



**HAL**  
open science

## Reconnaissance de sons d'eau pour l'indexation en activités de la vie quotidienne

Patrice Guyot, Julien Pinquier, Régine André-Obrecht

► **To cite this version:**

Patrice Guyot, Julien Pinquier, Régine André-Obrecht. Reconnaissance de sons d'eau pour l'indexation en activités de la vie quotidienne. Workshop - Alzheimer, Approche pluridisciplinaire - De la recherche clinique aux avancées technologiques - 2013, Jan 2013, Toulouse, France. pp. 147-161. hal-01142073

**HAL Id: hal-01142073**

**<https://hal.science/hal-01142073>**

Submitted on 14 Apr 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive TOULOUSE Archive Ouverte (OATAO)

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : <http://oatao.univ-toulouse.fr/>  
Eprints ID : 12795

**To cite this version** : Guyot, Patrice and Pinquier, Julien and André-Obrecht, Régine *[Reconnaissance de sons d'eau pour l'indexation en activités de la vie quotidienne](#)*. (2013) In: Workshop - Alzheimer, Approche pluridisciplinaire - De la recherche clinique aux avancées technologiques - 2013, 25 January 2013 - 25 January 2013 (Toulouse, France).

Any correspondance concerning this service should be sent to the repository administrator: [staff-oatao@listes-diff.inp-toulouse.fr](mailto:staff-oatao@listes-diff.inp-toulouse.fr)

# Reconnaissance de sons d'eau pour l'indexation en activités de la vie quotidienne

► Patrice GUYOT (IRIT, Toulouse), Julien PINQUIER (IRIT, Toulouse), Régine ANDRE-OBRECHT (IRIT, Toulouse).

■ **RÉSUMÉ** • L'évaluation de troubles dans la réalisation des activités quotidiennes est aujourd'hui utilisée dans le diagnostic des démences, mais se heurte à un manque d'outils objectifs. Pour pallier ce manque, le projet IMMED propose la réalisation de vidéo au domicile du patient et l'indexation automatique de ces vidéos en activités. Ces vidéos indexées permettent aux spécialistes de visualiser les patients effectuer des activités dans leur environnement habituel. Dans ce contexte, de nombreuses tâches quotidiennes ont un rapport avec l'eau : se laver les mains, faire la vaisselle, se brosser les dents, etc. Dans cet article, nous présentons deux méthodes de détection de sons d'eau pour la segmentation automatique en activité. La première méthode, basée sur des descripteurs acoustiques, permet la détection du flot d'eau. Pour reconnaître les autres types de sons d'eau, comme les gouttes, nous présentons également une approche originale qui s'appuie sur des modèles acoustiques des sons de liquides.

■ **MOTS-CLÉS** • Démence, Alzheimer, caméra, indexation, segmentation, activités de la vie quotidienne, acoustique, scènes sonores, eau, goutte, liquide

■ **ABSTRACT** • *With ageing of the population in the world, patients with Alzheimer's disease consequently increase. People suffering from this pathology show early modifications in their "activities of daily living". Those abilities modifications are part of the dementia diagnosis, but are often not reported by the patients or their families. The IMMED project investigates tele-monitoring technologies to support doctors in the diagnostic and follow-up of dementia illnesses such as Alzheimer. Specifically, water sounds are very useful to track and identify abnormal behaviors from everyday activities (e.g. hygiene, household, cooking, etc.). In this work, we propose two approaches to detect this type of sound events. The first one is based on acoustics descriptors and aim to recognize water flow events. The second one is based on a physical model of air bubble acoustics. It is able to recognize a wide variety of water sounds.*

■ **KEYWORDS** • *Dementia, Alzheimer, camera, indexing, segmentation, activities of daily living, acoustic, computational auditory scene analysis, water, drop, liquid*

## 1. Indexation en activités pour l'aide au diagnostic

### 1.1. Contexte général

Avec l'augmentation de l'espérance de vie, les cas de démences sont de plus en plus fréquents. Actuellement, le processus de diagnostic de ces démences ne permet pas d'identifier à temps toutes les personnes souffrantes dans la population. De nouveaux protocoles et outils de diagnostic semblent ainsi nécessaires pour permettre de sécuriser les conditions de vie de ces personnes âgées et d'éviter une dégradation de leur état de santé.

Les démences telles que la maladie d'Alzheimer sont habituellement diagnostiquées par l'intermédiaire d'indices sur des changements pathologiques dans la vie de tous les jours des patients. Pour évaluer ces changements, les médecins utilisent des examens neurologiques, des techniques d'image cérébrale, et principalement des tests neuropsychologiques.

Les tests neuropsychologiques sont basés sur des questionnaires sur support papier et des discussions qui ont pour but d'identifier des difficultés à réaliser certaines activités de la vie quotidienne. La déficience dans la réalisation de ces activités peut, en effet, être reliée à un déclin cognitif causé par les démences. Elle engendre également une perte d'autonomie du patient. Ainsi, le diagnostic de démence est déclaré à partir du résultat de ces tests et du constat de perte d'autonomie du patient.

Évaluer une perte d'autonomie fait donc partie du diagnostic. Toutefois, cette évaluation dépend d'outils subjectifs et de compétences à analyser clairement la situation ; elle se heurte aux dénis du patient et de ses proches, ou à l'anosognosie. Concrètement, les équipes médicales estiment le plus souvent des aptitudes énoncées par le patient lui-même ou par ses proches (Gaëstel et al, 2011).

### 1.2. Observation des activités quotidiennes à domicile

Le projet IMMED a été développé pour élaborer de nouvelles méthodes pour mesurer objectivement les effets de la perte d'autonomie chez les patients (Mégret et al, 2010). L'objectif est d'estimer de manière la plus complète les aptitudes du patient à exécuter dans son domicile des activités de la vie quotidienne, et d'évaluer les éventuels erreurs des patients au sein d'un environnement réel et habituel.

Dans le cadre de ce projet, nous avons développé un système de capture vidéo permet de filmer les activités instrumentales du patient. Ce dispositif, illustré à la figure 1, a été réalisé à partir d'une caméra portative fixée sur un gilet porté par le patient. Les tests réalisés sur des volontaires âgés sains ainsi que sur des patients souffrants de troubles légers n'ont pas mis en évidence de gêne des personnes vis-à-vis du dispositif. Lors de la réalisation effective des vidéos, un psychologue vient au

domicile du patient pour installer le dispositif et proposer au patient un ensemble d'activités d'intérêt à effectuer.



**Figure 7 • Dispositif d'acquisition audio-vidéo**

### **1.3. Indexation audio vidéo**

L'acquisition vidéo au domicile du patient fournit de longues séquences. Ces séquences peuvent de quelques dizaines de minutes à plus d'une heure, ce qui est trop long pour être visualisé entièrement par un spécialiste. De plus les scènes filmées n'ont d'intérêt que si l'autonomie du patient peut être évaluée. La navigation dans la vidéo est donc un point critique pour la viabilité de cette approche. Au sein du projet IMMED, une interface de navigation a été réalisée par l'intermédiaire d'un ensemble d'activités d'intérêt, défini préalablement par les médecins et indexées automatiquement. Cette interface permet au spécialiste de visualiser directement les scènes où l'autonomie du patient peut être évaluée, ou de regarder la vidéo de manière séquentielle (Karaman, 2011).

L'analyse automatique des vidéos est ainsi une étape préalable importante pour construire l'indexation nécessaire à la consultation des données par un médecin expert. Un des objectifs du projet est donc d'extraire des informations des flux audio et vidéo pour segmenter chaque vidéo en activité.

Concrètement, des informations audio et vidéo de bas niveau (bruit d'intérêts liés à des activités, mouvement de la caméra, ou encore localisation du patient dans la maison) sont fusionnées pour segmenter la vidéo en activité (Pinquier et al, 2012). Si ce type de segmentation automatique obtient de nos jours des résultats satisfaisants pour des types de données fortement structurées et de bonne qualité, comme le journal télévisé ou certaines émissions de radio, l'indexation en activités pose de nombreux problèmes sur les enregistrements issus de la caméra portée dans différents lieux.

## 1.4. Description de la problématique

Au sein du projet IMMED, nous avons principalement travaillé sur la détection automatique de bruit d'intérêts permettant d'inférer sur l'activité effectuée par le patient. La principale difficulté est que ces bruits d'intérêts sont le plus souvent recouverts par d'autres sons. Par exemple, le son dû à l'utilisation d'un robinet va être recouvert par les bruits de chocs de vaisselle lors de l'activité *faire la vaisselle*. A ces sons peuvent s'ajouter d'autres sons environnementaux indépendants de l'activité du patient, qu'ils soient intérieurs, par exemple un bruit de néon, ou extérieurs au domicile comme les bruits de circulation (Mesaros et al, 2010). De plus, le patient étant en discussion permanente avec le psychologue qui le guide sur les activités à réaliser, la parole est extrêmement présente dans le corpus IMMED. La voix du patient recouvre donc souvent les sons environnementaux que nous aimerions détecter. Enfin, à la différence de la plupart des projets comparables de l'état de l'art, par exemple (Chen et al, 2005), les enregistrements du projet IMMED proviennent de lieux différents, et sont effectués par l'intermédiaire d'une caméra portée. Ainsi, une même activité, par exemple *faire le café*, peut provoquer des sons très différents selon les objets et l'acoustique des lieux impliqués, mais aussi selon la position du patient et du microphone de la caméra.

Au sein de ce projet, nous avons travaillé sur de la détection automatique de sons d'eau. Ces sons ont, en effet, la particularité de présenter une certaine homogénéité indépendante des maisons. Nous pouvons donc supposer qu'il est possible de créer un système de détection robuste et indépendant du domicile. De plus, à la différence d'autres bruits de la maison liés aux activités, ces sons ont la particularité d'être présents sur un ensemble important d'activités d'intérêt définies par les médecins. Ainsi, les sons d'eau se retrouvent dans les activités d'hygiène (*se laver les mains, se laver les dents*), de nettoyage (*faire la vaisselle*), mais aussi les activités de cuisine (*faire le café*) ou d'entretien (*arroser les plantes*).

Identifier correctement un son d'eau constitue donc un élément bas niveau important pour l'indexation en activité. Nous proposons dans cet article deux approches de détection acoustique de sons d'eau. Une première approche se base sur l'utilisation d'un robinet et vise à reconnaître un jet d'eau continu que nous appelons flot d'eau. Une seconde approche basée sur l'acoustique des sons de liquide permet de reconnaître les autres types de sons d'eau.

## 2. Reconnaissance de flot d'eau

Nous allons décrire dans cette partie nos contributions à la reconnaissance automatique de sons de flot d'eau dans des vidéos d'activités effectuées au domicile du

patient. Nous appellerons « flot d'eau » les sons liés à l'ouverture d'un robinet ou d'une douche. Ces sons ont la particularité acoustique d'être plutôt stationnaires et bruités.

## 2.1. Etat de l'art

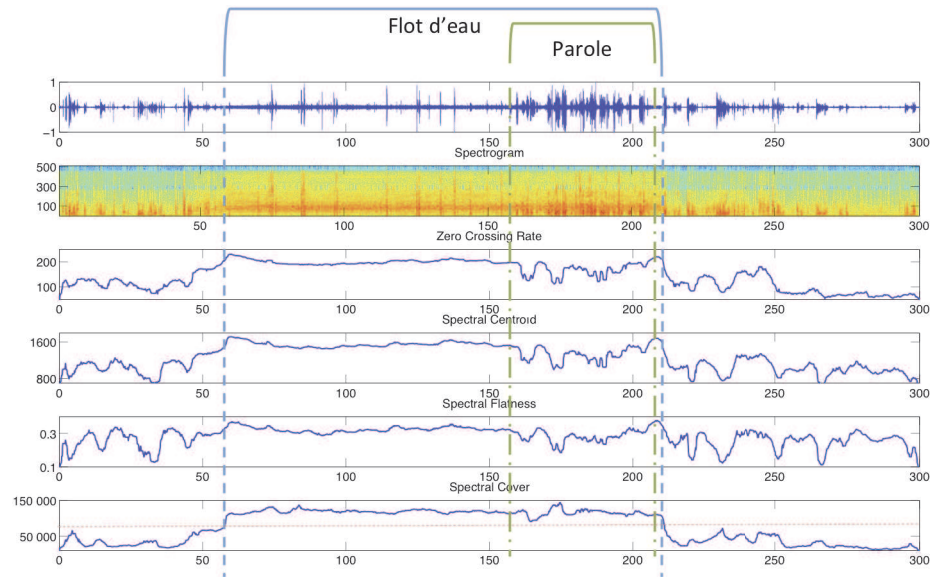
La reconnaissance de flot d'eau pour des applications médicales a déjà été abordée dans plusieurs études scientifiques. Certaines approches utilisent des capteurs placés directement sur les tuyaux pour détecter l'utilisation de l'eau (Fogarty et al, 2006). D'autres utilisent des microphones placés près du bassin pour reconnaître les activités liées à l'eau (Chen et al, 2005). Dans une étude plus proche de notre application, Taati utilise une caméra placée au-dessus d'un lavabo pour détecter l'activité *se laver les mains* (Taati et al., 2010).

Ces précédentes études présentent toutefois le point commun d'avoir été effectuées dans des lieux uniques. Ainsi les données servant à modéliser les sons d'eau et les données visant à tester les systèmes ont été enregistrées dans les mêmes conditions. Celles-ci permettent une utilisation satisfaisante des méthodes d'apprentissage automatique, par exemple des Machines à Vecteur de Support (SVM). Dans notre projet, chaque patient est filmé dans un lieu différent. Les données obtenues sont donc très hétérogènes. Aussi nous avons décidé de nous éloigner des algorithmes d'apprentissage automatique pour privilégier une approche bas-niveau et robuste au changement de domicile.

## 2.2. Descripteurs acoustiques

Dans un premier temps, nous avons comparé des descripteurs acoustiques qui pouvaient caractériser la forme bruitée et continue des sons de flot d'eau. La figure 2 nous montre une comparaison des valeurs de descripteurs sur une séquence audio de 3 minutes, extraite du corpus IMMED. Dans cette séquence, nous pouvons observer un long passage durant lequel le robinet est ouvert et le flot d'eau peut être entendu. Dans la dernière partie de ce passage, le patient discute avec le psychologue et la parole s'ajoute au flot d'eau.

Nous pouvons observer sur le spectrogramme la bande bruitée et continue correspondante au flot d'eau. Comme nous pouvons le voir sur la figure, des descripteurs usuels tels que le *Zero Crossing Rate* ou le *Spectral centroid* pourraient être bien adaptés à la détection de zone de flot d'eau, malheureusement leur courbe s'effondre en présence de parole. La courbe du *Spectral Flatness*, souvent utilisé pour décrire l'aspect bruité d'un son, est quant à elle très instable en dehors de la zone de flot d'eau, du fait de la présence importante de bruits dans le corpus (choc, frottement sur caméra).



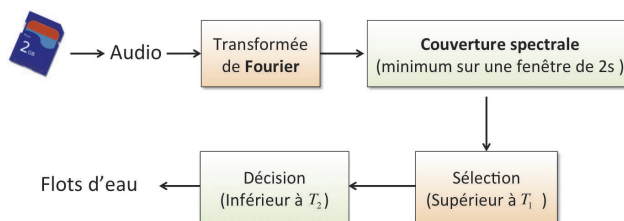
**Figure 8 • Comparaison d'un ensemble de descripteurs acoustiques**

Nous avons introduit un nouveau descripteur, appelé *Spectral Cover*, dont les caractéristiques permettent de détecter les sons de flots d'eau tout en étant robuste à la voix. Comme nous le voyons sur la figure 2, un simple seuillage suffit à isoler la partie du flot d'eau du reste de l'enregistrement. Ainsi, à partir de ce descripteur, nous avons pu définir un système de reconnaissance (Guyot et al, 2012).

### 2.3. Système de détection de flot d'eau

La figure 3 illustre le fonctionnement de notre système de détection. Dans un premier temps, la partie de l'audio est extraite de la vidéo. Nous calculons alors la couverture spectrale sur l'ensemble du flux audio. Un premier seuillage ( $T_1$ ) permet de sélectionner des segments pouvant correspondre à des sons d'eau. Un deuxième paramètre ( $T_2$ ) nous permet d'éliminer d'autres sons dont la couverture spectrale est trop élevée, par exemple les sons provoqués par un aspirateur. La sortie du système est constituée de segments de flots d'eau. Les paramètres de ce système ont été ajustés sur un fichier du corpus IMMED qui a été utilisé pour le développement.





**Figure 9 • Système de segmentation de flot d'eau**

Au final, ce système assez simple n'utilise pas d'apprentissage et peut être utilisé dans n'importe quel domicile. Il obtient pourtant des résultats supérieurs aux systèmes de l'état de l'art qui s'adaptent mal aux problématiques du projet IMMED.

#### 2.4. Construction d'un système hiérarchique

Dans une autre étude (Guyot, Valero et al, 2013), nous rajoutons à ce système une deuxième étape de classification basée sur de l'apprentissage pour raffiner la sortie de notre système. Nous utilisons comme étape de segmentation le système décrit ci-dessus, la sortie du système n'étant plus constituée de flots d'eau mais d'évènements clefs que nous classons dans la deuxième étape. Pour maximiser le rappel de ce système de segmentation, les paramètres ont été légèrement modifiés.

Pour la classification, nous utilisons les gammatones qui ont récemment montré de bonnes performances sur les sons environnementaux (Valero, 2012). Pour le système de classification automatique, plusieurs méthodes furent testées et les machines à vecteur de support furent finalement sélectionnées. Cette étape de classification permet de classer les évènements clefs comme flots d'eau ou autre.

#### 2.5. Expériences

Le tableau 1 présente les résultats de ces différents systèmes. Les expériences ont été effectuées sur un corpus de 20 fichiers du projet IMMED. Ces fichiers audio ont été enregistrés dans 20 domiciles différents. La durée totale de ces fichiers est de 7 heures et 30 minutes. Une vérité terrain, annotée manuellement, indexe 85 segments de son d'eau pour une durée totale de 25 minutes.

	Précision	Rappel	F-mesure
MFCC / GMM	35%	87%	45%
Gammatones / SVM	48%	86%	53%
Système spectral cover	54%	83%	66%
Système hiérarchique	79%	86%	82%

**Tableau 1 • Résultat des systèmes de reconnaissance de flot d'eau.**

Les deux premiers systèmes sont basés sur des techniques d'apprentissage automatique, les résultats étant obtenus par des méthodes de validation croisées dont les

vidéos de test et d'apprentissage proviennent de domiciles différents. Le premier système est un classique de l'état de l'art qui utilise des coefficients cepstraux (MFCC) et un mélange de lois gaussiennes (GMM). La deuxième ligne de ce tableau présente les résultats du système Gammatones/SVM, identique à celui utilisé dans notre système hiérarchique, mais cette fois-ci utilisé seul sans segmentation préalable.

Notre système, basé sur le *Spectral Cover* décrit dans la partie 2.3, est présenté sur la troisième ligne. Nous pouvons observer des résultats supérieurs aux systèmes classiques fondés sur de l'apprentissage, qui s'adaptent mal aux problématiques du projet IMMED. La dernière ligne présente les résultats du système hiérarchique décrit en 2.4. L'utilisation d'une étape de classification supplémentaire améliore largement les scores de reconnaissance. Nous supposons donc que la première étape de segmentation basée sur un paramètre bas-niveau permet de créer des données plus homogènes qui peuvent être par la suite utilisées au sein d'une classification automatique.

La détection de flots d'eau admet ainsi des scores satisfaisants pour que cette tâche soit incorporée dans un système plus large de reconnaissance d'activités. Toutefois certains sons d'eau ne sont pas détectés par cette méthode, même en privilégiant le rappel du système. En effet, les activités liées à l'utilisation de l'eau n'impliquent pas forcément un son de flot d'eau. Dans la partie suivante, nous allons caractériser plus finement les sons d'eau, en nous basant sur des éléments acoustiques.

### 3. Reconnaissance de sons de gouttes d'eau

#### 3.1. Etat de l'art : acoustique

L'acoustique des sons de liquide, est particulièrement de l'eau, a été étudiée depuis de nombreuses années (Bragg, 1921). Il peut être étonnant de constater que l'eau en elle-même ne produit pratiquement aucun son. C'est en général quand l'air est emprisonné dans l'eau que les sons d'eau peuvent être entendus. Ce phénomène est, par exemple, observable en pleine mer, quand les vagues deviennent bruyantes lorsqu'elles se cassent en emprisonnant de l'air.

Les sons de liquide viennent ainsi principalement de la vibration de bulle d'air dans l'eau. Ils sont donc constitués d'une multitude d'éléments sonores discrets dont la localisation dans le plan temps-fréquence est reconnue comme étant un indice perceptif permettant de reconnaître les sons de liquides (Geffen et al., 2011). La vibration de ces bulles d'air est étudiée par les acousticiens (Leighton, 1997), mais elle est également utilisée dans les techniques de synthèse sonore pour générer des sons de liquides divers à partir de la vibration unique d'une bulle d'air (Van den Doel, 2005).

### 3.2. Modélisation de la vibration de bulle d'air

La vibration d'une bulle d'air peut être modélisée par un système masse-ressort. La réponse impulsionnelle de l'oscillation radiale d'une bulle d'air sphérique est donné par :

$$u(t) = a \sin(2\pi ft)e^{-dt}$$

La fréquence de résonance  $f$  étant liée au diamètre de la bulle d'air, tout comme le facteur d'amortissement  $d$ .

Une bulle d'air apparaît généralement lors de la chute d'une goutte d'eau dans l'eau. Lorsque la bulle d'air monte à la surface, la masse d'eau qui recouvre la bulle diminue, et la fréquence de vibration de la bulle augmente. Ce phénomène constitue un indice perceptif important pour reconnaître les sons de liquide. Nous pouvons observer ce phénomène dans un son de goutte d'eau, comme sur la figure 4 où la montée en fréquence succède à l'impact de la goutte dans l'eau au temps 0.

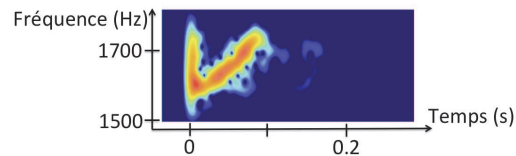


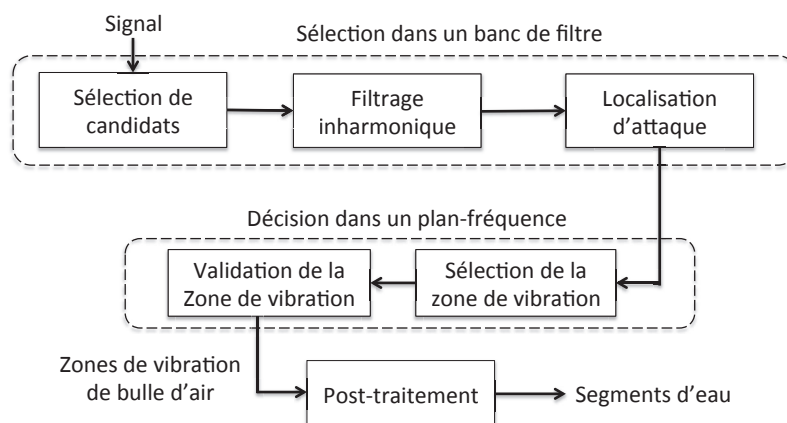
Figure 10 • Spectrogramme d'un son de goutte d'eau

### 3.3. Détection de bulles d'air

Nous avons ainsi proposé un système de reconnaissance audio basé sur les modèles acoustiques de vibration de bulles d'air dans l'eau (Guyot et al, 2013). L'objectif de cette approche est la détection des sons de bulles d'air dans un spectrogramme (au niveau fréquentiel) à partir d'hypothèse provenant des modèles physiques utilisés en acoustique et en synthèse sonore. Notre système se décompose en trois parties, une sélection de candidats dans un banc de filtre fréquentiel, une décision dans un plan temps-fréquence et une étape finale de post-traitement (cf. figure 5).

#### 3.3.1. Sélection dans un banc de filtre

Dans cette étape, les candidats correspondants à des points temps-fréquence sont sélectionnés dans un spectrogramme. A partir du modèle physique, nous supposons toute l'énergie d'une bulle est contenue dans une seule bande de fréquence du spectrogramme à un instant donné. Nous utilisons un banc de filtre fréquentiels pour sélectionner des candidats. Ces derniers sont des points du banc de filtre qui contiennent 85% de l'énergie du signal en cet instant.



**Figure 5 • Diagramme du système de détection de bulles d'air**

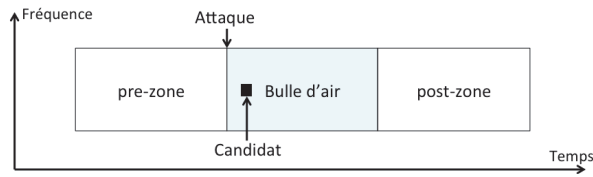
D'autre part, la fréquence de résonance des bulles est liée à leur rayon. Comme les grosses bulles n'apparaissent que rarement dans la nature (sauf quand un très gros objet tombe dans l'eau), nous supprimons les candidats dont la fréquence est inférieure à 800 Hz. Nous supprimons également les candidats situés sur des trames harmoniques dont la fréquence fondamentale est en dessous de ce seuil.

L'étape de sélection se termine par la localisation de l'attaque du son, associé au minimum local d'énergie situé entre le candidat et 100 millisecondes avant celui-ci. Nous supposons que le candidat trouvé correspond au début du son de bulle.

### 3.3.2. Décision dans un plan temps-fréquence

Dans cette phase, nous allons considérer l'ensemble du spectrogramme et associer une zone temps-fréquence à chaque bulle. À partir des modèles physiques de vibration, nous pouvons exprimer l'amortissement de la vibration en fonction de la fréquence de résonance, et ainsi associer des zones temporelles supérieures pour les grosses bulles qui provoquent des sons graves. La taille fréquentielle des zones est fixe et fixée à 500Hz, ce qui permet de contenir les variations fréquentielles de la bulle.

Nous considérons les zones situées avant (pré-zone) et après (post-zone) la bulle d'air (cf. figure 6). Comme les vibrations sont des éléments discrets, nous supposons que l'énergie sera principalement située dans la zone de vibration de la bulle.



**Figure 6 • Localisation dans une zone temps/fréquence**

Enfin, l'étape de post-traitement permet de supprimer les zones détectées isolées pour ne garder que des groupes de bulles qui apparaissent lors des activités liées à l'eau.

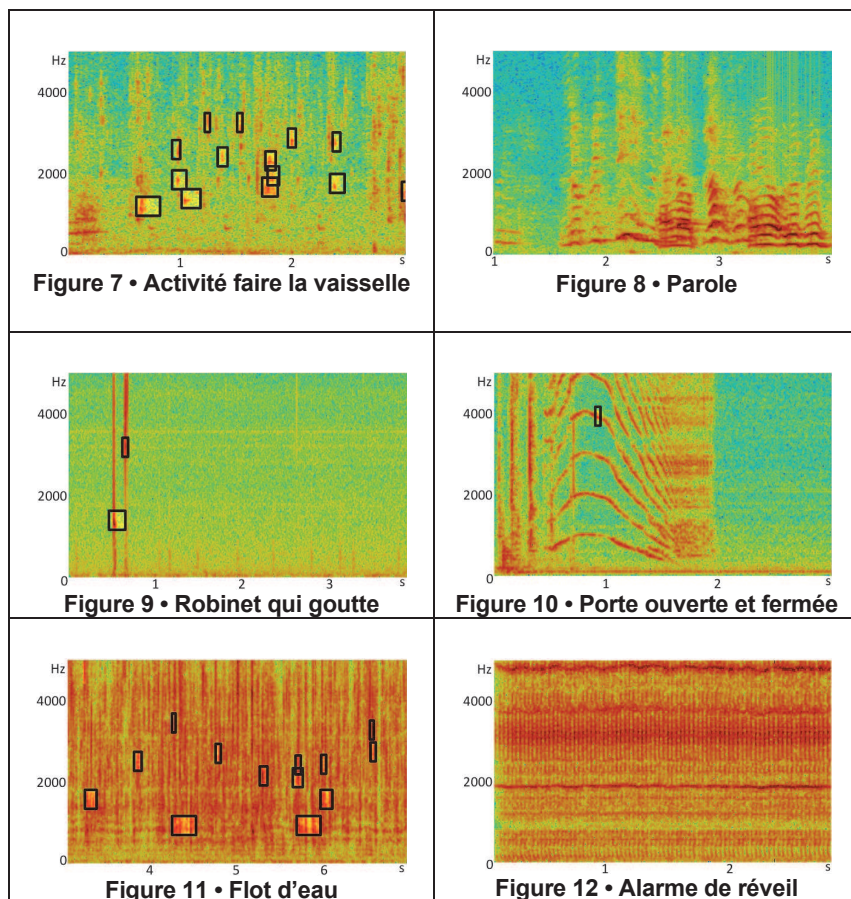
### 3.4. Développement

Nous avons constitué un corpus de développement à partir de deux extraits de 5 secondes du projet IMMED auxquels nous avons ajoutés des sons environnementaux issus du projet Freesound (Freesound, 2012). Ce corpus de 22 sons nous a permis de fixer les seuils de notre algorithme.

Nous pouvons voir sur les figures suivantes les résultats de cet algorithme avant l'étape de post-traitement. Les figures 7 et 8 sont issues du projet IMMED. Les bulles sont détectées en présence de l'activité *faire la vaisselle*. Les quatre autres sons sont issus de la base Freesound. Si notre méthode révèle quelques fausses alarmes (figure 7), l'étape de post-traitement est souvent suffisante pour les supprimer. La figure 8 est intéressante car elle montre que, même en présence d'un flot d'eau plutôt bruyé, certaines bulles ressortent acoustiquement et peuvent être détectées par notre système.

### 3.5. Expériences

Deux expériences ont été effectuées pour valider cette approche. Dans la première nous avons utilisé un corpus (accessible à la communauté scientifique<sup>1</sup>) de 45 fichiers, enregistrés dans une habitation et correspondant à des activités comme *faire la vaisselle* ou au bruit de certains objets. L'objectif de cette expérience était de classer les sons selon la présence d'eau ou non. En appliquant notre méthode de détection de bulles d'air, tous les sons de liquide ont été détectés (*bain, se laver les mains, tirer la chasse, robinet qui coule*). Les fausses alarmes (*scier, allumer une allumette*) révèlent des sons qui semblent être composés d'impacts multiples, localisés dans le plan temps-fréquence, et détectés par notre système.



La deuxième expérience a été effectuée sur le corpus IMMED. Nous avons choisi un fichier de 21 minutes dans lequel une personne âgée effectue diverses activités à la maison. Ce fichier contient 82 secondes de l'activité faire la vaisselle. Comme le flot d'eau est presque inaudible durant cette activité, les précédents systèmes ne pouvaient pas détecter correctement cette activité. L'application de notre méthode de détection de bulle d'air a permis de reconnaître 63 secondes de sons de liquides. Les sons de liquide reconnus sont liés à la manipulation de vaisselle immergée ou à l'utilisation d'une éponge.

#### 4. Conclusion

A travers différentes approches, nous avons développé des méthodes de détection de sons d'eau pour l'indexation en activités du quotidien. Une première approche,

basée sur le signal et les descripteurs acoustiques, permet de détecter les sons de flots d'eau liés à l'ouverture d'un robinet. Les études sur l'acoustique des sons de liquide nous ont permis de développer une seconde approche originale dans le but de détecter d'autres sons d'eau ne faisant pas intervenir de flot, par exemple les sons de gouttes d'eau. Ces deux systèmes nous permettent de reconnaître différentes activités liées à l'eau et pourront être fusionnés dans un système plus complexe d'indexation en activités de la vie quotidienne. Ce dernier système pourra ainsi être utilisé par les médecins pour visionner des activités de la vie quotidienne enregistrés dans des domiciles différents.

Les méthodes de détection de sons d'eau développées peuvent également être utilisées pour d'autres applications liées à l'utilisation de l'eau au quotidien, par exemple la détection de gaspillage d'eau (Terao et al., 2008).

Une taxonomie des sons de liquide du quotidien est actuellement à l'étude dans l'objectif de valider ces deux approches de détection sur un ensemble éclectique de sons de liquide du quotidien. Cette recherche précisera les catégories sonores impliquées dans la perception des sons de liquide. Les résultats nous permettront de comparer nos deux approches de reconnaissance de sons de liquide avec le fonctionnement de la perception humaine. Ces conclusions pourront nous amener à proposer d'autres applications à nos systèmes comme l'identification des activités liées à l'utilisation de l'eau.

---

<sup>i</sup> BBC-Sound-Effect-Library. Disc #3: Household

## BIBLIOGRAPHIE

(Bragg, 1921)

Bragg W.H. (1921). *The World of Sound*, Bell, London.

(Chen et al, 2005)

Chen J., Kam A., Zhang J., Liu N., Shue L. (2005). Bathroom activity monitoring based on sound. *Pervasive Computing, PERVASIVE*.

(Fogarty et al, 2006)

Fogarty J., Au C., Hudson S. (2006). Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. *Proceedings of the 19th annual ACM symposium on User interface software and technology*.

(Freesound, 2012)

Freesound. Collaborative database of creative-commons licensed sound. <http://www.freesound.org> (consulté le 15 décembre 2012).

(Gaëstel et al, 2011)

---

Gaëstel Y., Onifade-Fagbemi C., Trophy F., Karaman S., Benois-Pineau J., Mégret R., Pinquier J., André-Obrecht R., Dartigues J.F. (2011). Autonomy at home and early diagnosis in alzheimer disease: usefulness of video indexing applied to clinical issues. the immed project. *Journal of Alzheimer's and Dementia*, vol. 7, no. 4, pp. 245.

(Geffen et al., 2011)

Geffen M.N., Gervain J., Werker J.F., Magnasco M.O. (2011). Auditory perception of self-similarity in water sounds. *Frontiers in Integrative Neuroscience*, vol. 5.

(Guyot et al, 2012)

Guyot P., Pinquier J., André-Obrecht R. (2012). Water flow detection from a wearable device with a new feature, the spectral cover. *Proceedings of the 10th International Workshop on Content-Based Multimedia Indexing*, CBMI. IEEE.

(Guyot et al, 2013)

Guyot P., Pinquier J., André-Obrecht R. (2013). Water sound recognition based on physical models. *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing*, ICASSP. IEEE.

(Guyot, Valero et al, 2013)

Guyot P., Valero X., Pinquier J., Alías F. (2013). Two-step detection of water sound events for the diagnostic and monitoring of dementia. *Proceedings of the International Conference on Multimedia and Expo*, ICME. IEEE.

(Karaman, 2011)

Karaman S., Benois-Pineau J., Mégret R., Pinquier J., Gaëstel Y., Dartigues J.-F. (2011). Activities of daily living indexing by hierarchical hmm for dementia diagnostics. *Content-Based Multimedia Indexing*, CBMI. IEEE.

(Leighton, 1997)

Leighton T. G. (1997). The acoustic bubble, vol. 10, *Academic Press*, 1997.

(Mégret et al, 2010)

Mégret R., Dovgalecs V., Wannous H., Karaman S., Benois-Pineau J., El Khoury E., Pinquier J., Joly P., André-Obrecht R., Gaëstel Y. (2010) The immed project: wearable video monitoring of people with age dementia. *Proceedings of the international conference on Multimedia*. ACM.

(Mesaros et al, 2010)

Mesaros A., Heittola T., Eronen A., Virtanen T. (2010). Acoustic event detection in real-life recordings. *Proceedings of the 18th European Signal Processing Conference*, EUSIPCO. IEEE.

(Pinquier et al, 2012)

Pinquier J., Karaman S., Letoupin L., Guyot P., Mégret R., Benois-Pineau J., Gaestel Y., Dartigues J-F. (2012). Strategies for multiple feature fusion with hierarchical hmm: application to activity recognition from wearable audiovisual sensors. *Proceedings of the 21st International Conference on Pattern Recognition*, ICPR. IEEE.

(Taati et al., 2010)

Taati B., Snoek J., Giesbrecht D., Mihailidis A. (2010). Water flow detection in a handwashing task. *Canadian Conference on Computer and Robot Vision (CRV)*. IEEE, pp. 175–182.

(Terao et al., 2008)

Terao Y., Mita A. (2008). Robust water leakage detection approach using the sound signals and pattern recognition. *Proceedings of Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*. Society of Photo-Optical Instrumentation Engineers.

(Valero, 2012)



---

Valero, X., Alías, F. (2012) Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. *Proceedings of Transactions on Multimedia*, vol. 14, pp. 1684-1689. IEEE

(Van den Doel, 2005)

Van den Doel K. (2005). Physically based models for liquid sounds. *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 534–546.