'onde est alors normalisée par rapport à son maximum, et on s'arrange aussi pour que la moyenne des amplitudes soit nulle. De la même façon, on applique un post-traitement aux paramètres calculés pour chacun des objets sonores. On a spécifié, au paragraphe 2.2.3, que les vecteurs des paramètres de classification doivent être de même taille. Étant donné que chaque objet sonore est de taille différente, on a choisi de fixer une taille égale à 50 trames pour chacun de nos paramètres de classification calculés. Cette valeur correspond à une durée d'événement sonore de 1,6 s. Pour l'image spectrale de la *DSP* en bandes critiques de *Mel*, ce ré-échantillonnage se fait pour chacune des bandes critiques,  $param_{class}$  sera donc un vecteur de taille :  $20 \cdot 50 = 1000$ . Pour l'image de l'énergie des coefficients en ondelettes, le vecteur  $param_{class}$  aura une taille de :  $8 \cdot 50 = 400$ . Les vecteurs  $param_{class}$  représentatifs du centre de masse, de la largeur de bande et de la fréquence de coupure spectrale auront une taille égale à 50. De plus, il est important de préciser que les images de la *DSP* et de l'énergie des coefficients en ondelettes sont normalisées entre 0 et 1. Tous les paramètres sont en effet normalisés ce qui les rend cohérents d'un vecteur à l'autre. La pondération de chaque élément au sein d'un vecteur est la même, c'est-à-dire qu'ils ont tous le même poids dans la décision. Cette étape est indispensable afin de rendre cohérente la classification de nos objets sonores. Par exemple, on considère 2 objets sonores ayant des énergies fréquentielles très différentes, mais réparties de la même façon. Si l'on omet la normalisation de nos données, ces objets sonores seraient classés dans 2 classes différentes. Les classes créées par l'algorithme des *k moyennes* seraient uniquement fonction de l'énergie

fréquentielle sans prendre en compte la répartition en fréquence de celle-ci. Pour ce qui est du centre de masse, de la largeur de bande et de la fréquence de coupure spectrale, ils sont déjà tous normalisés entre 0 et  $F_s/2=8kHz$ .

#### 4.2.2 Création des dictionnaires de références

C'est la phase d'apprentissage de notre système de reconnaissance des objets sonores. On considère alors 5 types d'événement sonore : un bruit de manipulation de casserole, d'ustensile, d'assiette, d'écoulement d'eau et de claquement de porte. On a choisi ces événements sonores car ils sont représentatifs des actions les plus fréquentes réalisées par une personne au sein de la cuisine. On attribue alors un dictionnaire de références à chacun des objets sonores. Les références des dictionnaires sont créées grâce à l'algorithme des *k-moyennes* dont le principe a été expliqué au paragraphe 2.2.3. On obtient alors 5 dictionnaires constitués d'un nombre fixe de références. Dans notre étude, on décide d'attribuer 4 références à chacun des dictionnaires. Ceci pour prendre en compte les différentes formes d'onde associées à un même type d'événement. Ainsi, on aura 4 références qui constitueront le dictionnaire associé à un bruit de manipulation d'ustensile, et ceci pour chacun des objets sonores considérés dans l'étude.

#### 4.2.3 Classification des objets sonores

C'est la phase d'évaluation de notre système de reconnaissance des objets sonores. La classification se fait à partir des paramètres de classification calculés pour chaque objet sonore. Ces paramètres sont alors comparés avec les dictionnaires des références. On attribue à chaque objet sonore une action sonore qui correspondra à l'étiquette du dictionnaire dont la référence sera la plus proche de notre observation. La classification se base sur la minimisation de l'erreur quadratique expliquée au paragraphe 2.2.3. La Figure 4-5 décrit les différentes étapes de la classification d'un objet sonore.

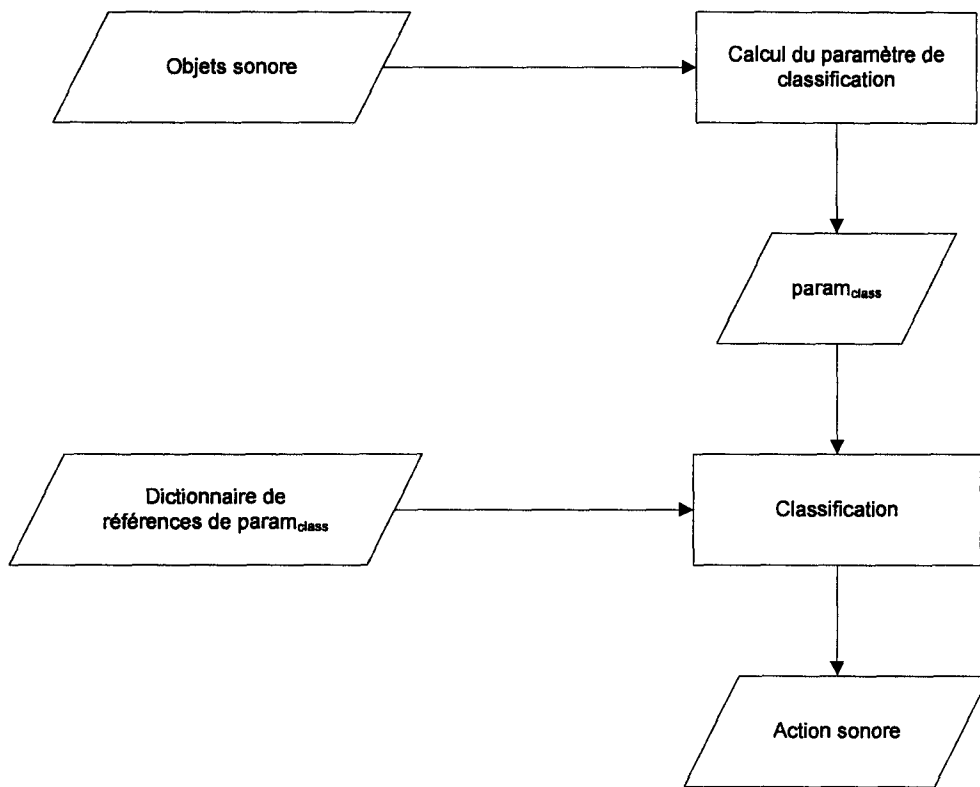


Figure 4-5 : Schéma de principe de la classification des objets sonores

## CHAPITRE 5. TEST ET ANALYSE

### 5.1 Test du système de détection des objets sonores

Afin de vérifier la robustesse de notre algorithme de détection des objets sonores, on applique en entrée de notre système un enregistrement d'une durée totale de 4 minutes et 45 secondes résultant de la concaténation des événements sonores enregistrés dans la cuisine. Celui-ci est successivement constitué de bruits de manipulation de casserole, d'assiette, de cuisson d'aliment, de manipulation d'ustensile, d'écoulement d'eau du robinet et de claquement de porte. On fait remarquer que l'enregistrement test comprend des bruits de cuisson d'aliment qui sont pris en compte uniquement par le système de détection des événements sonores. Le système de reconnaissance ne traitera pas ce type d'événement sonore. La Figure 5-1 présente l'enregistrement de test.

Dans un premier temps, on utilise l'énergie des coefficients en ondelettes pour une décomposition en ondelettes de niveau 3 avec une ondelette de *Daubechies 3*. Par souci d'homogénéité, le paramètre absolu est l'énergie globale des coefficients en ondelettes. Pour le calcul du seuil de détection des accroches sonores, on prend une valeur de coefficient de lissage temporel  $\gamma$  égale à  $0,9$ . Les résultats sont donnés à la Figure 5-2.

Dans un second temps, on reprend la même expérimentation, mais cette fois on choisit comme paramètre la *DSP* découpée en 20 bandes critiques de *Mel*. Du même coup, le paramètre de détection absolu est alors l'énergie globale de la *DSP*. La valeur du coefficient de lissage temporel est la même que précédemment. Les résultats sont présentés à la Figure 5-3.

Enfin, on a appliqué notre algorithme de détection à l'enregistrement sonore résultant d'une pré-expérimentation menée au sein du laboratoire. Une personne a suivi un scénario dans lequel elle devait effectuer un repas. La partie de l'enregistrement sonore étudié dure 6 minutes. Le paramètre de détection choisi est l'énergie des coefficients en ondelettes pour une décomposition en ondelettes de niveau 3 avec une ondelette de *Daubechies 3*. Les résultats sont présentés à l'Annexe A. L'enregistrement continu est tracé en gris et les objets sonores capturés en noir.

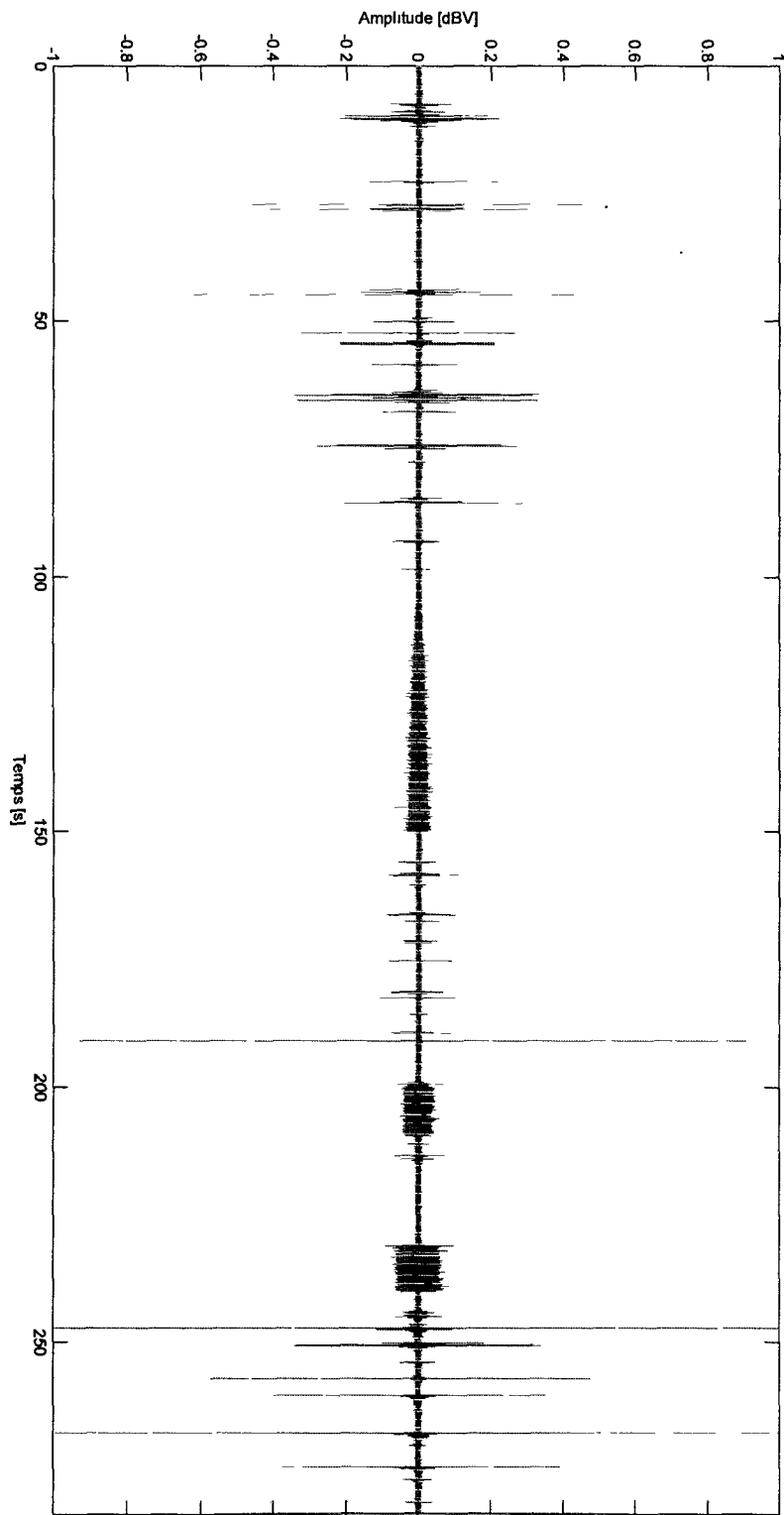


Figure 5-1 : Enregistrement test

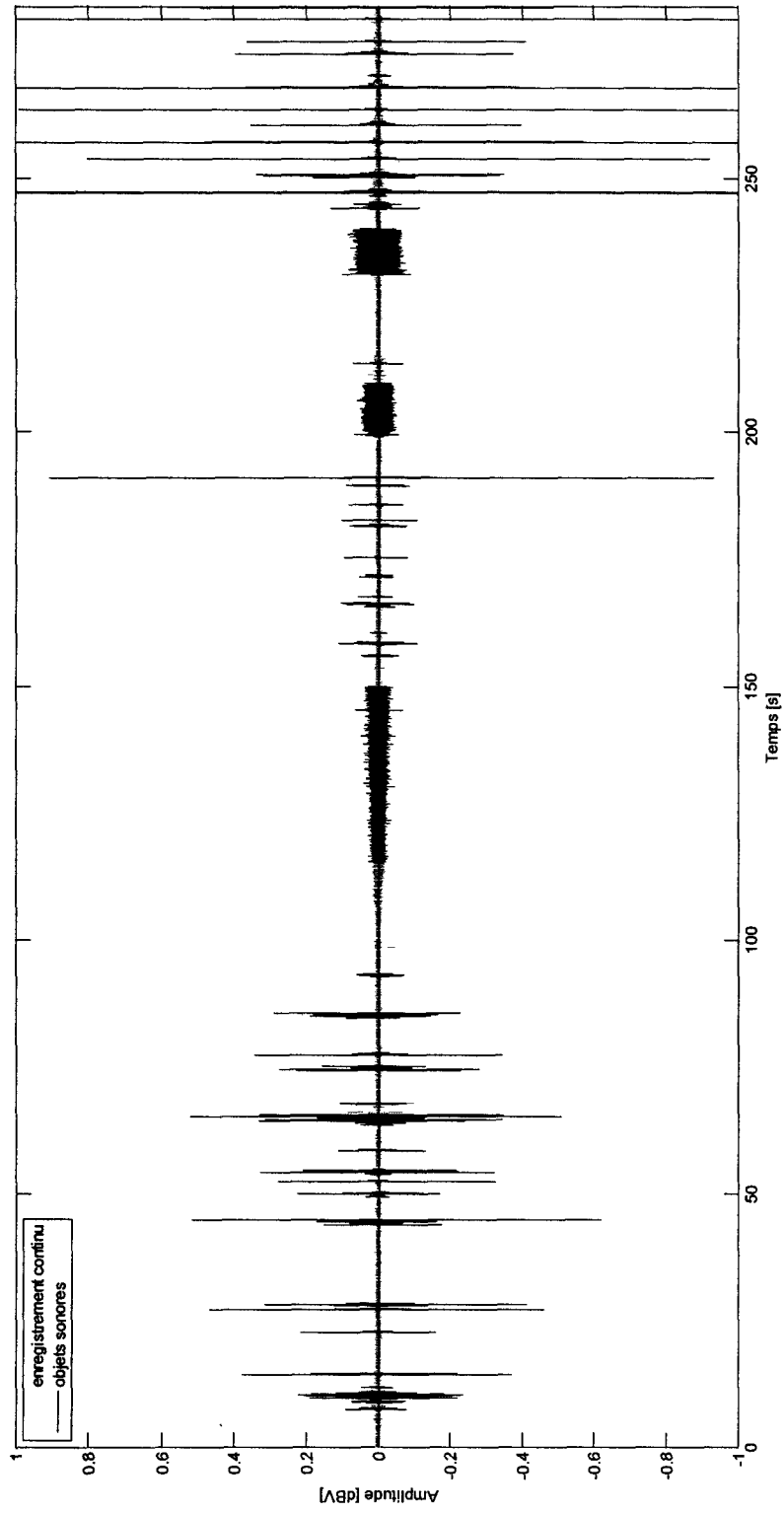


Figure 5-2 : Résultats de la 1<sup>ère</sup> simulation de capture des objets sonores sur l'enregistrement test

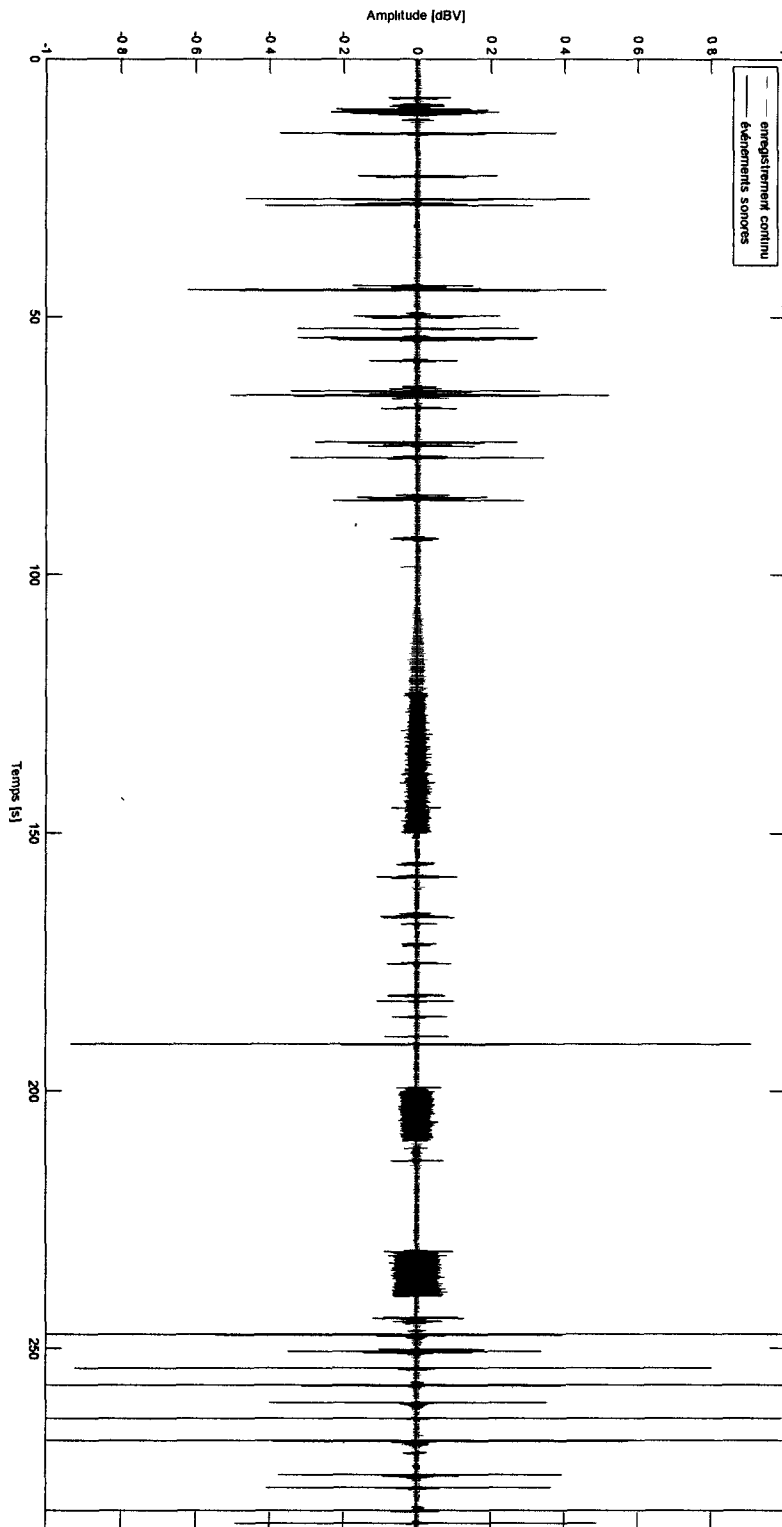


Figure 5-3 : Résultats de la 2<sup>ème</sup> simulation de capture des objets sonores sur l'enregistrement test



Au regard des 2 simulations de la Figure 5-2 et de la Figure 5-3, l'algorithme de détection des événements sonores donne des résultats globalement semblables quelque soit le paramètre de détection utilisé. On remarque que le système de détection des objets sonores permet d'effectuer une capture efficace des objets sonores de différentes natures. D'une part, le système détecte aussi bien les événements sonores de faible que de forte amplitude. Ainsi, les bruits de faible amplitude tels que la manipulation d'ustensile, d'écoulement d'eau et de cuisson sont aussi bien détectés que les bruits de manipulation de casserole, d'assiette et de claquement de porte qui ont une plus forte amplitude. D'autre part, le système de détection s'adapte aux différentes caractéristiques fréquentielles des événements sonores considérés. Ainsi, les bruits de type stationnaire comme la cuisson d'aliment et l'écoulement d'eau sont détectés de la même manière que des bruits fluctuants de manipulation de casserole et d'ustensile tout comme les bruits impulsifs de claquement de porte.

Aux paragraphes 5.2 et 5.3, on discutera plus en détail des résultats des 2 simulations effectuées afin de faire ressortir les avantages et les inconvénients de chacun des paramètres de détection utilisés. Ceci afin de dresser un comparatif entre l'utilisation de la *DSP* en bandes critiques de *Mel* et l'énergie des coefficients en ondelettes dans le contexte de la détection des événements sonores.

## 5.2 Résultats de la 1<sup>ère</sup> simulation

On présente 3 différents types d'objets sonores capturés par le système. On considère alors un son fluctuant de manipulation d'ustensiles, un son stationnaire d'écoulement d'eau du robinet, et un son impulsif de claquement de porte. Puis, on étudie le paramètre absolu de détection et le paramètre de détection des accroches pour chacun de ces sons, ainsi que leurs seuils correspondants. Aux Figures 5-4 (c), 5-5 (c) et 5-6 (c), on prend comme objet d'étude l'énergie des coefficients en ondelettes de niveau 3 à l'aide d'une ondelette de *Daubechies 3*, c'est-à-dire  $d_2$  si l'on se réfère à la Figure 3-14.

La Figure 5-4 représente la capture d'un bruit correspondant à une manipulation d'assiette ainsi que les paramètres et seuils de détection associés.

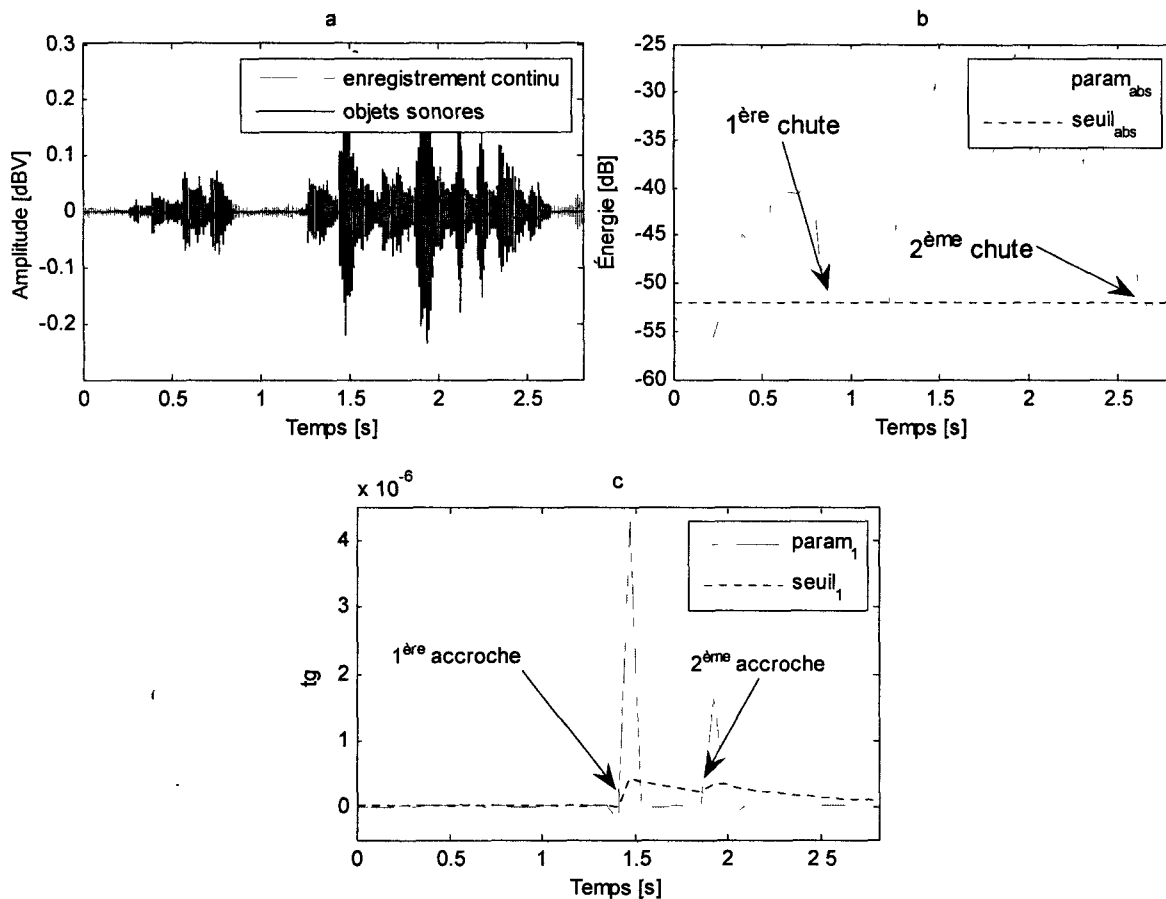


Figure 5-4 : Capture d'une manipulation d'assiette (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$ , (b)  $param_1$ ,  $seuil_1$  (c) pour la 1<sup>ère</sup> simulation

Le système a extrait 2 objets sonores de cette partie de l'enregistrement. Ces sons correspondent à des formes d'onde fluctuantes. On définit un signal fluctuant comme une forme d'onde présentant plusieurs accroches sonores rapprochées, par exemple des assiettes qui s'entrechoquent entre elles. Il est possible de discerner les 2 accroches sur le tracé du paramètre et de la fonction de seuillage de détection des accroches sonores. Celles-ci permettent de commencer l'enregistrement de l'objet sonore puisque les 2 seuils de détection sont dépassés. Les chutes sonores du 1<sup>er</sup> et du 2<sup>ème</sup> objet sonore sont détectées par la fonction de seuillage absolu.

La Figure 5-5 illustre la capture d'un écoulement d'eau du robinet et les paramètres et seuils de détection associés.

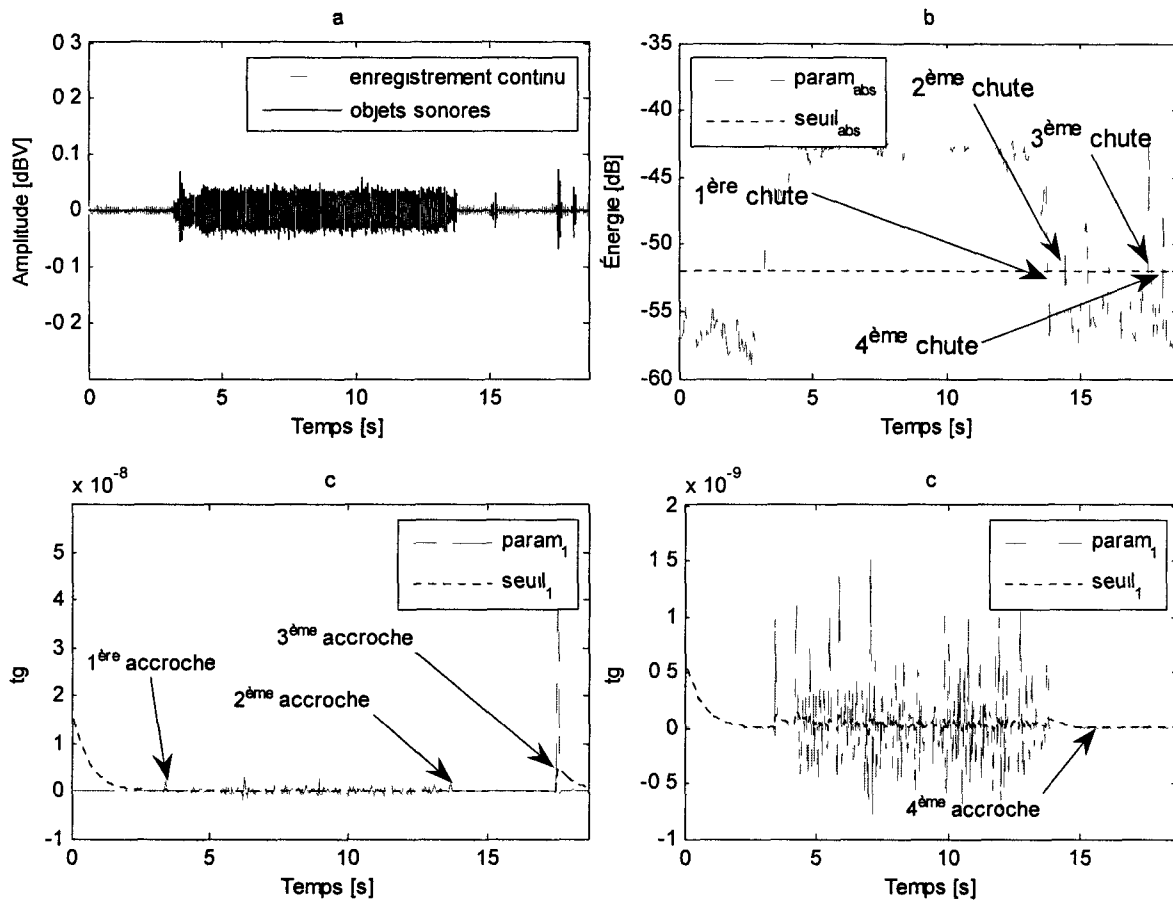


Figure 5-5 : Capture d'un écoulement d'eau (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$ , (b)  $param_1$ ,  $seuil_1$  (c) et (d) pour la 1<sup>ère</sup> simulation

Le système extrait de cet enregistrement 4 objets sonores. Le 1<sup>er</sup> objet correspond à un écoulement d'eau stationnaire, et les 3 objets suivants à l'évacuation du drain de l'évier qui sont des bruits de nature impulsive. La Figure 5-5 (d) présente l'énergie des coefficients en ondelettes de niveau 4 c'est-à-dire  $d_3$  qui permet la détection de l'accroche du 4<sup>ème</sup> objet sonore. Ces bruits ne sont pas révélateurs d'une action de l'habitant, mais ils sont tout de même détectés par notre système. Pour le 1<sup>er</sup> objet, on remarque que l'accroche de début du signal sonore d'écoulement d'eau est détectée de façon nette, tout comme sa chute. Les chutes sonores de ces événements sont reconnues grâce à la fonction de seuillage absolue. On fait remarquer que l'écoulement d'eau possède une multitude d'accroches.

La Figure 5-6 illustre la capture d'un événement correspondant à un claquement de porte et les paramètres et seuils de détection associés.

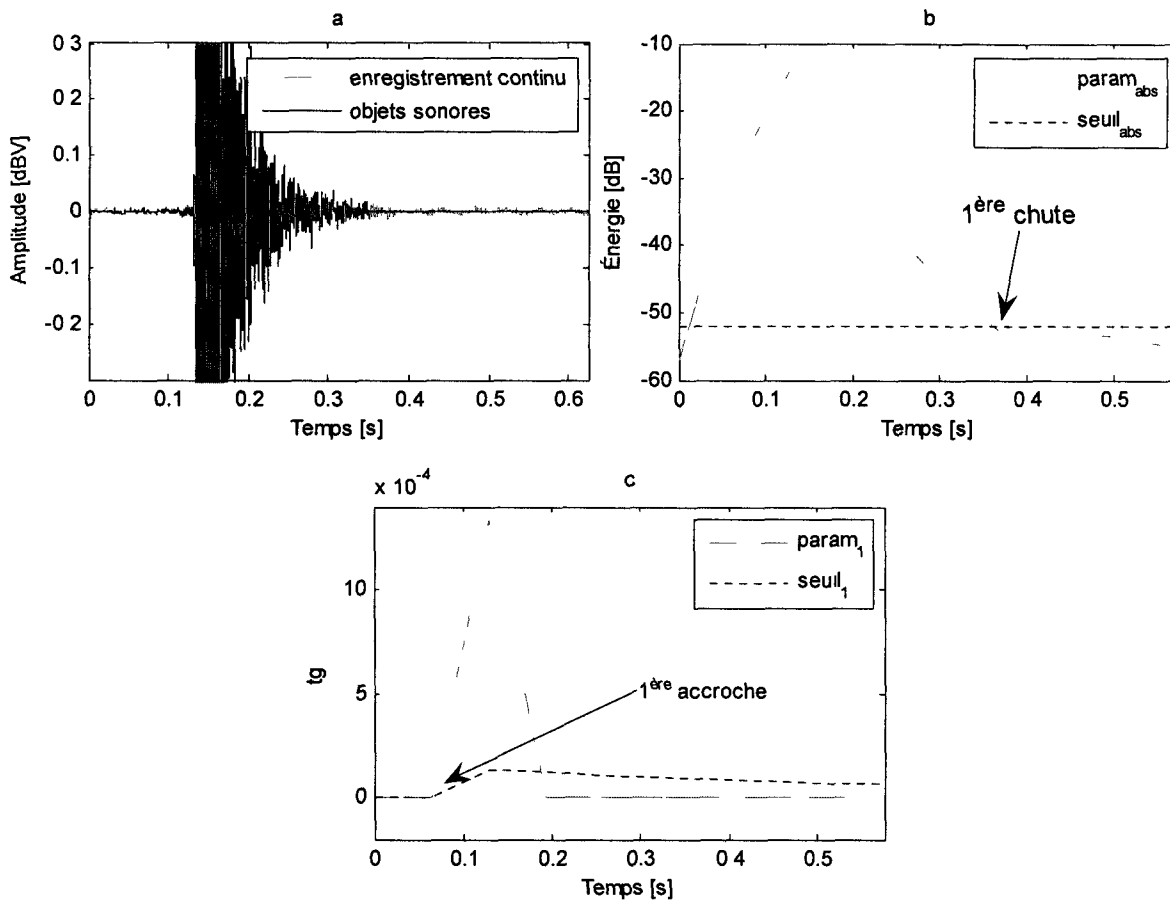


Figure 5-6 : Capture d'un claquement de porte (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$ , (b)  $param_1$ ,  $seuil_1$  (c) pour la 1<sup>ère</sup> simulation

On voit alors que pour un bruit impulsif de courte durée, la détection des accroches et chutes sonore est facilement mise en évidence par les paramètres et seuils de détection. Néanmoins, le début de la capture de l'événement sonore se situe environ *10 ms* avant le début réel du claquement de porte. Ce phénomène n'a pu être visualisé sur les précédents enregistrements, mais c'est un problème qui se pose pour tous les événements sonores détectés. C'est une conséquence de la longueur de trame d'analyse  $N_t$  et de le chevauchement choisis qui conditionnent la précision temporelle de la capture de nos événements. Or, la précision temporelle de notre système de détection est de *16ms*. Dans ce cas, le segment sonore qui déclenche le début de l'enregistrement de l'objet est constitué de *10ms* de silence et de seulement *6ms* de bruit correspondant à l'accroche sonore.

### 5.3 Résultats de la 2<sup>ème</sup> simulation

De la même façon que pour la 1<sup>ère</sup> simulation, on présente les résultats du système de détection pour les différents types d'objets sonores aux Figures 5-7, 5-8 et 5-9. Pour cette 2<sup>nd</sup> simulation, on a utilisé en entrée de notre système de détection la *DSP* en 20 bandes critiques de *Mel*, le paramètre de détection présenté aux Figures 5-7 (c), 5-8 (c) et 5-9 (c) correspond à la 3<sup>ème</sup> bande critique de la *DSP* entre 284 et 426 *Mel*, soit 201 et 322 *Hz*.

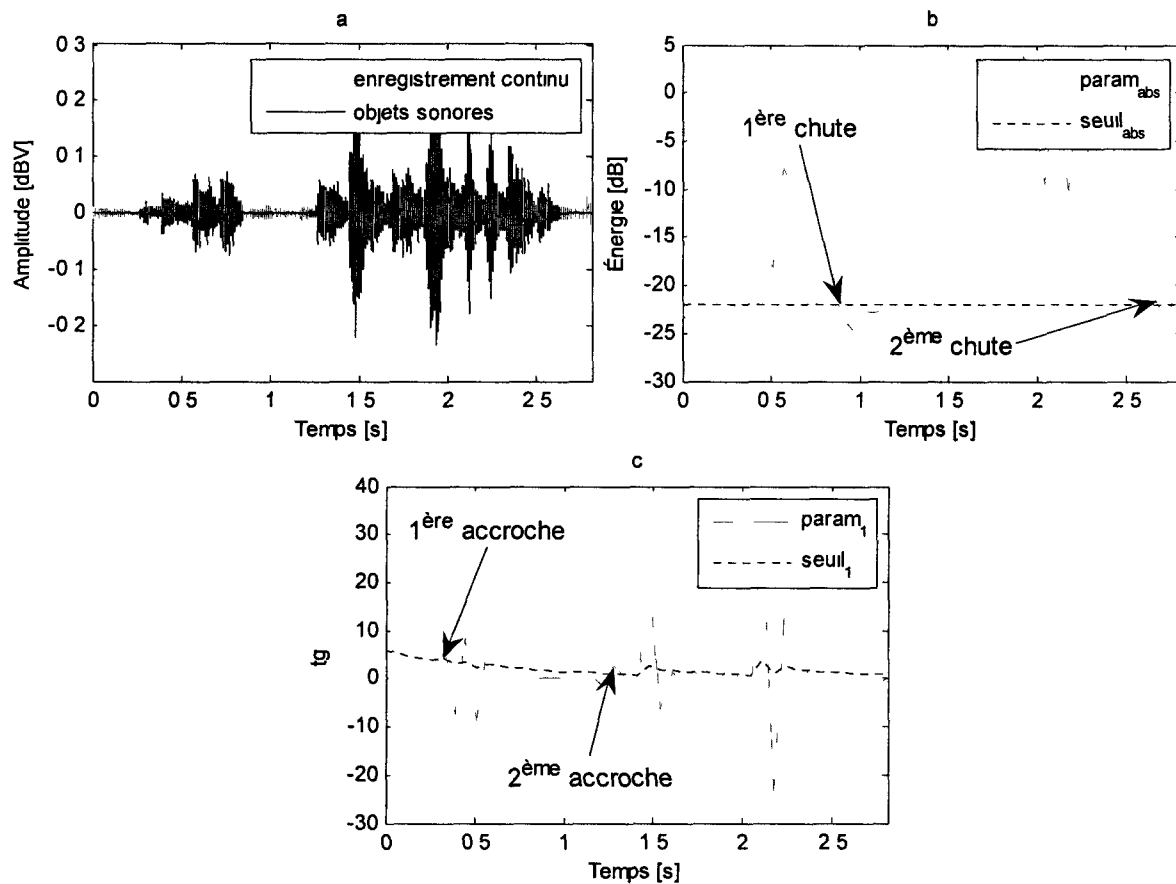


Figure 5-7 : Capture d'une manipulation d'assiette (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$  (b)  $param_r$ ,  $seuil_r$  (c) pour la 2<sup>ème</sup> simulation

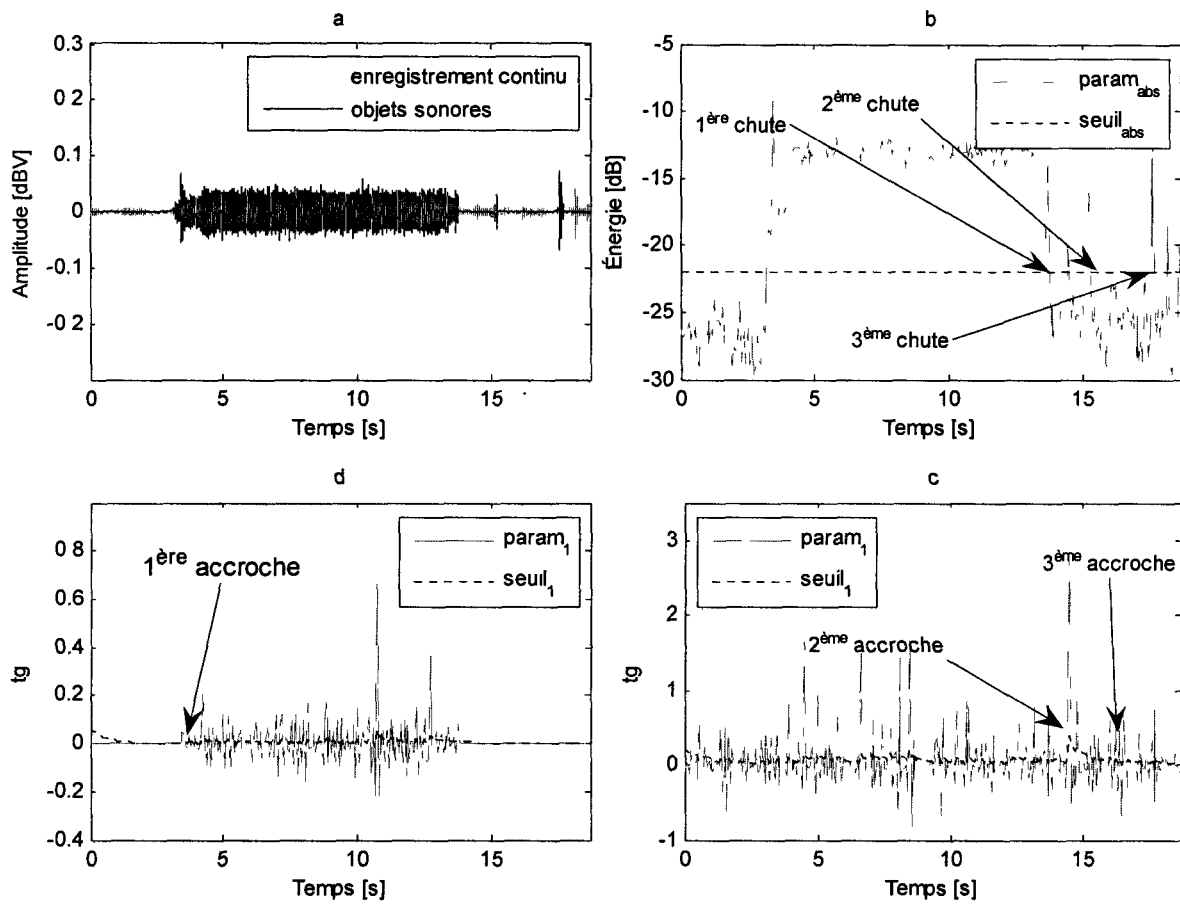
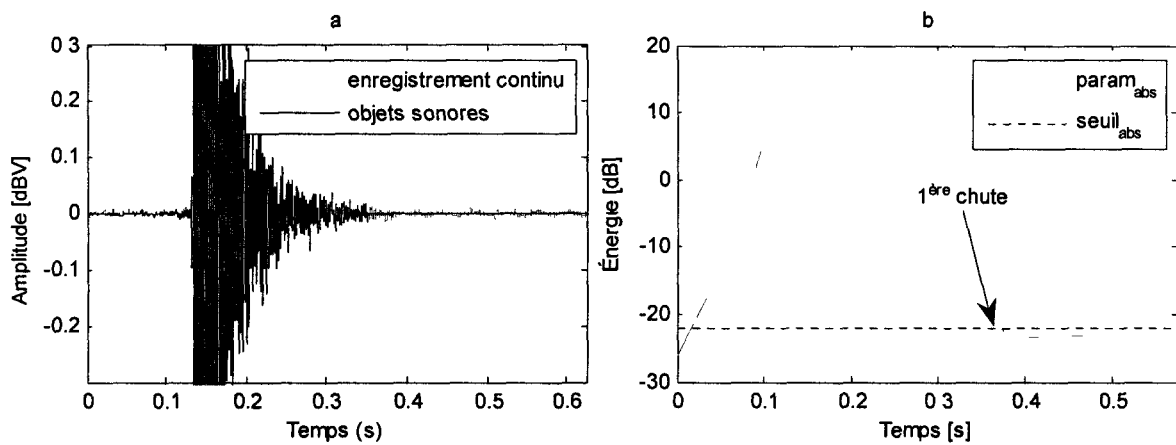


Figure 5-8 : Capture d'un écoulement d'eau (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$ , (b)  $param_1$ ,  $seuil_1$  (c) et (d) pour la 2<sup>ème</sup> simulation



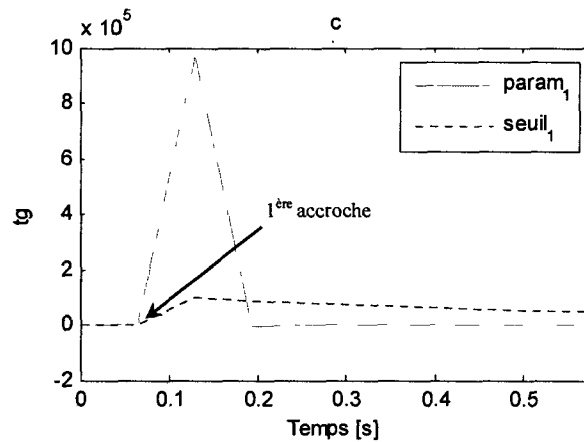


Figure 5-9 : Capture d'un claquement de porte (a) en fonction de  $param_{abs}$ ,  $seuil_{abs}$ , (b)  $param_1$ ,  $seuil_1$  (c) pour la 2<sup>ème</sup> simulation

Au vue des Figures 5-7, 5-8 et 5-9 de la 2<sup>nd</sup> simulation, on remarque que les résultats sont similaires à ceux de la 1<sup>ère</sup> simulation présentée au paragraphe 5.2. Seul l'un des bruits impulsifs d'évacuation d'eau n'a pas été détecté pour la 2<sup>nd</sup> simulation. De plus, la détection de l'accroche du bruit d'écoulement d'eau du robinet n'est pas possible à l'aide de l'information apportée par la 3<sup>ème</sup> bande critique de la DSP à la Figure 5-8 (c). En revanche cette accroche sonore est détectée grâce à l'information contenue dans la 19<sup>ème</sup> bande critique à la Figure 5-8 (d).

## 5.4 Analyse des résultats des simulations

De manière générale, on observe que les paramètres de détection absolue des événements sonores ont des allures et des valeurs très semblables pour les 2 cas d'étude. Les paramètres de détection des accroches sonores ont des formes similaires, en revanche leur plage dynamique est totalement différente pour les 2 simulations présentées. Ceci s'explique par une différence d'implémentation des paramètres de détection pour ces 2 simulations, mais l'allure de ces paramètres reste la même puisqu'elle est le reflet du contenu fréquentiel des objets sonores qui est indépendant des méthodes d'analyse utilisées. De plus, les valeurs de ces paramètres diffèrent selon les types d'événements considérés.

Ainsi, pour la 1<sup>ère</sup> simulation les valeurs des paramètres de détection ont des valeurs situées aux alentours de  $\pm 10^{-9} dB^2$  pour un écoulement d'eau,  $\pm 10^{-6} dB^2$  pour un bruit d'assiette et de  $10^{-3} dB^2$  pour un bruit de porte de placard. On sait que le paramètre  $a_3$  nous renseigne sur le

contenu fréquentiel du signal entre  $F_s/4$  et  $F_s/2$ , soit 2 et 4 kHz. Il est alors normal que le paramètre de détection des accroches sonores soit le plus élevé pour un bruit de claquement de porte, puisque ce signal possède une forte énergie dans les basses fréquences bien que son accroche sonore soit pleine bande. Le signal sonore d'un écoulement d'eau du robinet possède une accroche pleine bande, mais de faible énergie. C'est pour cela que son paramètre de détection des accroches présente la plus faible valeur. Pour ce qui est du signal sonore de manipulation d'assiette, c'est un bruit qui possède aussi une accroche pleine bande, mais dont l'énergie basse fréquence est plus faible que pour un bruit de porte de placard.

Si l'on considère la 2<sup>nd</sup> simulation, les valeurs du paramètre de détection des accroches sont de plus fortes valeurs de manière générale. Elles se situent entre  $\pm 10 \text{ dB}^2$  et  $\pm 50 \text{ dB}^2$  pour un bruit de casserole, entre  $\pm 3 \text{ dB}^2$  pour un écoulement d'eau et de  $10^6 \text{ dB}^2$  pour un bruit de claquement de porte. Ces différences des valeurs du paramètre d'accroche sonore s'expliquent de la même façon que pour la 1<sup>ère</sup> simulation.

Il est aussi évident que la *DSP* en 20 bandes critiques de *Mel* fait ressortir certaines caractéristiques spectrales qui sont bien plus fines que la décomposition en ondelettes de niveau 3. Ainsi, l'énergie des coefficients en ondelettes de niveau 3 correspondant à  $a_3$  a une précision fréquentielle de 1 kHz alors que la 3<sup>ème</sup> bande critique a une précision fréquentielle de 121 Hz. Par exemple, il est impossible de détecter l'accroche sonore d'un écoulement d'eau du robinet à l'aide de l'information apportée par la 3<sup>ème</sup> bande critique de *Mel* de la *DSP*.

De manière plus générale, on observe que le nombre d'accroches sonores présent dans un objet sonore est dépendant de la nature de l'événement sonore. Ainsi, pour un bruit impulsif tel qu'un claquement de porte on détecte une unique accroche, tandis que pour un signal de type fluctuant comme la manipulation d'assiette, on distingue 3 et 7 accroches pour les 2 objets sonores détectés. Enfin, pour un bruit d'écoulement d'eau de type stationnaire, le nombre d'accroches sonores est très élevé, supérieur à la centaine. Il serait aisé de calculer le nombre d'accroches sonores à l'aide du taux de passage par zéro présenté au paragraphe 3.3.2.

Puisque les résultats des simulations présentés sont semblables, il serait alors intéressant d'étendre notre étude à l'évaluation de la complexité des méthodes de détection des



événements sonores afin de permettre une implémentation en temps réel du système de détection des objets sonores.

## 5.5 Résultats du système de reconnaissance des objets sonores

Dans cette partie, on expose les résultats du système de reconnaissance des objets sonores. Les Tableaux 5-1, 5-2, 5-3, 5-4 et 5-5 montrent les résultats des taux obtenus par notre système de reconnaissance des objets sonores suivant le paramètre de classification utilisé. On précise que la reconnaissance des événements sonore est effectuée à partir de la base d'évaluation de notre système, qui est différente de la base d'apprentissage utilisée pour la création des dictionnaires. On précise que les dictionnaires des références ont été créés à partir de 15 objets sonores pour les bruits de claquement de porte, de manipulation d'ustensile, de casserole et d'assiette et de 10 objets sonores pour les bruits d'écoulement d'eau du robinet. De plus, les objets sonores utilisés pour le système de reconnaissance ont tous une taille supérieure ou égale à 0,256 s, soit 7 fenêtres d'étude. En effet, cette hypothèse est obligatoire afin que le ré-échantillonnage des paramètres de classification des objets sonores ne soit pas incohérent, ce qui est le cas pour des événements sonores de plus petite taille.

Tableau 5-1 : Résultats du système de reconnaissance des objets sonores à l'aide de l'énergie des coefficients en ondelettes

		Reconnu comme					
		Assiette	Casserole	Porte	Robinet	Ustensile	Taux de reconnaissance
Action de	Assiette	22/40	7	2	1	8	55,00%
	Casserole	3	11/26	6	6		42,31%
	Porte	6		43/49			87,76%
	Robinet				10/10		100,00%
	Ustensile	2	3		3	35/43	81,40%

Tableau 5-2 : Résultats du système de reconnaissance des objets sonores à l'aide de la DSP en bandes critiques

		Reconnu comme					
		Assiette	Casserole	Porte	Robinet	Ustensile	Taux de reconnaissance
Action de	Assiette	25/40	9	3	1	2	62,50%
	Casserole	5	13/26	4		4	50,00%
	Porte		2	47/49			95,92%
	Robinet				10/10		100,00%
	Ustensile	1	6		3	33/43	76,74%

Tableau 5-3 : Résultats du système de reconnaissance des objets sonores à l'aide du centre de masse de la DSP

		Reconnu comme					
		Assiette	Casserole	Porte	Robinet	Ustensile	Taux de reconnaissance
Action de	Assiette	24/40	13		1	2	60,00%
	Casserole	8	15/26	1	1	1	57,69%
	Porte	4		44/49	1		89,80%
	Robinet				10/10		100,00%
	Ustensile	1	10		6	26/43	60,47%

**Tableau 5-4 : Résultats du système de reconnaissance des objets sonores à l'aide de la largeur de bande de la DSP**

		Reconnu comme					
		Assiette	Casserole	Porte	Robinet	Ustensile	Taux de reconnaissance
Action de	Assiette	25/40	13			2	62,50%
	Casserole	7	16/26	2		1	61,54%
	Porte	7	1	41/49			83,67%
	Robinet				10/10		100,00%
	Ustensile	2	2		2	37/43	86,05%

**Tableau 5-5 : Résultats du système de reconnaissance des objets sonores à l'aide de la fréquence de coupure de la DSP**

		Reconnu comme					
		Assiette	Casserole	Porte	Robinet	Ustensile	Taux de reconnaissance
Action de	Assiette	26/40	11		1	2	65,00%
	Casserole	8	15/26	1		2	57,69%
	Porte	9		40/49			81,63%
	Robinet				10/10		100,00%
	Ustensile		3		6	34/43	79,07%

Les résultats obtenus aux Tableau 5-1 et 5-2 montrent que les objets sonores tels qu'un bruit de claquement de porte et un écoulement d'eau du robinet sont reconnus de façon assez fiable, en se basant sur l'étude de l'énergie des coefficients en ondelettes ou la *DSP* en 20 bandes critiques de *Mel*. Les taux de reconnaissance pour ces objets sonores de types impulsif et stationnaire sont situés entre 90 et 100%. Néanmoins, pour ces mêmes méthodes, il existe une forte confusion entre les bruits de casserole, d'ustensile et d'assiette; leurs taux de reconnaissance sont entre 60 et 85%. La répartition de l'énergie fréquentielle au cours du temps est différente d'un événement sonore à l'autre pour ces types de signaux sonores. En effet, les formes d'onde de ces objets sonores sont assez variables d'un événement à l'autre pour une même classe d'objets sonores. Ces objets sonores sont parfois confondus avec des bruits impulsifs de claquement de porte. En effet, si l'on pose une assiette ou une casserole sur un plan de travail, une table ou une cuisinière; la forme d'onde résultante correspondra à un bruit de type impulsif. De façon générale, le système n'arrive pas à discriminer les objets sonores de type fluctuant entre eux. Les événements sonores fluctuants correspondent aux manipulations d'assiettes, d'ustensiles et de casseroles qui possèdent plusieurs accroches sonores dépendamment de comment on les manipule. On imagine alors que si l'on ajoute d'autres événements sonores de types stationnaires et impulsifs, ceux-ci seraient alors probablement confondus avec les objets sonores tels que des claquements de porte et des écoulements d'eau.

Aux Tableaux 5-1, 5-2, 5-3, 5-4 et 5-5, on observe que le taux de reconnaissance obtenu pour l'objet sonore stationnaire d'écoulement d'eau est toujours égal à 100%. Pour ce qui est d'un bruit de claquement de porte, le taux de reconnaissance est entre 80 et 90%. Pour ce qui est des bruits de type fluctuants, les résultats restent semblables à ceux des Tableau 5-1 et Tableau 5-2 pour les mêmes raisons énoncées précédemment. Il est à noter que les bruits d'ustensile ont un taux de reconnaissance légèrement supérieur qui est dû aux caractéristiques hautes fréquences de ces objets sonores.

## CHAPITRE 6. CONCLUSION

L'étude menée portant sur la détection des événements sonores s'est limitée à des signaux sonores correspondant à des actions simples réalisées dans la cuisine telles qu'une manipulation d'assiette, d'ustensile, de casserole, un écoulement d'eau ou encore la cuisson d'aliments dans un environnement non bruité. Ces sons ont été sélectionnés car ils sont représentatifs de la majorité des actions réalisées par une personne au sein de la cuisine. En effet, une *AVQ* comme faire un repas ou du café nécessite de récupérer du café, des casseroles et des ustensiles dans des placards et de faire cuire des aliments ou bouillir de l'eau. L'ensemble de ces événements sonores engendrés par ces actions simples ont été considéré dans notre étude. Cependant, il serait possible, par extension, d'appliquer notre système de détection des événements sonores à tout type de son ayant lieu dans n'importe quelle pièce de l'habitat.

En effet, la preuve de concept de la faisabilité d'un système de détection des événements sonores dans un environnement non bruité à l'aide d'un unique microphone par pièce a été faite. Le système de détection des objets sonores se basant sur l'énergie des coefficients en ondelettes ou la *DSP* en bandes critiques de *Mel* permet une capture des objets sonores avec une bonne précision temporelle. Pour juger des limites du système de détection des objets sonores, il serait alors souhaitable d'identifier les omissions dans la détection des événements sonores, ou encore de discerner les vraies détections des fausses détections, c'est-à-dire si l'objet sonore correspond à une action réelle de la personne [7]. De plus, l'algorithme de détection des objets sonores nécessite seulement le calcul de 2 fonctions de seuillage. Une fonction de seuillage des accroches sonores et une fonction de seuillage absolu pour les chutes sonores. Il serait également pertinent d'étudier la faisabilité de l'implémentation d'un tel système en temps réel. À partir de là, un déploiement d'un réseau de microphone à l'intérieur de l'habitat pourrait permettre de tester le système de détection des événements sonores dans un environnement sonore réel et dans un appartement au complet. Néanmoins, l'algorithme de détection des événements sonores réagit de façon correcte vis-à-vis d'un environnement sonore réel comme le montre les résultats présentés à l'Annexe A.

Le système de reconnaissance des événements sonores utilise la distribution des énergies fréquentielles des événements sonores au cours du temps : l'énergie des coefficients en ondelettes, la *DSP* en bandes critiques et le centre de masse, la largeur de bande ou la fréquence de coupure spectrale. À noter que le système de reconnaissance des objets sonores se base et est dépendant de la détection de ces mêmes objets sonores. Ce système a prouvé son efficacité pour la reconnaissance des événements sonores répétitifs tels que les bruits de claquement de porte ou d'écoulement d'eau. Ces événements sonores sont détectés et reconnus de façon fiable par notre système car les formes d'onde de ces sons varient peu d'une fois à l'autre. La création de modèles pour l'étude de ces 2 types de signaux, stationnaire et impulsif, peut paraître être une bonne méthode. Par contre, les résultats obtenus par cette méthode ne permettent pas la discrimination des objets sonores de type fluctuant tels que la manipulation d'assiette, de casserole et d'ustensile. Par exemple, sortir une casserole d'un meuble en l'entrechoquant avec les rangements ou bien avec d'autres casseroles, prendre une seule fourchette ou plusieurs fourchettes dans le tiroir, saisir une assiette d'une pile, ou bien poser une assiette sur une table, sont autant de scénarios possibles et différents que les formes d'ondes qu'ils vont générer. On ne peut donc pas créer autant de dictionnaires de références que de scénarios possibles. Il serait alors intéressant d'adapter notre classification des objets sonores non pas sur l'objet qui a généré un bruit, mais sur le type de bruit qui est généré. Par exemple décider si un bruit est de type impulsif, fluctuant ou stationnaire. À partir de là, une approche hiérarchique visant à avoir plus de détail sur la nature de l'action réalisée par l'habitant pourrait donner de meilleurs résultats [3, 26]. De plus, cette approche pourrait prendre en compte la séparation des sources sonores. Ainsi, si une personne fait la cuisine en faisant fonctionner la hotte, il serait utile de séparer les événements sonores de différents types [20, 34]. En effet, lors de la préparation d'un repas, l'habitant peut effectuer plusieurs actions qui généreront des bruits qui s'ajouteront à celui du fonctionnement de la hotte.

Afin d'améliorer aussi bien la détection que la reconnaissance des événements sonores, il serait judicieux d'investiguer plus en profondeur les méthodes de caractérisation du son. Par exemple, les paramètres psycho-acoustiques qui sont des grandeurs physiques du son qui représente la sensation auditive que provoque un son [15, 16, 35]. Les informations utilisées par le système de détection et de reconnaissance des sons sont des images spectrales. De la même façon, l'utilisation de techniques de traitement de l'image se basant sur la perception

visuelle des images pourrait permettre d'obtenir de meilleurs résultats pour ce qui est de la détection et de la reconnaissance des objets sonores [5, 10, 31].

Si l'on se replace dans le contexte de l'identification des *AVQ*, le système réalisé répond en partie à la problématique posée. Certes, à l'aide des objets sonores extraits, il est seulement possible de reconnaître certaines actions réalisées par une personne au sein d'un habitat. En revanche, on est capable, grâce au système de détection des objets sonores de savoir combien d'événements sonores ont été provoqués par l'habitant, et de quelle pièce ils proviennent. Il serait donc envisageable de déterminer un taux d'activité de la personne dans chacune des pièces de l'habitat. En corrélant ces informations avec les périodes de la journée, on pourrait déterminer un certain nombre d'*AVQ* réalisé par l'habitant. Par exemple, un fort taux d'activité de la personne dans la cuisine sur l'heure du midi serait fortement représentatif de l'*AVQ* correspondant à la réalisation d'un repas. De plus, il ne faut pas oublier que l'identification d'*AVQ* au sein du laboratoire *DOMUS* ne se concentre pas uniquement sur les informations sonores provenant des microphones, mais sur les informations apportées par l'ensemble des capteurs dont est équipé l'appartement-témoin. Par conséquent, des informations provenant des autres capteurs seront fusionnées avec celle de notre système afin de renforcer l'identification des *AVQ*. De cette façon, si le taux d'activité de la personne révèle une situation de risque ou de crise pour l'habitant, il sera possible de prévenir un intervenant extérieur [4, 15, 27].

Pour finir, la présente étude témoigne de la nécessité de l'analyse continue des données sonores dans un contexte de surveillance à domicile. Plus particulièrement, le système conçu fait la preuve de concept de l'efficacité de l'utilisation des signaux sonores dans le but d'identifier les *AVQ* réalisées par l'habitant. La contribution principale du document réside dans la réalisation d'un système de détection des événements sonores qui nous renseigne à une échelle macroscopique des *AVQ* réalisées par la personne au sein de l'habitat.





## ANNEXE A – RÉSULTATS D'EXPÉRIMENTATION

Les résultats de l'expérimentation présentés à la Figure A-1 démontrent que le système de détection des événements sonores est efficace, cependant certains événements de faibles amplitudes ne sont pas détectés. Ceci est compréhensible puisque l'enregistrement a été effectué à l'aide du microphone dans la cuisine et que les événements qui se sont produits dans la salle à manger, par exemple, n'ont pu être détectés. Si l'on avait effectué un enregistrement à l'aide d'un réseau de microphones distribués dans l'appartement tel que décrit à la Figure 1-1, ce problème serait aisément résolu. Ainsi, les microphones de chacune des pièces reconnaîtraient les événements sonores qui ont lieu dans leurs pièces respectives. On remarque aussi que le système n'est pas capable de détecter des événements sonores simultanés. On voit alors qu'à la 50<sup>ème</sup> seconde de l'enregistrement sonore, des événements sonores impulsifs et stationnaires se chevauchent. L'événement sonore impulsif est alors parfaitement détecté alors que la détection de l'accroche sonore de l'événement stationnaire est retardée de 1,5 s.

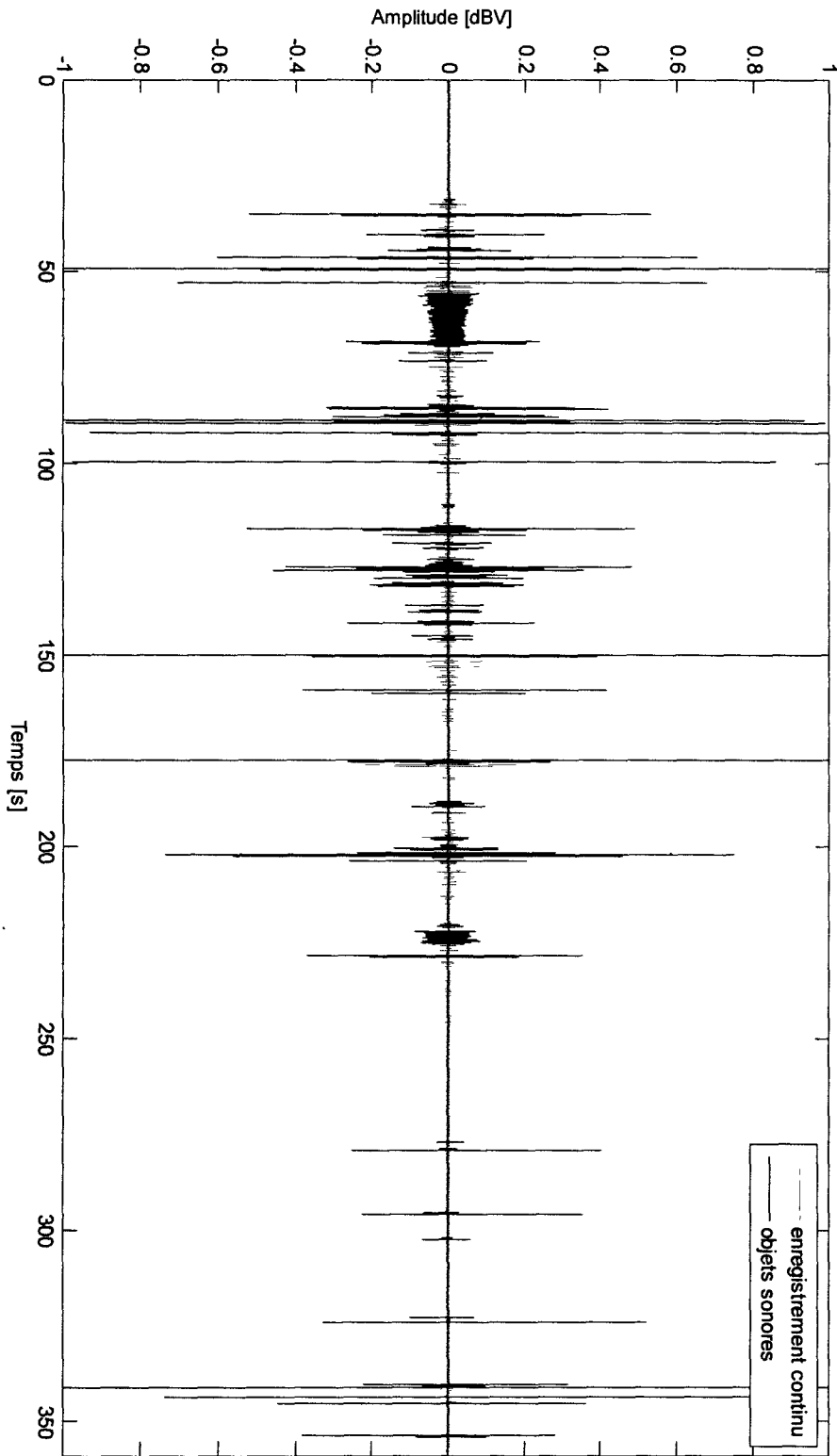


Figure A-1 : Résultats de la capture des objets sonores sur l'enregistrement expérimental

## LISTE DES RÉFÉRENCES

- [1] Aribé, O., Meana, H. & Miyatake, M., *Environmental sounds recognition system using the speech, recognition system techniques*, Electrical and Electronics Engineering, 2nd International Conference on, 2005, pp. 13-16.
- [2] Ashiya, T., H.M. & Nakagawa, M., *IOSES: an indoor observation system based on environmental sounds recognition using a neural network*, Transactions of the Institute of Electrical Engineers of Japan, Part C, 1996, pp. 341-349.
- [3] Atrey, P., Maddage, N. & Kankanhalli, M., *Audio Based Event Detection for Multimedia Surveillance*, Acoustics, Speech and Signal Processing, ICASSP Proceedings, IEEE International Conference on, 2006, Vol. 5, pp. 813-816.
- [4] M. Azlan, I. Cartwright, *Multi-modal monitoring of the aged in their own homes*, Proc. ICOST, 3rd. International Conference on Smart Homes and Health Telematics, 2005, pp. 1-8.
- [5] Chendeb, M., Khalil, M. & Duchêne, J., *Methodology of wavelet packet selection for event detection*, Signal Process., Elsevier North-Holland, Incorporation, 2006, Vol. 86, pp. 3826-3841.
- [6] Clarkson, B. & Pentland, A., *Extracting context from environmental audio*, Wearable Computers, Digest of Papers, Second International Symposium on, 1998, pp. 154-155.
- [7] Clavel, C., Ehrette, T. & Richard, G., *Events Detection for an Audio-Based Surveillance System*, Multimedia and Expo, ICME, IEEE International Conference on, 2005, pp. 1306-1309.
- [8] Cowling, M. & Sitte, R., *Comparison of techniques for environmental sound recognition*, Pattern Recognition , Elsevier Science Incorporation, 2003, Vol. 24, pp. 2895-2907.
- [9] Cristani, M., Bicego, M. & Murino, V., *On-line adaptive background modelling for audio surveillance*, Pattern Recognition, ICPR, Proceedings of the 17th International Conference on, 2004, Vol. 2, pp. 399-402.
- [10] De Silva, G., Anavi, S., Yamasaki, T. & Aizawa, K., *A Sensor Network for Event Retrieval in a Home Like Ubiquitous Environment*, Acoustics, Speech and Signal Processing, ICASSP, IEEE International Conference on, 2007, Vol. 4, pp.1393-1396.
- [11] Gersho, A., Gray, R.M., 1992, « Vector quantization and signal compression », Kluwer Academic Press., Norwell, MA, USA.

- [12] Giannakopoulos, T., Pikrakis, A. & Theodoridis, S., *A novel efficient approach for audio segmentation*, Pattern Recognition, ICPR, 19th International Conference on, 2008, pp. 1-4.
- [13] Harma, A., McKinney, M. & Skowronek, J., *Automatic surveillance of the acoustic activity in our living environment*, Multimedia and Expo. ICME, IEEE International Conference on, 2005, pp. 4-9.
- [14] Hattori, Y., Ishihara, K., Komatani, K., Ogata, T. & Okuno, H., *Repeat recognition for environmental sounds*, Robot and Human Interactive Communication, ROMAN, 13th IEEE International Workshop on, 2004, pp. 83-88.
- [15] Head acoustics, 2010, [http://www.head-acoustics.de/downloads/fr/Application\\_Notes/ArtemiS/Analyses\\_Psychoacoustiques\\_I\\_04\\_10f.pdf](http://www.head-acoustics.de/downloads/fr/Application_Notes/ArtemiS/Analyses_Psychoacoustiques_I_04_10f.pdf), page consulté le 1<sup>er</sup> septembre 2010.
- [16] Head acoustics, 2008, [http://www.head-acoustics.de/downloads/fr/Application\\_Notes/Psychoakustische\\_Analysen\\_II\\_12\\_08f.pdf](http://www.head-acoustics.de/downloads/fr/Application_Notes/Psychoakustische_Analysen_II_12_08f.pdf), page consulté le 1<sup>er</sup> septembre 2010.
- [17] Hitti, E., Lucas, M.F., *Wavelet-packet basis selection for abrupt changes detection in multicomponent signals*, Proceedings of EUSIPCO, 1998, pp. 1841-1844.
- [18] Institut nationale de la santé et de recherche médicale, 2005, <http://infodoc.inserm.fr/serveur/vieil.nsf/397fe8563d75f39bc12563f60028ec43/91f6979fd28b9d6a80256b430057aa7f?OpenDocument>.
- [19] Istrate, D. & Castelli, E., *Multichannel Sound Acquisition with Stress Situations Determination for Medical Supervision in a Smart House*, Proceedings of the 4th International Conference on Text, Speech and Dialogue, Springer-Verlag, 2001, pp. 266-272.
- [20] Janku, L., *Artificial Perception: Auditory Decomposition of Mixtures of Environmental Sounds - Combining Information Theoretical and Supervised Pattern Recognition Approaches*, SOFSEM, 2004, pp. 229-240.
- [21] Jhing-Fa W., Jia-Ching W., Tze-Hsuan H., Cheng-Shu H., *Home environmental sound recognition based on MPEG-7 features*, Micro-NanoMechatronics and Human Science, IEEE International Symposium on, 2003, Vol. 2, pp. 682-685.
- [22] Jia-Ching W., Hsiao-Ping L., Jhing-Fa W., Cai-Bei L., *Robuust Environmental Sound Recognition for Home Automation*, Automation Science and Engineering, IEEE Transactions on, 2008, Vol. 5(1), pp. 25-31.
- [23] Kejun M., Rovnyak, S. M., Chee-Mun O., *Dynamic event detection using wavelet analysis*, Power Engineering Society General Meeting, IEEE, 2006, pp. 1-7.

- [24] Laurent, H., Hitti, E. & Lucas, M.-F., *Abrupt changes detection in the time-scale and in the time-frequency planes: a comparative study*, Time-Frequency and Time-Scale Analysis, Proceedings of the IEEE-SP International Symposium on, 1998, pp. 501-504.
- [25] Louradour, J., 2003, *Recherche de slogans et de logos dans les clips publicitaires*, Rapport de stage, DEA IIL, Université Paul Sabatier.
- [26] Ma, L., Milner, B. & Smith, D., *Acoustic environment classification*, ACM Trans. Speech Lang. Process., ACM, 2006, Vol. 3, pp. 1-22.
- [27] Moncrieff, S., Venkatesh, S., West, G. & Greenhill, S., *Multi-modal emotive computing in a smart house environment*, Pervasive Mob. Comput., Elsevier Science Publishers B. V., 2007, Vol. 3, pp. 74-94.
- [28] Mallat, S., 2000, « Une exploration des signaux en ondelettes », Édition de l'école polytechnique/Ellipse, France.
- [29] Radhakrishnan, R., Divakaran, A. & Smaragdis, A., *Audio analysis for surveillance applications*, Applications of Signal Processing to Audio and Acoustics, IEEE Workshop on, 2005, pp. 158-161.
- [30] Shi-Huang, C., & Wang, J.-F., *Speech Enhancement Using Perceptual Wavelet Packet Decomposition and Teager Energy Operator*, J. VLSI Signal Process. Syst., Kluwer Academic Publishers, 2004, Vol. 36, pp. 125-139.
- [33] Shu-Ching C., Chen, M., Zhang, C. & Shyu, M.-L., *Exciting Event Detection Using Multi-level Multimodal Descriptors and Data Classification*, Proceedings of the Eighth IEEE International Symposium on Multimedia, IEEE Computer Society, 2006, pp. 193-200.
- [32] Sitte, R. & Willets, L., *Non-speech environmental sound identification for surveillance using self-organizing-maps*, Proceedings of the Fourth conference on IASTED International Conference: Signal Processing, Pattern Recognition, and Applications, ACTA Press, 2007, pp. 281-286.
- [33] Smeaton, A.F. & McHugh, M., *Towards event detection in an audio-based sensor network*, Proceedings of the third ACM international workshop on Video surveillance & sensor networks, ACM, 2005, pp. 87-94.
- [34] Wen, X., Shi, Y.-Y. & She, B., *Separation of impulsive acoustical events*, Acoustics, Speech and Signal Processing, Proceedings, ICASSP, IEEE International Conference on, 2004, Vol. 2, pp. 733-736.
- [35] Yasukawa, H., Itai, A. & Takasuka, T., *Sound Scene Analysis with Psycho-acoustics Parameter*, Communications and Information Technologies, ISCIT, International Symposium on, 2006, pp. 527-530.

- [36] Zhan, Y., Miura, S., Nishimura, J. & Kuroda, T., *Human Activity Recognition from Environmental Background Sounds for Wireless Sensor Networks*, Networking, Sensing and Control, IEEE International Conference on, 2007, pp. 307-312.