Contribution à l'analyse de l'environnement sonore et à la fusion multimodale pour l'identification d'activités dans le cadre de la télévigilance médicale

Dan Istrate

Université d'Evry Val d'Essonne
Ecole doctorale Sciences et Ingénierie

Mémoire d'Habilitation à Diriger des Recherches

Contribution à l'analyse de l'environnement sonore et à la fusion multimodale pour l'identification d'activités dans le cadre de la télévigilance médicale

Dan Istrate

Jury

M. Jacques Demongeot	Professeur des Universités, UJF Grenoble	Rapporteur
M. Hervé Rix	Professeur Emérite, Université de Nice Sophia-Antipolis	Rapporteur
M. Alain Pruski	Professeur des Universités, Université de Metz	Rapporteur
M^{me}. Hélène Pigot	Professeure Titulaire, Université de Sherbrooke	Examinatrice
M. Eric Walter	Directeur de recherche CNRS, LSS Supélec–Paris XI	Examinateur
M. Didier Aubert	Directeur de recherche, IFSSTAR	Examinateur
M. James Crowley	Professeur, Grenoble INP	Examinateur
M. Etienne Colle	Professeur des Universités, Université d'Evry Val d'Essonne	Examinateur
M. Said Mammar	Professeur des Universités, Université d'Evry Val d'Essonne	Directeur des travaux

Université d'Evry Val d'Essonne
Ecole doctorale Sciences et Ingénierie

Mémoire d'Habilitation à Diriger des Recherches

Contribution à l'analyse de l'environnement sonore et à la fusion multimodale pour l'identification d'activités dans le cadre de la télévigilance médicale

Dan Istrate

Jury

M. Jacques Demongeot	Professeur des Universités, UJF Grenoble	Rapporteur
M. Hervé Rix	Professeur Emérite, Université de Nice Sophia-Antipolis	Rapporteur
M. Alain Pruski	Professeur des Universités, Université de Metz	Rapporteur
M^{me}. Hélène Pigot	Professeure Titulaire, Université de Sherbrooke	Examinatrice
M. Eric Walter	Directeur de recherche CNRS, LSS Supélec–Paris XI	Examinateur
M. Didier Aubert	Directeur de recherche, IFSSTAR	Examinateur
M. James Crowley	Professeur, Grenoble INP	Examinateur
M. Etienne Colle	Professeur des Universités, Université d'Evry Val d'Essonne	Examinateur
M. Said Mammar	Professeur des Universités, Université d'Evry Val d'Essonne	Directeur des travaux

A mon épouse et mon fils

A mes parents

Remerciements

Ce manuscrit n'aurait pas existé sans l'accompagnement de mes collègues professionnels, mes amis et ma famille.

Tout d'abord, je tiens à remercier les professeurs M. Alain Pruski, M. Jacques Demongeot et M. Hervé Rix pour m'avoir fait l'honneur d'être rapporteurs de ce manuscrit. Je remercie les professeurs Mme. Hélène Pigot, M. Eric Walter, M. Didier Aubert, M. James Crowley, M. Etienne Colle et M. Said Mammar d'avoir accepté de faire participer au jury et d'examiner ce travail.

Je remercie M. Said Mammar de m'avoir conseillé et avoir eu l'amabilité de relire ce manuscrit. Je remercie vivement, M. Nacef Berkoukchi (Directeur de l'ESIGETEL entre 2003 et 2011) de m'avoir accueilli, soutenu et encouragé dans mes démarches depuis mon arrivée à l'ESIGETEL. Je remercie M. Eric Parlebas (Directeur de l'ESIGETEL) pour m'avoir accordé sa confiance dans mon projet de recherche.

Je remercie chaleureusement M. Jérôme Boudy (TSP) pour sa grande gentillesse et pour toutes ces années de collaboration fructueuse et d'échanges particulièrement riches. Je remercie Mme. Bernadette Dorizzi (TSP) de m'avoir invité à co-encadré des doctorants. Je remercie aussi particulièrement M. Jean Louis Baldinger (TSP), de nos échanges sur des questions très variées.

Je voudrais aussi remercier M. Thierry Joubert (Société Theoris) de son ouverture vers la recherche et plus particulièrement, pour tous les moments de discussions et partage d'informations dans tous les domaines.

Je remercie M. Eugen Iancu (Université de Craiova), qui a su me transmettre les notions de fondamentales pendant mes années d'études et a été très ouvert une fois que je suis devenu à mon tour enseignant, en participant à une collaboration très sincère et fructueuse. Un mot aussi pour Mme. Doicaru Elena qui m'a formé à l'électronique et avec qui j'ai eu des collaborations scientifiques ces dernières années.

Je remercie les étudiants que j'ai pu encadrés pendant leur thèses de doctorat : Hamid, Amine, Paulo, Toufik ; les post-doctorants : Hamid, Jamal et Joan ; et tous les autres étudiants en stage, en échange Erasmus et en projet. Un chaleureux remerciement à Hamid pour les avancées scientifiques obtenues dans le domaine de télévigilance médicale. Il a su faire preuve d'une grande disponibilité tout au long de notre collaboration.

Je voudrais remercier tous mes collègues de l'ESIGETEL avec lesquels j'ai eu des échanges très riches. Je remercie le personnel administratif pour sa constante bonne humeur et son efficacité.

Depuis la soutenance de ma thèse de doctorat j'ai travaillé avec plusieurs collègues et j'ai encadré plusieurs doctorants, stagiaires, post-doctorants, etc. que je voudrais tous leur témoigner ma reconnaissance et les remercier, même s'ils ne se trouvent pas cités.

Je tiens à exprimer mes plus vifs remerciements à ma famille pour m'avoir soutenu, encouragé et compris durant toutes ces années.

Sommaire

I. Curriculum Vitae 9
I.1. Etat civil 9
I.2. Diplômes universitaires 9
I.3. Situation professionnelle actuelle 10
I.4. Parcours professionnel 10
I.5. Thématiques de recherche 10
I.6. Responsabilités pédagogiques et administratives 11
I.7. Bilan des encadrements 11
I.8. Bilan des publications 11
I.9. Participations aux projets de recherche 11
I.10. Représentations / Rayonnement /membre de sociétés savantes / expertise 12
II. Introduction générale du document 13
III. Activités de recherche et d'encadrement 14
III.1. Résumé 14
III.2. Activités d'encadrement 16
III.2.1. Encadrement de Post-doctorants 17
III.2.2. Encadrement de doctorants 19
III.2.3. Encadrement de Masters 2 21
III.2.4. Encadrement Ingénieur de recherche 23
III.2.5. Encadrement stagiaires fin d'études ingénieur 23
III.3. Administration de la recherche 25
III.3.1. Responsabilités scientifiques 25
III.3.2. Rayonnement 27
III.3.3. Participation à des groupes de recherche 27
III.3.4. Relecture d'articles, activités éditoriales 28
III.3.5. Expertise 28
III.3.6. Collaborations nationales et internationales 28
III.4. Activités d'enseignement 30
III.5. Responsabilités pédagogiques 34
III.6. Développements pédagogiques 34
III.7. Collaborations internationales 35
III.8. Retombées de la recherche sur l'enseignement 35

IV. Synthèse des activités de recherche....36
IV.1. La télévigilance médicale....36
IV.2. La reconnaissance des sons....38
IV.3. La fusion de données multimodales....43
IV.4. La segmentation et la reconnaissance du locuteur....47
IV.5. Les projets de recherche....51
IV.5.1. Projet CompanionAble....51
IV.5.2. Projet QuoVADis....52
IV.5.3. Projet INEASE-CAMED....53
IV.5.4. Projet Sweet-Home....54
V. Conclusions. Perspectives....56
VI. Annexe....59
VI.1. Références....59
VI.2. Liste des publications et communications....63
VI.2.1. Participation à des ouvrages....63
VI.2.2. Articles de revues internationales avec comité de lecture....64
VI.2.3. Articles de revues nationales avec comité de lecture....64
VI.2.4. Conférences internationales avec actes et comité de lecture....65
VI.2.5. Conférences nationales avec actes et comité de lecture....68
VI.2.6. Communications industriels....69
VI.2.7. Communications sans acte....69
VI.2.8. Rapports de recherche....69
VI.3. Articles joints....71

I. Curriculum Vitae

I.1. Etat civil

Nom : Istrate
Prénom : Mircea Dan
Date et lieu de naissance : le 04 février 1976 à Craiova – Roumanie
Nationalité : française
Situation de famille : marié, 1 enfant
Adresse personnelle : 60 Rue de Montgeron, 91800 Brunoy

I.2. Diplômes universitaires

Décembre 2003 — **Docteur en sciences** de l'INP de Grenoble
Spécialité : Signal, Image, Parole, Télécommunications
École Doctorale : École Doctorale d'Électronique, Electrotechnique, Automatique, Télécommunications, Signal (EEATS)
Titre : Détection et Reconnaissance des sons de la vie courante pour une application de télésurveillance médicale
Laboratoire : Communication Langagière et Intéraction Personne-Système (CLIPS – IMAG)
Jury : M. James Crowley (INPG) – Président du jury
M. Gaël Richard (ENST) – Rapporteur
M. Michael Ansorge (Université de Neuchâtel) – Rapporteur
M. Eric Castelli – Directeur
M. Laurent Besacier – Co-Directeur
M. Pierre Yves Coulon (INPG) – Examinateur

Juillet 2000 — **Diplôme d'études approfondies (DEA)** de l'INPG
Spécialité : Signal, Image, Parole, Télécommunications
École Doctorale : École Doctorale d'Électronique, Electrotechnique, Automatique, Télécommunications, Signal (EEATS)
Mention : Assez bien
Titre : Validation d'un algorithme de localisation, vers la reconnaissance des sources de bruit

Septembre 1999 — **Diplôme d'ingénieur** de l'Université de Craiova (Roumanie), Faculté d'Automatique, Ordinateurs et Electronique
Spécialité : Electronique
Mention : Très bien (9.37/10)
5ème année effectuée à l'ENSERG validée en Roumanie

I.3. Situation professionnelle actuelle

Enseignant-Chercheur à l'Ecole Supérieure d'Ingénieurs en Informatique et Génie des Télécommunications (ESIGETEL) – Fontainebleau-Avon
Recruté en septembre 2005.

ESIGETEL
1 Rue du Port de Valvins
77210 Avon Cedex
Tel. 01 60 72 70 51
Fax. 01 60 72 11 32

I.4. Parcours professionnel

09/2004 – 08/2005	**Post-Doctorant** au laboratoire d'informatique d'Avignon (LIA) *Recherche* : LIA *Enseignement* : vacations à l'IUP Avignon
10/2003 -08/2004	**Attaché Temporaire d'Enseignement et de Recherche** (ATER) *Recherche* : CLIPS-IMAG *Enseignement* : IUT1 Grenoble
10/2000-09/2003	**Doctorant**, **Allocataire de recherche** au laboratoire CLIPS-IMAG *Recherche* : CLIPS-IMAG *Enseignement* : vacations à l'ENSERG

I.5. Thématiques de recherche

Les principales thématiques de recherche que j'ai abordées sont :

- **Télévigilance médicale** : détection de chute ou de situation de détresse
- **Biométrie** : reconnaissance du locuteur

Les principaux axes scientifiques sont :

- La détection d'événements sonores : *Transformée en ondelettes (DWT).*
- Reconnaissance des sons : *Mélange de distributions de Gauss (GMM), machines à support vecteur (SVM), paramètres acoustiques basés sur la DWT.*
- Reconnaissance des expressions de détresse : *Modèles de Markov cachés (HMM), HTK.*
- Reconnaissance du locuteur mono ou multicanal : *Mélange de distributions de Gauss (GMM).*
- Fusion de données : *Logique floue, réseaux d'évidence, théorie de Dempster Schaffer.*
- Mise en œuvre temps réel : agrégation des temps creux.

I.6. Responsabilités pédagogiques et administratives

Responsable du laboratoire de Recherche et Innovation Technologique (LRIT) de l'ESIGETEL depuis 2009
Responsable de la Voie d'Approfondissement « Technologies des systèmes embarqués » (TSE) en 3ème année, 2005-2011
Responsable du Domaine d'Application « Télémédecine et Systèmes d'Information en Santé » créé en 2011

I.7. Bilan des encadrements

La synthèse des encadrements que j'ai effectués :

- 1 thèse de doctorat soutenue en 2010
- 3 thèses de doctorat en cours
- 3 post-doctorants
- 4 stages Masters 2
- 1 Ingénieur de recherche
- 5 stages de fin d'étude d'ingénieur

I.8. Bilan des publications

Mes publications après la soutenance de la thèse se répartissent comme suit :

- 7 Chapitres d'ouvrages
- 3 Articles de revues internationales avec comité de lecture
- 5 Articles de revues nationales avec comité de lecture
- 38 Conférences internationales avec actes et comité de lecture
- 8 Conférences nationales avec actes et comité de lecture
- 3 Communications industrielles
- 2 Communications sans actes
- 17 Rapports de recherche (rapports de projet européen ou national)

I.9. Participations aux projets de recherche

Depuis mon recrutement à l'ESIGETEL, j'ai participé à plusieurs projets de recherche :

- En tant que responsable du projet :
  - 1 projet international de type PHC
- En tant que responsable scientifique de l'ESIGETEL dans le projet :
  - 1 projet européen
  - 3 projets nationaux
  - 1 projet international de type PHC
- En tant que participant :
  - 1 projet FEDER

I.10. Représentations / Rayonnement /membre de sociétés savantes / expertise

- Organisation et présentation au Colloque STIC et Santé « Les technologies numériques au service de la santé et du mieux-vivre partagé », le 31 mars 2011 à l'ESIGETEL.
- Membre de :
  - GDR ISIS (Information, Signal, Images et Vision) depuis 2006
  - GDR STIC Santé depuis 2006
  - Réseau français de compétences en télésanté CATEL depuis 2010
  - Grappe d'entreprises Sol'lage anciennement Réseau de compétences en Gérontechnologies Charles Foix depuis 2008
  - Société Française des Technologies pour l'Autonomie et Gérontechnologies (SFTAG) depuis sa date de création en 2007.
  - International Society for Gerontechnology depuis 2009
  - IEEE depuis 2009
  - Conférence des Grandes Ecoles (CGE), Commission Recherche et Transferts depuis 2009
- Relecture d'articles de plusieurs revues et conférences IEEE.
- Expert scientifique européen à la campagne AAL 2008.
- Expert ANR (plus spécialement CONTINT 2010).

II.Introduction générale du document

Je suis enseignant-chercheur à l'ESIGETEL depuis 2005. J'ai soutenu ma thèse de doctorat en 2003 sur la « Détection et Reconnaissance des sons de la vie courante pour une application de télésurveillance médicale ». J'ai poursuivi dans un premier temps cette recherche dans le cadre d'une année d'ATER à l'IUT1 de Grenoble, année pendant laquelle j'ai terminé la mise en œuvre temps réel et l'évaluation du système d'analyse de l'environnement sonore. Par la suite, j'ai effectué une année de post-doctorat au LIA sur la segmentation en locuteur multi-canal où j'ai appliqué mes connaissances en prétraitement du signal et classification par GMM. Une fois recruté à l'ESIGETEL, j'ai créé l'équipe ANASON dans laquelle j'ai mis en place deux axes de recherche, en continuité de mes travaux précédents : analyse de l'environnement sonore et reconnaissance du locuteur et un nouvel axe - la fusion de données. Ce dernier axe a été choisi afin de permettre à la fois de fusionner différentes méthodes pour le traitement du son mais aussi le couplage du système avec d'autres types de capteurs dans le cadre de la télévigilance médicale. Sur chaque axe de recherche, j'ai recherché et j'ai monté plusieurs collaborations avec des partenaires académiques et industriels. Tous les stages et les thèses de doctorat, ont été co-encadrés avec ces partenaires. Ces activités de recherche ont été portées par des projets nationaux, européens ou internationaux. J'ai participé activement au montage et par la suite au bon déroulement des projets de recherche financés par l'ANR, l'Europe ou le ministère des affaires étrangères et européennes.

Ma motivation de soutenir une Habilitation à diriger des recherches (HDR) vient du fait que ces 6 ans d'activité après ma thèse m'ont permis de monter une équipe de recherche, d'encadrer des doctorants et Masters 2 et de m'impliquer dans des projets de recherche à l'échelle nationale, européenne et internationale. Mes travaux se sont déroulés dans le domaine de la télévigilance médicale et du maintien à domicile sur plusieurs axes originaux : utilisation de l'environnement sonore (sons de la vie courante), fusion de données entre plusieurs capteurs, contrôle d'accès. Depuis deux ans, j'ai aussi bénéficié de l'expérience d'assumer la responsabilité du laboratoire de recherche de l'ESIGETEL (10 enseignants-chercheurs, 1 chercheur post-doctoral et 4 doctorants). Il est important pour la suite de la ma carrière d'avoir la reconnaissance de mes pairs à travers l'obtention HDR qui me permettra de mener plus loin mes activités de recherche et d'encadrement.

La télévigilance médicale représente un enjeu de la société d'aujourd'hui parce que l'espérance de vie augmente dans tous les pays et les prévisions statistiques annoncent un nombre important de personnes âgées (17% de 60-74 ans en 2030) ou très âgées (12% de plus 75 ans en 2030). Grâce à la progression de la médecine ces personnes peuvent être maintenues plus longtemps à leur domicile mais sont plus fragiles et nécessitent donc des solutions techniques permettant de faciliter la tâche des aidants et d'augmenter le confort de ces personnes. La télévigilance ne se propose pas de remplacer la présence humaine mais de faciliter la tâche et d'offrir une sécurité accrue pour les personnes âgées. Les enjeux de recherche du domaine consiste dans la contrainte de faire fonctionner 24h/7j des systèmes dans l'environnement domestique, et sans être intrusifs ou difficiles à installer. L'analyse sonore est soumise aux contraintes de l'acquisition sonore distante, à la présence des bruits provenant de l'extérieur et à la grande variabilité des sons à reconnaître. La fusion de données doit traiter des signaux de natures différents (binaires ou continus), avec des périodicités différentes et de types différents (périodiques ou asynchrones).

Le chapitre III présente un résumé de mes activités de recherche, d'encadrement et de valorisation. Le chapitre IV présente les activités d'enseignement et ainsi que les responsabilités pédagogiques et administrative et le chapitre V la synthèse de mes activités de recherche. Six articles les plus représentatifs sont joints en annexe.

III. Activités de recherche et d'encadrement

III.1. Résumé

Mes activités de recherche après la thèse de doctorat se sont déroulées dans le cadre du domaine de la télévigilance médicale et ont concerné l'analyse de l'environnement sonore, la fusion de données multimodales et la reconnaissance du locuteur.

Après avoir soutenu ma thèse de doctorat en 2003 et avoir passé une année comme chercheur postdoctoral, j'ai développé dans le cadre du laboratoire de recherche de l'ESIGETEL, créé à mon arrivée au sein de l'école, un axe de recherche sur la télévigilance médicale à travers l'analyse de l'environnement sonore et la fusion de données multimodales. J'ai encadré 5 stagiaires ingénieurs en fin d'études, 3 stagiaires en Master 2, 3 doctorants et 3 post-doctorants. J'ai participé au montage de plusieurs projets de recherche et je suis le responsable scientifique pour l'ESIGETEL dans le cadre d'un projet européen (FP7), 2 projets nationaux (ANR) et responsable d'un projet PHC de collaboration France-Roumanie. Depuis 2009, en tant que responsable du laboratoire de recherche de l'ESIGETEL, j'ai encadré l'activité de recherche de 4 équipes :

- **ANASON** (2 EC, 3 Doctorants, 1 Post-Doctorant) – spécialisée en reconnaissance des sons et des mots clés et la fusion de données multimodales (l'équipe de laquelle j'en fais partie)
- **Loc'In** (4EC) – spécialisée dans la localisation à l'intérieur des bâtiments (indoor) en utilisant des techniques radios
- **SITR** (2EC, 1 Doctorant) – spécialisée en traitement temps réel et recherche d'informations
- **NetS** (2EC) – sécurité de la transmission des données à travers le réseau

Mes activités de recherche se regroupent en 3 thématiques autour de la télévigilance médicale :

- *La reconnaissance des sons de la vie courante et des expressions de détresse*
- *La fusion de données multimodales*
- *La reconnaissance du locuteur*

La reconnaissance des sons de la vie courante et des expressions de détresse est une thématique que j'ai commencée à développer pendant ma thèse de doctorat au laboratoire CLIPS-IMAG (actuellement laboratoire d'informatique de Grenoble - LIG) et en collaboration avec le laboratoire TIMC. Depuis mon recrutement à l'ESIGETEL j'ai approfondi cette thématique en créant l'équipe de recherche ANASON que je dirige depuis 2005. L'activité de recherche concerne plus particulièrement la chaine d'analyse sonore permettant d'écouter en continu l'environnement sonore, d'extraire les signaux utiles et de les classer en temps réel. Pour la détection d'événements sonores une

adaptation automatique de l'algorithme basé sur la transformée en ondelettes, proposé pendant la thèse de doctorat, a été réalisée. Deux approches ont été investiguées pour la classification des sons de la vie courante : l'utilisation des mélanges de distributions de Gauss (GMM) et l'utilisation des machines avec des vecteurs de support (SVM).

Les apports au niveau de la classification basée sur les GMM ont concerné l'adaptation à l'environnement bruité et la modélisation des classes adaptée à la taille des bases de données. Ce travail a été effectué à travers l'encadrement de 2 stages de post-doctorat (Jamal Eddine Rougui et Joan Mouba), il a donné lieu à plusieurs publications dans des conférences internationales [CI3, CI6, CI7] et à un rapport de projet [D4.4].

L'approche SVM est étudiée dans le cadre d'une thèse de doctorat (Mohamed Amine Sehili Janvier 2010 – Janvier 2013) et vise à proposer l'utilisation des SVM à la reconnaissance des sons. Une comparaison GMM/SVM est en cours. Ce travail a donné naissance jusqu'à maintenant, à 2 publications dans des conférences internationales [CI1, CI5] et une dans une revue nationale [RN3].

Une attention toute particulière a été donnée à la mise en œuvre temps réel du système, en étudiant les problématiques d'ordonnancement des tâches dans le cadre d'une collaboration inter-équipes à travers un stage de Master 2 (Frédéric Fauberteau) et en optimisant et parallélisant les traitements. Ce travail a donné naissance à un article de conférence internationale pour la partie ordonnancement [CI11] et à un article de revue [O4] et un article de conférence internationale [CI20] pour la partie sonore.

Les travaux de recherche sur la fusion de données multimodales a démarré en 2006 dans le cadre d'une collaboration avec l'équipe Intermedia (Bernadette Dorizzi, Jérôme Boudy et Jean Louis Baldinger) de Télécom SudParis. Cette collaboration a été conduite à travers 2 thèses de doctorat (Hamid Medjahed, Paulo Cavalcante), plusieurs stages de Master 2 et plusieurs projets de recherche impliquant d'autres partenaires. La détection de situations de détresse pour les personnes âgées vivant seules à la maison est une thématique complexe nécessitant l'utilisation de plusieurs types de capteurs pour garantir une fiabilité élevée. En partant du fait que l'écoute de l'environnement sonore à lui seul ne peut pas assurer une détection fiable de toutes les situations de détresse, je me suis intéressé à la fusion de données issues d'autres types de capteurs.

La collaboration avec Télécom SudParis et INSERM U558 Toulouse m'as permis de réunir trois modalités de détection de situations de détresse : capteur mobile RFPAT [1] (Télécom SudParis), capteur infrarouges (INSERM) et capteur sonore intelligent. L'absence de base de données de situations de détresse m'a amené dans une première étape à étudier l'adaptation de la logique floue pour la fusion de données multimodales. Une application temps réel a été proposée, développée et évaluée dans le cadre d'un projet de recherche ANR (QuoVADis). Ce travail a été valorisé par un chapitre de livre et plusieurs publications dans des conférences internationales. Une deuxième possibilité, basée sur la théorie de Dempster-Shafer et les réseaux d'évidence est actuellement étudiée. Les premiers résultats obtenus sont encourageants et ont donné lieu à 2 publications dans des conférences internationales [O2, CsA1].

[1] RFPAT = capteur mobile sur le patient conçu et réalisé par TSP

La reconnaissance du locuteur est une thématique de recherche que j'ai commencée à étudier pendant mon stage postdoctoral au LIA et que j'ai continuée à l'ESIGETEL dans le cadre d'une collaboration avec Université de Reims Champagne-Ardenne (Michel Herbin et Valeriu Vrabie). En effet, j'ai proposé pendant mon post-doc une méthode d'utilisation des signaux sonores multicanal dans la segmentation en locuteurs, qui s'est avérée très prometteuse [CI22, CI23]. La collaboration avec URCA est centrée autour de l'encadrement de plusieurs stages (Mohamed Chenafa, Nicolae Florin Iancu, Kamel Benhamida). Ce travail a été publié dans 2 revues [O5, RN2], 1 conférence internationale [CI17] et 1 conférence nationale [CN8].

Je me suis aussi intéressé à la reconnaissance du locuteur dépendante du texte pour le contrôle d'accès. Une méthode de fusion de plusieurs techniques de reconnaissance a été proposée, développée et évaluée. Cette méthode est fondée sur une première étape de fusion entre un système d'identification du locuteur indépendant du texte et un système de reconnaissance de mots clés. Une fois un locuteur identifié, une étape de vérification du locuteur sur un autre texte est mise en place. La combinaison de ces trois systèmes nous a permis l'amélioration des performances de reconnaissance par rapport à un système état de l'art de type reconnaissance du locuteur dépendante du texte. Le travail a été publié dans un article de revue nationale et dans plusieurs conférences internationales [O5, RN2, CI16, CN8].

Tous ces travaux de recherche se sont déroulés partiellement ou totalement dans le cadre de plusieurs projets de recherche :

- Projet **CompanionAble** (Integrated Cognitive Assistive & Domotic Companion Robotic Systems for Ability & Security) Projet Européen (FP7) de type IP, depuis janvier 2008.
- Projet **Sweet-Home** (Système Domotique d'Assistance au Domicile) Projet ANR VERSO 2009, depuis novembre 2009.
- Projet **QuoVADis** (Aide à Distance à la Vie Quotidienne pour des personnes âgées atteintes de troubles cognitifs) Projet ANR TecSan 2007, janvier 2008 – juin 2011.
- Projet **INEASE-CAMED** (INtelligent Environment for the ASsitance of Elders. Computer Assisted MEdical Diagnosis) Projet PHC Brancusi Roumanie, janvier 2009 – décembre 2010.
- Projet **IE4IL** (Intelligent Environments for Independent Living) Projet SAFETI Afrique du Sud, avril 2007 – décembre 2010.
- Projet **Telepat** (Télésurveillance de patients à domicile) – Projet RNTS 2003, de 2003 à 2006. Participation depuis 2005.

III.2. Activités d'encadrement

Depuis mon recrutement à l'ESIGETEL en 2005, j'ai commencé à encadrer des stages de Master 2 et des doctorants. Tous les encadrements de stages, doctorants ou post-doctorants se sont déroulés dans le cadre des collaborations avec des partenaires académiques ou industrielles et la plupart avec un financement issu d'un projet de recherche. Je considère très importante cette activité d'encadrement scientifique qui permet de réaliser un transfert de savoir-faire vers les générations futurs. Le nombre de stages encadrés est présenté dans le tableau ci-dessous et les sections suivantes décrivent les sujets des stages.

Type	Nombre	Commentaires
Post-Doctorants	3	
Doctorants	4	1 a soutenu en 2010 et 3 sont en cours
Master 2	4	
Fin d'étude d'école d'ingénieurs	5	
Ingénieur de recherche	1	

III.2.1. Encadrement de Post-doctorants

- **Novembre 2008 – Septembre 2009**

Post-Doctorant : Jamal Eddine Rougui
Sujet : Amélioration de la reconnaissance des sons de la vie courante en présence du bruit
Financement : Contrat CDD, Projet CompanionAble
Résumé : Dans le cadre du projet CompanionAble, Jamal Eddine Rougui a eu comme mission l'évaluation et l'amélioration du système de reconnaissance des sons de la vie courante basé sur GMM, en présence du bruit. La première phase du post-doctorat a été dédiée à l'automatisation des seuils utilisés dans l'algorithme de détection de signaux utiles, basé sur la transformée en ondelettes. Dans la deuxième phase, une évaluation en présence de différents types de bruits a été menée pour déterminer les bruits de la vie courante les plus difficiles pour le système. L'adaptation des modèles GMM à la taille de la base de données disponible a été proposée en permettant l'amélioration des résultats. Cette étude a été menée en étroite collaboration avec AKG [2] (partenaire du projet CompanionAble) qui fournit le microphone CMT (Coincidence Microphone Technology). Ce microphone permet la localisation de la source sonore utilisant les caractéristiques spectrales des 4 capsules microphones qui le composent. Un algorithme de traitement multicanal spécifique pour le CMT a été proposé.
Parallèlement à ces travaux, Jamal Eddine Rougui a participé à la mise en œuvre en C et l'intégration du système dans l'architecture CompanionAble à travers des réunions de travail du projet. Une réorganisation du code en C a été nécessaire pour répondre aux besoins du projet.
Valorisation : 2 publications dans des conférences internationales [CI6, CI7]
Position actuelle : Ingénieur de recherche LINA GRIM, Université de Nantes

[2] AKG = partenaire industriel autrichien spécialisé dans la construction de microphones

- **Octobre 2009 – Aout 2010**

Post-Doctorant : Joan Mouba
Sujet : Traitement sonore multicanal, étude des paramètres acoustiques, intégration du système
Financement : Contrat CDD, ESIGETEL
Résumé : Dans le cadre du même projet CompanionAble, Joan Mouba a eu comme mission l'intégration et la participation aux évaluations du projet CompanionAble. Il a été confronté à la fois aux problématiques spécifiques à l'intégration de modules de différents participants mais aussi aux problématiques d'adaptation du système de reconnaissance des sons à l'environnement réel des maisons de test. Il a aussi réalisé des travaux avec Télécom ParisTech et Télécom SudParis (Jugurta Montalvao) sur des problématiques liées aux paramètres acoustiques en étudiant les paramètres de type « *Ensemble Interval Histogram* » (EIH). Ses travaux ont concerné aussi l'adaptation par apprentissage des modèles des sons pour le microphone CMT.

Valorisation : une publication dans une conférence internationale [CI4]
Position actuelle : Formateur Epignosis Center, Paris

- **Février 2010 – Juillet 2011**

Post-Doctorant : Hamid Medjahed
Sujet : Fusion de données multimodales, évaluation des performances
Financement : Contrat CDD, CompanionAble, Sweet-Home et ESIGETEL
Résumé : Dans le cadre du projet CompanionAble, Hamid Medjahed, a eu comme mission l'adaptation et l'intégration du système de fusion de données multimodales pour la détection de situations de détresse et la localisation de la personne. Nous avons amélioré la fusion de données basée sur la logique floue en l'adaptant aux diverses modalités du projet et en rajoutant des règles spécifiques. Hamid Medjahed s'est intéressé plus particulièrement à la localisation de la personne à travers les capteurs infrarouges et de contact.
Dans le cadre du projet QuoVADis, Hamid Medjahed, a eu comme mission la fusion de données avec intégration dans l'architecture globale. La technique de fusion de données utilisée a été la logique floue proposée dans sa thèse de doctorat. Dans une première étape il a adapté le système basé sur la logique floue aux modalités du projet : capteur mobile RFPAT, capteurs infrarouges, dalles actimétriques, capteurs sonores intelligents. Dans une 2ème étape, il a intégré ce système dans l'architecture globale permettant l'envoi d'alarme vers le serveur du SAMU. Il a aussi participé à la formation des opérateurs du SAMU à l'utilisation du système pour permettre l'évaluation de celui-ci. L'évaluation globale du système de télévigilance a été très positive (4 sur une échelle de 5).
Dans le cadre du projet Sweet-Home, Hamid Medjahed, avec son expérience précédente, a aidé à l'enregistrement et à l'indexation de la base de données sonore et capteurs.
Valorisation : 1 chapitre de livre [O1], 2 revues [O2, RI3], 4 publications dans des conférences internationales [CI2, CI4, CsA1, CsA2], 1 conférences nationales [CN2] et 3 rapports de projet [D4.21, L4.3, L7.3].

III.2.2. Encadrement de doctorants

- **Septembre 2006 – Janvier 2010**

Doctorant : Hamid Medjahed
Directeur de thèse : Bernadette Dorizzi
Encadrement : **D. Istrate (50%)**, J. Boudy (30%), Bernadette Dorizzi (10%) et François Steenkeste (10%)
Sujet : Identification de situations de détresse par la fusion de données multimodales pour la télévigilance médicale à domicile
Financement : ESIGETEL
Ecole doctorale : Sciences et Ingénierie, Spécialité : Science de l'ingénieur de l'Université d'Evry (Site : TSP)
Résumé : Dans cette thèse, qui s'inscrit dans le cadre de la télévigilance médicale, un nouveau système, à plusieurs modalités, nommé EMUTEM (Environnement Multimodale pour la Télévigilance Médicale) a été développé. Il combine et synchronise plusieurs modalités ou capteurs, grâce à une technique de fusion de données multimodale basée sur la logique floue. Ce système peut assurer une surveillance continue de la santé des personnes âgées.

Outre la nouvelle approche de fusion, l'originalité de ce système réside dans sa flexibilité à combiner plusieurs modalités de télévigilance médicale. Il offre un grand bénéfice aux personnes âgées en surveillant en permanence leur état de santé et en détectant d'éventuelles situations de détresse.

Les travaux de recherche ont porté sur :

- L'analyse bibliographique de la fusion de données;
- Choix d'une technique de fusion adaptée à la problématique de la télévigilance médicale ;
- Proposition et développement d'un système de fusion basé sur la logique floue ;
- L'intégration du système sur la plateforme de laboratoire ;
- L'enregistrement d'une base de données ;
- L'évaluation de l'algorithme.

Date de soutenance : 19 janvier 2010
Valorisation : 1 chapitre de livre [O3], 1 revue nationale [RN4], 4 publications dans des conférences internationales [CI8, CI9, CI11, CI15], 3 conférences nationales [CN1, CN5, CN6], 3 rapports de projet [D4.3, D4.5, L4.2] et une communication industrielle [ComI2].
Position actuelle : post-doc de Janvier 2010 à juillet 2011.

- **Janvier 2010 – Janvier 2013**

Doctorant : Mohamed Amine Sehili
Directeur de thèse : Bernadette Dorizzi
Encadrement : **D. Istrate (70%)**, J. Boudy (20%), Bernadette Dorizzi (10%)
Sujet : Reconnaissance de sons d'effraction et de détresse dans un contexte domotique
Financement : Projet Sweet-Home
Ecole doctorale : Sciences et Ingénierie, Spécialité : Science de l'ingénieur de l'Université d'Evry (Site : TSP)
Résumé : Ces travaux de recherche ont lieu dans le cadre du projet SWEET-HOME et vise à adapter et évaluer d'autres algorithmes que les GMM pour la reconnaissance des sons d'effraction, de détresse mais aussi de la vie courante. Le choix des algorithmes devra tenir compte de la qualité des signaux (fréquence d'échantillonnage) et de la

puissance de calcul nécessaire tenant compte qu'à terme l'application devra tourner dans un contexte domotique. L'algorithme est implémenté en C d'une façon modulaire pour permettre son portage facile sur différentes cibles. Une des problématiques importantes de ces travaux de recherche est liée à la présence du bruit et à la grande gamme dynamique des signaux à reconnaître.

Les premiers travaux réalisés visent à comparer les performances de classification des SVM avec ceux obtenus avec des GMM. Les premiers résultats sont en faveur des GMM mais l'utilisation des SVM séquentielles est envisagée.

Dans le cadre du projet, une étude des classes de sons utiles pour l'application a été effectuée. Une étude des caractéristiques des sons à reconnaître est en cours (spectre, forme, durée,…).

Un premier test du système sur la plateforme du laboratoire LIG (porteur du projet) est prévu en décembre 2011.

L'évaluation à mi-parcours par le comité de thèse a été positive et favorable à la continuation de la thèse.

Valorisation : 2 publications dans des conférences internationales [CI1, CI5] et une revue nationale [RN3].

- **Octobre 2009 – Octobre 2012**

Doctorant : Paulo Armando Cavalcante Aguilar
Directeur de thèse : Bernadette Dorizzi
Encadrement : J. Boudy (55%), **D. Istrate (35%)**, Bernadette Dorizzi (10%)
Sujet : Méthodes de détection automatique de situations de détresse fondées sur la fusion de signaux vitaux et de localisation issus des capteurs intégrés dans un environnement de type Smart-Home : Application au suivi à distance de personnes dépendantes à domicile
Financement : Projet CompanionAble
Ecole doctorale : Sciences et Ingénierie, Spécialité : Science de l'ingénieur de l'Université d'Evry (Site : TSP)
Résumé : Ces travaux de recherche se dérouleront dans le prolongement de ceux de Hamid Medjahed sur la fusion de données. Le principal but de la thèse est d'obtenir un algorithme capable de prédire la possibilité d'apparition d'une chute en se basant sur une détection de rupture dans le comportement de la personne. Les travaux se baseront sur des nouvelles modalités de détection comme les accéléromètres et les caméras vidéo. La prédiction de situations de détresse sera envisagée à travers des méthodes d'analyse à long-terme des mesures issues des capteurs.

Les travaux de recherche effectués jusqu'à maintenant ont porté sur :
- L'analyse bibliographique des techniques de fusion de données ;
- L'étude de techniques basées sur les réseaux d'évidence et la théorie de Dempster-Shafer ;
- La proposition et le développement d'un système de fusion basé sur les réseaux d'évidence ;
- Une première évaluation du système proposé ;
- L'enrichissement de la base de données enregistrée par Hamid Medjahed;

Valorisation : 1 chapitre de livre [O2], une publication dans une conférence internationale [CsA1], une publication dans une conférence nationale [CN2].

- **Septembre 2011 – Octobre 2014**

Doctorant : Toufik Guettari
Directeur de thèse : Badr-Eddine Benkelfat
Encadrement : **D. Istrate** (25%), J. Boudy (25%), B.-E. Benkelfat (50%)
Sujet : Analyse des informations générées par une installation domotique dans un contexte d'assistance à l'autonomie de personnes âgées. Mise à disposition des résultats.
Financement : CIFRE, Legrand
Ecole doctorale : Sciences et Ingénierie, Spécialité : Science de l'ingénieur de l'Université d'Evry (Site : TSP)
Résumé : Il s'agit, à partir de la collecte des trames domotiques, d'être en mesure de produire des pré-alertes en cas de doute sur un risque vital pour la personne et de fournir des informations sur l'occupant du logement, susceptibles d'intéresser les aidants familiaux, sociaux et médicaux. Une information de localisation, de l'environnement de la personne (statuts, ouvrants, occultants, éclairage, température,…), de l'activité de la personne (au sens calme ou agité) et de la posture de la personne (couché, assis, debout…) devra être extraite. Le but est de proposer des dispositifs d'analyse pouvant s'intégrer à une installation domotique existante et permettant d'exploiter l'ensemble des autres équipements domotiques de l'installation ; Une contrainte spécifique de la thèse est la limitation des paramétrages spécifiques à l'installation. Un deuxième but est d'être en mesure de détecter la présence de visiteurs ainsi que les absences « normales » des personnes de leur logement afin d'éviter les détections intempestives.

III.2.3. Encadrement de Masters 2

- **Février 2008 – Juin 2008**

Etudiant : Toufik Guettari
Intitulé du Master : Traitement de l'Information et exploitation des données (TRIED), INT- UVSQ
Sujet : Etude d'approches de fusion conçues pour les données audio et vidéo : Application aux données de télévigilance
Financement : Projet QuoVADis
Résumé : Ce stage de Master 2 a visé l'étude des méthodes de fusion de données utilisées dans l'analyse des enregistrements audio-vidéo en vue d'une adaptation aux problématiques de télévigilance médicale. Il s'est intéressé à la fusion de scores de différents systèmes de classification avec pondération mutuelle des vraisemblances des sorties (en utilisant l'entropie des sorties des classifieurs).
La technique de fusion étudiée a été testée en simulation avec les 3 entrées : capteur mobile, capteurs infrarouges et capteur sonore. L'algorithme a été validé et évalué sur une application uniquement sonore pour améliorer les performances de reconnaissance en présence du bruit. En effet, plusieurs modèles pour la même classe de sons à différents rapports signal sur bruit (RSB) ont été créés. Le signal à reconnaitre a été classé en rapport avec tous les modèles et l'algorithme de fusion a permis de pondérer les différentes sorties en fonction de leurs taux de vraisemblance. Cette technique a permis de passer le nombre de fichiers mal classés de 6 /95 à 0/95.
Position actuelle : 2009-2011, Ingénieur de Recherche à TSP. Depuis septembre 2011, doctorant Télécom Sud-Paris - Legrand - ESIGETEL

- **Février 2008 – Juin 2008**

Etudiant : Kamel Benhamida
Intitulé du Master : STIC de l'URCA
Sujet : Reconnaissance du visage dans le cadre d'une application d'identification
Financement : ESIGETEL
Résumé : Ce stage de Master 2 a visé l'étude de méthodes de reconnaissance de visage 2D pour une application de contrôle d'accès dans le cadre d'une collaboration entre le CReSTIC de l'URCA et l'ESIGETEL. La méthode de l'analyse en composantes principales a été mise en œuvre et évaluée dans une première étape sur une base de données de visage mise à disposition sur le web. Dans une deuxième étape, il a participé à l'enregistrement au CReSTIC d'une base de données contenant des images de visages, prises en lumières ambiante, infrarouge et thermique. Une première approche de classification multi-spectrale a été étudiée.

- **Février 2008 – Juin 2008**

Etudiant : Frédéric Fauberteau
Intitulé du Master : Science Informatique, Spécialité : Logiciels des réseaux de l'Université Paris-Est Marne-la-Vallée
Sujet : Agrégation des temps creux pour l'économie d'énergie des capteurs
Financement : Projet QuoVADis
Résumé : Ce stage de Master 2 a visé l'étude de méthodes d'ordonnancement des tâches dans un réseau de capteurs dans le cadre d'une application de télévigilance médicale. Ces travaux de recherche ont abordé le problème de traitement et de fusion de données issues de plusieurs capteurs en partant de l'hypothèse de la distribution des calculs sur les différents processeurs embarqués pour fiabiliser le système et réduire la consommation. Rappelons que l'approche classique consiste à faire circuler l'information brute des capteurs vers un serveur central d'analyse. Une méthode de détection et de prédiction des temps creux (quand le processeur est en attente) a été proposée en vue de dédier ce temps à des traitements des données. Cette méthode permettra de réduire la quantité d'information à faire suivre vers le nœud central donc une optimisation de la bande passante et implicitement de la consommation des capteurs. La méthode proposée a été évaluée sur un simulateur.
Valorisation : 1 publication en conférence internationale [CI11].
Position actuelle : Doctorant à l'université de Marne-la-Vallée

- **Février 2008 – Juin 2008**

Ingénieur : Mohamed Chenafa
Sujet : Etude de la multimodalité biométrique dans le cadre d'une application d'identification
Financement : ESIGETEL
Résumé : Dans un premier temps, une étude bibliographique a été menée sur les différentes modalités biométriques qui peuvent être utilisées pour le contrôle d'accès (reconnaissance d'empreinte digitale, la reconnaissance vocale et la reconnaissance faciale). Dans un deuxième temps, les travaux de recherche ont été consacrés à la reconnaissance automatique du locuteur. Dans ce contexte, une nouvelle méthode de reconnaissance a été proposée. Elle utilise la fusion de données (reconnaissance du locuteur et reconnaissance du mot). Une évaluation du système en présence du bruit a été menée, en utilisant différents types de bruits pouvant être présents à l'entrée d'une salle (bruit de pas, ventilateur,…).

Valorisation : 1 chapitre de livre [O5], une publication en conférence internationale [CI17], 1 revue nationale [RN2] et 1 conférence nationale [CN8].
Position actuelle : Ingénieur Axa Banque

III.2.4. Encadrement Ingénieur de recherche

- **Septembre 2009 – Juillet 2011**

Ingénieur : Toufik Guettari
Sujet : Localisation de la personne à travers des capteurs infrarouges et capteurs de contact
Financement : Projet CompanionAble
Résumé : Une étude bibliographique des méthodes de localisation a été effectuée dans un premier temps. Le positionnement des capteurs infrarouges et de contact de porte pour permettre une meilleure localisation avec une bonne résolution a été étudié pour les différents lieux d'expérimentation : APHP Broca et Smart Homes (SmH) Eindhoven. L'algorithme proposé combine l'information issue des capteurs infrarouges avec celle venant du microphone CMT (Coïncidence Microphone Technology) et celle issue de la caméra vidéo. Les problématiques spécifiques à cette tâche sont :

- La précision très différente des différents capteurs : la pièce ou une zone de la pièce pour les capteurs infrarouges, 15° d'azimut pour le microphone CMT et quelques mètres pour la caméra
- La présence de l'information des différents capteurs n'est garantie en permanence. Le CMT seulement si la personne parle et la caméra uniquement si la personne est dans la seule pièce surveillée par caméra
- La fiabilité de l'information. Le CMT envoie une information de localisation en permanence mais elle n'est fiable que si la personne est en train de parler ; dans ce cas le signal de localisation du CMT a été filtré par l'information issue du capteur sonore intelligent qui détecte la présence de la parole.

Ce module a été mise en œuvre en temps réel et intégré dans l'architecture du projet CompanionAble. Il a été utilisé pour les différents tests du projet.
Valorisation : 2 publications en conférences internationales [CI4, CsA1] et 2 en conférences nationales [CN2, CN6].
Position actuelle : 2009-2011, Ingénieur de Recherche à TSP. Depuis septembre 2011, doctorant Télécom Sud-Paris - Legrand - ESIGETEL

III.2.5. Encadrement stagiaires fin d'études ingénieur

- **Mars 2009 – Juin 2009**

Etudiant : Nicolae Iancu Florin
Intitulé de l'école d'ingénieurs : Faculté d'Automatique, Ordinateurs et Electronique, Université de Craiova - Roumanie
Sujet : Système biométrique basé sur la reconnaissance du locuteur en utilisant des multi-classifieurs
Financement : ESIGETEL
Résumé : Ce stage de fin d'études d'ingénieur a représenté a été réalisé dans la suite de celui de M. Mohamed Chenafa sur la reconnaissance du locuteur. Il a visé le couplage du système de reconnaissance du locuteur par fusion de plusieurs techniques avec un système de détection automatique de présence de parole et de segmentation en mots.

Une évaluation de l'influence de la détection et de la segmentation automatique de la parole a été menée en absence et en présence du bruit. Le système a été mise en œuvre en temps réel en langage C.
Valorisation : 1 publication en revue nationale [RN2].

- **Mars 2009 – Juin 2009**

Etudiant : Badea Bogdan
Intitulé de l'école d'ingénieurs : Faculté d'Automatique, Ordinateurs et Electronique, Université de Craiova - Roumanie
Sujet : Système biométrique basé sur la reconnaissance du visage
Financement : ESIGETEL
Résumé : Ce stage de fin d'études d'ingénieur a visé la mise en œuvre en temps réel de la détection du visage en utilisant une caméra IP. Le stage a débuté avec un état de l'art de méthodes utilisées dans la détection du visage. Le choix a été fait d'évaluer une méthode qui combine une détection dans l'espace des couleurs de la couleur de peau et la reconnaissance de texture de la peau en se basant sur une transformée en ondelettes multidimensionnelle. L'algorithme a été mis en œuvre en C et il a été évalué premièrement sur la base de données de visage enregistrée précédemment au CReSTIC et dans une deuxième étape en temps réel avec la caméra IP.

- **Mars 2010 – Juin 2010**

Etudiant : Ica Bogdan Adrian
Intitulé de l'école d'ingénieurs : Faculté d'Automatique, Ordinateurs et Electronique, Université de Craiova - Roumanie
Sujet : Reconnaissance des expressions de détresse pour une application de télévigilance médicale
Financement : Projet QuoVADis
Résumé : Ce stage de fin d'études d'ingénieur a visé la mise en œuvre en utilisant les outils « open source » HTK d'un système capable à reconnaître des expressions de détresse prédéfinies. L'apprentissage et l'évaluation du système ont été effectués en utilisant un corpus de parole contenant 20 locuteurs, 60 expressions de détresse et 60 expressions normales. Le système a donnée entièrement satisfaction sur cette base. Il a été intégré dans le capteur sonore intelligent et testé en temps réel en couplage avec le système de détection automatique de la parole (détection du signal suivi d'une classification son/parole).

- **Mars 2010 – Juin 2010**

Etudiant : Ioja Andrea
Intitulé de l'école d'ingénieurs : Faculté d'Automatique, Ordinateurs et Electronique, Université de Craiova - Roumanie
Sujet : Envoi d'alarme à travers IP ou GPRS pour une application de télévigilance médicale
Financement : Projet Sweet-Home
Résumé : Ce stage de fin d'études d'ingénieur a visé la conception et la réalisation d'un système capable au moment de l'envoi d'une alarme par le système de télévigilance de choisir la méthode la plus appropriée pour garantir l'envoi de l'alarme. Classiquement, l'alarme est envoyée à travers Internet en utilisant le protocole TCP/IP. Dans le cas de l'installation chez une personne âgée à domicile, l'internet passe par l'ADSL qui ne garantit pas en permanence le lien. Le système développé commence par vérifier la qualité du lien Internet et si nécessaire l'envoi de l'alarme se fera à travers un modem

GPRS sous forme d'un SMS ou envoi de données. L'arrivée correcte de l'alarme est aussi vérifiée. Le système a été codé en C en utilisant un modem GSM/GPRS.

- **Mars 2011 – Juin 2011**

Etudiant : Stefan Todorov
Intitulé de l'école d'ingénieurs : Université "Saints Cyril et Méthode" de Véliko Tirnovo, Bulgarie
Sujet : Conception et Développement d'un agent intelligent pour la télévigilance médicale
Financement : Projet Sweet-Home
Résumé : Ce stage de fin d'études d'ingénieur a visé la conception et la réalisation d'un agent intelligent réalisant la reconnaissance des expressions de détresse pour être intégré dans un système DSS (Decision Support System). La reconnaissance des expressions de détresse a été réalisée en utilisant le moteur de reconnaissance « open source » Sphinx. Le taux de détection d'alarme obtenu a été de 86 % sur les premiers tests effectués.

III.3. Administration de la recherche

Depuis mes travaux de recherche de ma thèse de doctorat qui se sont déroulés dans le cadre d'un projet de recherche national (ANR), en continuant avec le post-doc lui aussi dans le cadre d'un projet national, la plupart de mes activités ont été en lien avec un projet de recherche. Depuis mon arrivée à l'ESIGETEL, j'ai cherché à monter des collaborations avec des partenaires académiques et industrielles qui m'ont permis de participer activement au montage de projets. Par la suite j'ai participé aux activités de ces projets (autant comme activité scientifique et d'encadrement que comme activité administrative).

Type de projet	Nombre	Responsabilités
National (ANR)	3	Responsable scientifique pour l'ESIGETEL pour l'ensemble
FEDER	1	Participant
Européen (FP7)	1	Responsable scientifique pour l'ESIGETEL
International (PHC)	2	Pour l'un responsable du projet et l'autre responsable scientifique pour l'ESIGETEL

III.3.1. Responsabilités scientifiques

Mes responsabilités dans les projets de recherche depuis mon recrutement à l'ESIGETEL ont été :

- **2009-2010 – Responsable coté français du Projet** « PHC Brancusi » **INEASE-CAMED** (*INtelligent Environment for the ASsitance of Elders. Computer Assisted MEdical Diagnosis*). Cette collaboration France-Roumanie a porté sur l'étude et le développement d'applications de télévigilance médicale et du diagnostic assisté par ordinateur. J'ai été l'acteur principal du montage de ce projet. Du point de vue technique, outre la direction des recherches sur la télévigilance médicale, j'ai aussi initié une collaboration avec la partie roumaine sur la prédiction de la

glycémie en fonction des valeurs précédentes. Deux articles ont été publiés dans ce cadre en conférences nationales et internationales [CI12, CN7].

- **2008-2012 – Responsable scientifique de l'ESIGETEL** dans le projet FP7 **CompanionAble** (*Integrated Cognitive Assistive & Domotic Companion Robotic Systems for Ability & Security*) de type IP (Integration Project). Ce projet vise à réaliser la synergie entre la robotique et l'intelligence ambiante pour permettre aux personnes âgées de vivre à domicile de façon autonome aussi longtemps que possible. J'ai participé de façon active au montage de ce projet et je dirige les recherches de l'ESIGETEL dans ce cadre. La principale tâche à laquelle j'ai participé concerne la reconnaissance des sons, des expressions de détresse et des commandes vocales ; je suis responsable d'une tâche de ce projet (T4.1.6). A travers les travaux de Jamal Eddine Rougui et Joan Mouba, que j'ai encadrés, 2 modules sonores (SSI – *Sound Speech Input* et SSA - *Sound Speech Analysis*) ont été développés et intégrés dans l'architecture du projet. A travers le stage de Toufik Guettari que j'ai encadré, un module (LA – *Localisation Abstraction*) permettant la localisation utilisant différentes modalités a été développé. Hamid Medjahed, post-doctorant que j'ai encadré, a développé avec Toufik Guettari le module de fusion permettant de générer l'alarme (PSI – *Person State Integrator*). La thèse de Paulo Cavalcante, que je co-encadre, étudie une autre approche de fusion de données en utilisant les réseaux d'évidence.
- **2008-2011 - Responsable scientifique de l'ESIGETEL** dans le projet ANR-TecSan 2007, **QuoVADis** (*Aide à Distance à la Vie Quotidienne pour des personnes âgées atteintes de troubles cognitifs*). J'ai participé au montage de ce projet et je dirige les recherches de l'ESIGETEL dans le cadre de ce projet. J'ai participé principalement à l'adaptation du système sonore pour ce projet et dans la partie fusion de données à travers la thèse de Hamid Medjahed. Le système de fusion de données développé par Hamid Medjahed a été évalué par le SAMU 92 (partenaire du projet). Sous ma direction, l'ESIGETEL a aussi pris la responsabilité du développement de l'architecture informatique et des interfaces Homme-Machine qui était initialement dévolue à l'ESIEE.
- **2010-2012 - Responsable scientifique de l'ESIGETEL** dans le projet VERSO 2009, **SWEET-HOME** (*Système Domotique d'Assistance au Domicile*). J'ai fait partie du noyau qui a monté ce projet (LIG et ESIGETEL) et je dirige les recherches de l'ESIGETEL dans le cadre de ce projet. La principale tâche à laquelle je participe par le biais de la thèse de doctorat de Mohamed Amine Sehili est celle de la reconnaissance de sons de la vie courante, de détresse et d'intrusion.
- **2011 – Participant projet** FEDER **MEDIATAGS.** Le projet MEDIATAGS a comme but de combiner l'utilisation des tags 2D avec de la localisation à travers les ondes radios (WiFi, Bluetooth) pour faciliter les visites touristiques. Le site choisi pour le prototype est celui des grottes souterraines de Provins. J'ai participé au montage de ce projet. Actuellement, je participe à la conception de l'architecture du système.
- **2007-2010 Participant au projet** SAFETI **IE4IL** (*Intelligent Environments for Independent Living*). Ce projet contribue à lever des obstacles qui constituent un fossé entre les utilisateurs et leur profil environnemental, et l'intégration dans la société et l'économie pour une vraie vie autonome. J'ai participé aux différents séminaires organisés par le projet avec mon apport sur la télévigilance médicale.
- **2003-2006 – Projet Telepat** (*Télésurveillance de patients à domicile*) - RNTS 2003. Admis comme participant à partir de 2005 pour étudier le rajout de la modalité sonore dans le système de télévigilance

J'ai créé l'équipe ANASON en 2005 à mon arrivée à l'ESIGETEL et j'en assure la responsabilité depuis. Quand, en 2009, le laboratoire de recherche de l'ESIGETEL a fait le choix de situer toutes ses équipes de recherche autour de l'axe principale « STIC Santé », je suis devenu le responsable du laboratoire et je dirige 4 équipes de recherche composées de 10 enseignants-chercheurs, 4 doctorants, 2 post-doctorants et 4 stagiaires/an. Ma fonction porte notamment sur l'organisation de réunions mensuelles avec toutes les équipes de recherche et la définition de stratégies de recherche. J'ai aussi organisé annuellement un séminaire recherche du laboratoire.

III.3.2. Rayonnement

- L'article « Medical Telemonitoring System Based on Sound Detection and Classification » publié dans la revue IEEE Transactions on Information Technology in Biomedicine (10:2, Avril 2006) a été **sélectionné et publié dans Yearbook of Medical Informatics 2007**.
- Invitation à participer à la table ronde « La gérontechnologie en questions » dans le cadre du séminaire « **La sécurité des soins en gérontologie » organisé par L'Institut Universitaire de Gérontologie Yves Memin et Faculté de Médecine Pierre et Marie Curie**, le 16 mai 2011. Présentation « *Télévigilance et détection de situations de détresse par écoute de l'environnement sonore* ».
- **Organisation et présentation au Colloque STIC et Santé « Les technologies numériques au service de la santé et du mieux-vivre partagé »** le 31 mars 2011. Présentation « Télévigilance médicale pour personnes âgées ».
- Présentation par Hamid Medjahed, post-doctorant que j'encadre, à la journée « **Avancées en Fusion de données » de GDR ISIS** du 11 février 2010. Titre de la présentation « La Fusion de Données Multimodale pour la Télévigilance Médicale à Domicile ».
- Co-présentation avec Jérôme Boudy aux **Journées Médicales de l'Hôpital de Moinesti (Roumanie)** le 29 juillet 2011. Titre « Remote Medical Monitoring based on several Sensors Combination for Elderly Persons in a Smart Home context ».
- **Chercheur associé** CLIPS-IMAG 2005-2007

III.3.3. Participation à des groupes de recherche

Je suis membre de :

- GDR ISIS (Information, Signal, Images et Vision) depuis 2006
- GDR STIC Santé depuis 2006
- Réseau français de compétences en télésanté CATEL depuis 2010
- Grappe d'entreprises Sol'iage anciennement Réseau de compétences en Gérontechnologies Charles Foix depuis 2008
- Société Française des Technologies pour l'Autonomie et Gérontechnologies (SFTAG) depuis sa date de création en 2007.
- International Society for Gerontechnology depuis 2009
- IEEE depuis 2009
- Conférence des Grandes Ecoles (CGE), Commission Recherche et Transferts depuis 2009

III.3.4. Relecture d'articles, activités éditoriales

J'ai été relecteur pour plusieurs revues internationales et conférences :

- IEEE Transactions on Biomedical Engineering, en 2010
- IEEE Transactions on Audio, Speech and Language Processing (2010)
- IEEE Sensors (2009)
- La Conférence IEEE EMBC (International Conference of the IEEE Engineering in Medicine and Biology Society), tous les ans depuis 2006 à ce jour.
- « International Conference on System Theory and Control » (ICSTC) 2010
- Plusieurs rapports du projet européen CompanionAble

III.3.5. Expertise

J'ai participé en tant qu'**expert scientifique européen à la campagne AAL 2008**. Je suis aussi **expert ANR** (plus spécialement CONTINT 2010).

III.3.6. Collaborations nationales et internationales

La plupart de mes activités de recherche se sont déroulées dans le cadre de collaborations académiques, industrielles et institutionnelles, nationales et internationales.

III.3.6.1. Collaborations nationales académiques

- **2005 à ce jour** – Collaboration avec **Télécom SudParis**, équipe Intermedia (J. Boudy, B. Dorizzi, J.L. Baldinger) sur l'axe de la télévigilance médicale : fusion de données multimodales pour la détection de situations de détresse, fusion de données pour la localisation de la personne, analyse de l'environnement sonore. Co-Réalisation d'un démonstrateur basé sur les 3 modalités (capteur mobile, capteurs infrarouges, son) à Télécom SudParis 2006 et à l'ESIGETEL en 2011. Montage d'un accord cadre de collaboration recherche Télécom SudParis-ESIGETEL, signé en 2011. Montage et co-participation à plusieurs projets de recherche : Telepat (2005-2006), CompanionAble (2009-2012), QuoVADis (2009-2011), IE4IL (2009-2010), INEASE-CAMED (2009-2010). Co-encadrement de 3 thèses de doctorat : Hamid Medjahed (soutenue en 2010), Mohamed Amine Sehili (2010-2013) et Paulo Cavalcante (2009-2012) et plusieurs stages de Master 2 ou d'ingénieurs. Plusieurs publications communes.
- **2008 à ce jour** – Collaboration avec le laboratoire **IBISC Evry** (Informatique, Biologie Intégrative et Systèmes Complexes) équipe HANDS (E. Colle et P. Hoppennot) sur le thème de l'intelligence ambiante. Montage et co-participation à deux projets de recherche : CompanionAble, QuoVADis.
- **2006 à ce jour** – Collaboration avec **INSERM U558** (F. Steenkeste) sur la détection de chute et localisation à travers des capteurs infrarouges. Montage et co-participation au projet de recherche QuoVADis. Co-encadrement d'une thèse de doctorat : Hamid Medjahed (soutenue en 2010). Plusieurs publications communes.
- **2005 à ce jour** – Collaboration avec le **Laboratoire d'Informatique de Grenoble (LIG)** (ancien CLIPS – IMAG) sur la reconnaissance des sons en temps réel. Montage et coparticipation au projet de recherche Sweet-Home (M. Vacher). Plusieurs publications communes.

- **2009 à ce jour** – Collaboration le laboratoire LTCI UMR 5141 de **Télécom ParisTech** (G. Chollet) sur la analyse de l'environnement sonore, reconnaissance des expressions de détresse. Participation conjointe au projet CompanionAble et participation dans l'équipe d'encadrement du stage Master 2 de Daniel Caon.
- **2006 à 2008 et 2011 à ce jour** – Collaboration avec le laboratoire **CReSTIC** (Centre de Recherche en STIC) de l'URCA (Université de Reims Champagne-Ardennes) (V. Vrabie et M. Herbin) sur la reconnaissance du locuteur par fusion d'algorithmes (2006-2008) et la détection de ruptures dans les données de télévigilance médicale (2011 à ce jour). Co-encadrement de stagiaires de Masters 2 ou ingénieurs. Plusieurs publications communes.

III.3.6.2. Collaborations nationales institutionnelles

- **2007 à ce jour** – Collaboration avec l'**Hôpital Européen Georges Pompidou** (HEGP) (P. Espinoza) sur l'amélioration de la transmission du son dans le cadre des téléconsultations. Participation au projet Telegeria. Montage de projets de recherche.
- **2006 à ce jour** – Collaboration avec le **SAMU 92** (M. Baer, A. Ouzgouler, T. Loeb) sur la définition des besoins des systèmes de détection de situation de détresse et le test de ces systèmes. Montage et coparticipation au projet de recherche ANR QuoVADis.
- **2007 à ce jour** – Collaboration avec l'**Hôpital Broca** (A.S. Rigaud) sur le test du système de télévigilance médicale. Montage et coparticipation aux projets de recherche européen CompanionAble et ANR QuoVADis.

III.3.6.3. Collaborations nationales industrielles

- **2006 à ce jour** – Collaboration avec **Theoris** (T. Joubert, Y.Balere, C. Fontaine) sur l'intégration sur système embarqué du système de reconnaissance des sons de la vie courante. Montage et coparticipation au projet de recherche Sweet-Home.
- **2008 à ce jour** – Collaboration avec **Legrand** (M. Teissier, P. Doré) sur la domotique et plus particulièrement les capteurs infrarouges, de porte et les actionneurs. Montage et coparticipation au projet de recherche CompanionAble. Co-encadrement d'une thèse de doctorat CIFRE depuis de septembre 2011.
- **2010 à ce jour** – Collaboration avec **TV77** (E. Roussel) concernant l'intégration de système sonore de détection de situations de détresse dans les décodeurs fibre optique. Montage et participation au projet Mediatags.
- **2011 à ce jour** – Collaboration initiée avec **Axon Cable** sur la participation à la plateforme de télévigilance, actuellement en cours de montage dans la région Champagne-Ardennes.
- **2011 à ce jour** – Collaboration avec **Ynamics** sur la réalisation d'un dispositif déclenchant une alarme au moment de la reconnaissance d'un mot clé prédéfini ou d'un son spécifique pour la protection des petits commerçants. Dépôt d'un dossier de demande de bourse de thèse à la ville de Paris cette année.
- **2011 à ce jour** – Collaboration avec **Overscan** pour la réalisation d'un système capable à détecter la présence ou l'absence d'un son spécifique dans un signal.

III.3.6.4. Collaborations internationales académiques

- **2010 à ce jour - Universidade Federal de Sergipe - UFS (Brésil)** – J. Montalvao – Stage à Télécom SudParis et ESIGETEL. Nous avons effectué des travaux de recherche sur l'utilisation des paramètres EIH (Ensemble Interval Histogram) pour la reconnaissance des sons et sur la comparaison GMM/réseaux de Parzen pour la classification des sons. Ces travaux ont donné lieu à 1 publication dans une conférence internationale [CI4] et 2 publications en conférences nationales [CN3, CN4].
- **2008 à ce jour - Université de Craiova – Faculté d'Automatique, Ordinateurs et Electronique (Roumanie)** – E. Iancu – La collaboration a lieu à travers le projet INEASE-CAMED sur plusieurs sujets : prédiction de la glycémie et reconnaissance des sons du cœur. Ces travaux ont donné lieu à 2 publications en conférences nationales et internationales [CI12, CN7].
- **2007 – University of Virginia (Computer Science department) –** G. Virone – la collaboration a vise l'introduction de l'analyse sonore dans le simulateur d'activité journalières. Les travaux ont été publiés dans une conférence internationale [CI18].

III.3.6.5. Collaborations internationales industrielles

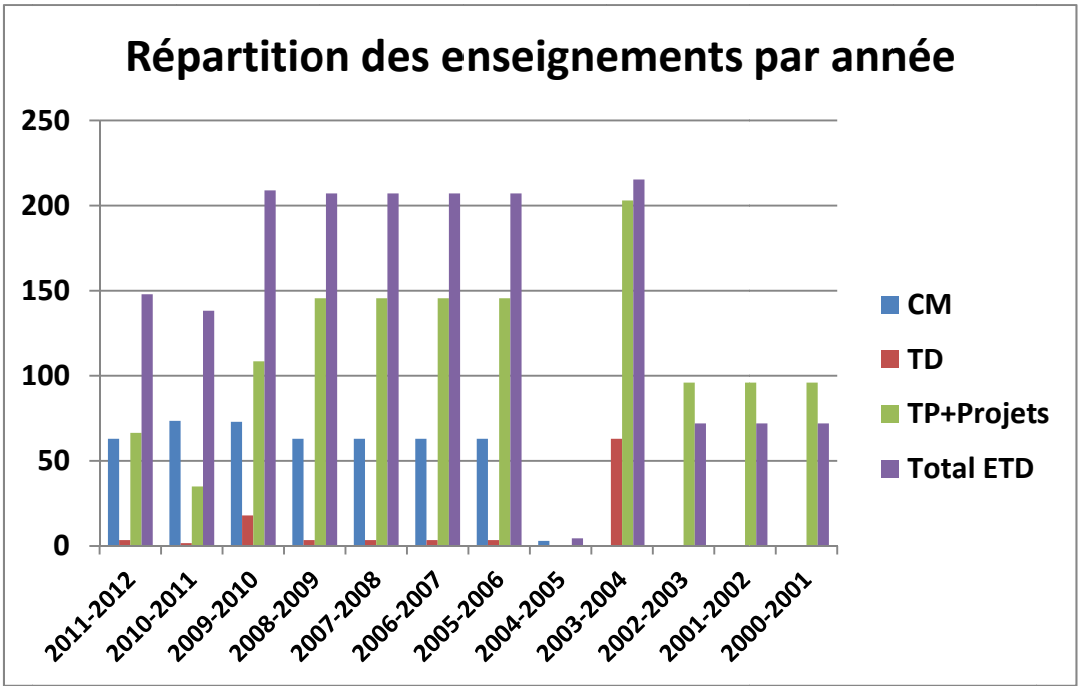
- **2008 à ce jour** – Collaboration avec **AKG** (M. Opitz, M.M. Persy) sur les algorithmes utilisant le microphone CMT (Coïncidence Microphone Technology) pour la reconnaissance des sons et des expressions de détresse. Cette collaboration a été valorisée à travers deux publications dans des conférences internationales [CI4, CI6].

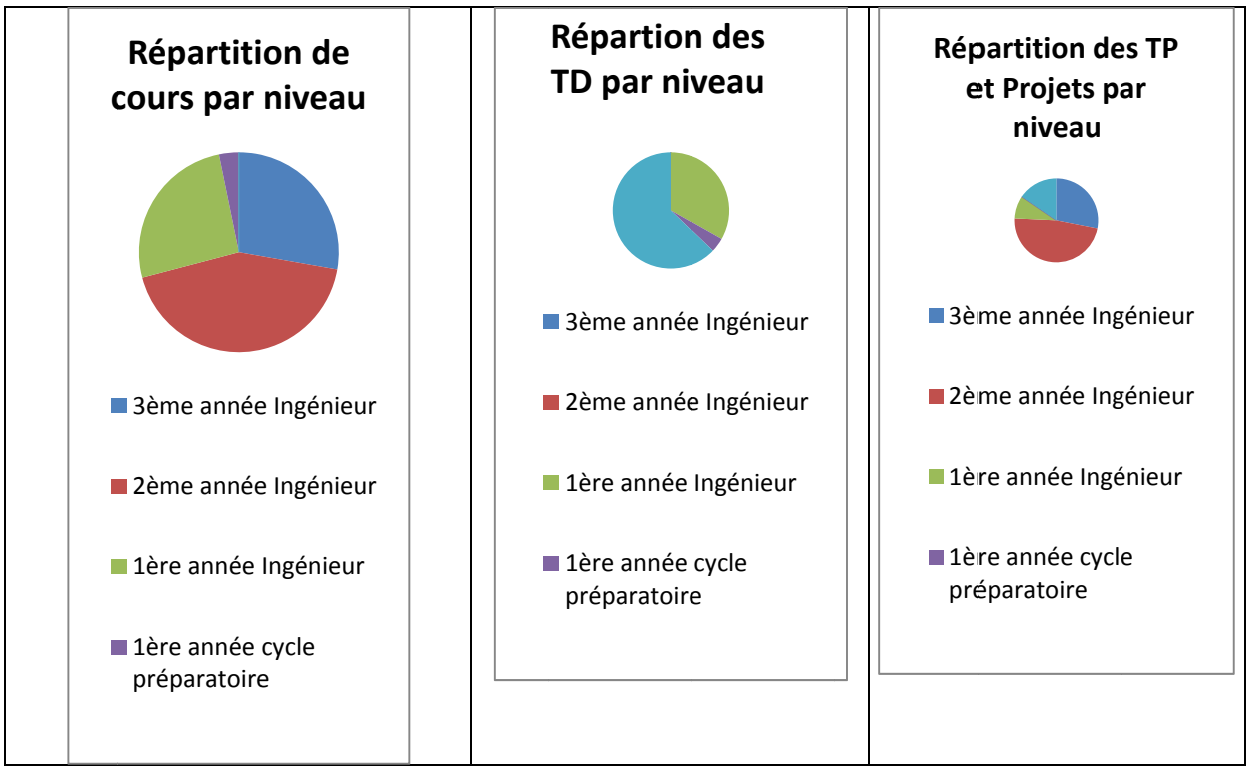
III.4. Activités d'enseignement

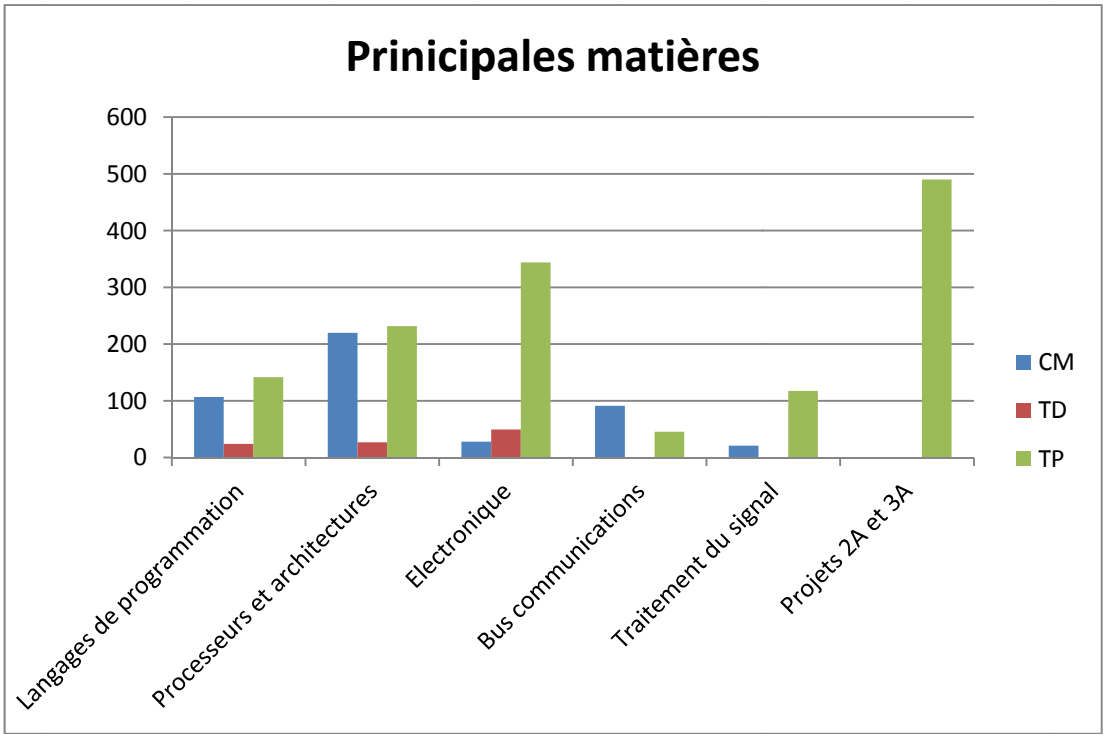
Mes activités d'enseignement sont résumées dans le tableau ci-dessous. Mon expérience de l'enseignement remonte au début de mes travaux de doctorat avec des vacations à l'ENSERG (96h ETD/an). Une fois la thèse soutenue, j'ai poursuivi ces activités avec un poste ATER à l'IUT1 de Grenoble où j'ai effectué 196 h ETD. L'année suivante, alors que j'étais en stage post-doctoral, j'ai pu enseigner un module à l'IUP d'Avignon.

Depuis 2005, en tant qu'enseignant-chercheur permanent à l'ESIGETEL, j'ai effectué 210h ETD/an de 2005 à 2010 et 150h ETD depuis 2010 et jusqu'à présent, ayant pu bénéficier d'une réduction de service pour portage de projets de recherche. Depuis 2005, j'interviens aussi à l'EPITA, à Télécom SudParis et à l'Université de Rennes.

Les graphiques qui suivent représentent la répartition CM/TD/TP par année, la répartition des CM/TD/TP par niveau d'étude et par type de matières.


Répartition des enseignements par année
250
200
150
100
50
0
2011-2012
2010-2011
2009-2010
2008-2009
2007-2008
2006-2007
2005-2006
2004-2005
2003-2004
2002-2003
2001-2002
2000-2001
CM
TD
TP+Projets
Total ETD


Répartition de cours par niveau
3ème année Ingénieur
2ème année Ingénieur
1ère année Ingénieur
1ère année cycle préparatoire
Répartion des TD par niveau
3ème année Ingénieur
2ème année Ingénieur
1ère année Ingénieur
1ère année cycle préparatoire
Répartition des TP et Projets par niveau
3ème année Ingénieur
2ème année Ingénieur
1ère année Ingénieur
1ère année cycle préparatoire



Année	Matière	Niveau	Lieu	Heures Eq. TD
2011-2012 prévisionnel	Protocole CAN (CM, TP)	3ème année ingénieur 20 étudiants	ESIGETEL	148h
	Domotique (CM, TP)			
	Fusion de données (CM, TP)			
	Flexray (CM)			
	Microcontrôleurs (CM, TP)	2ème année ingénieur 20 étudiants		
	C++ (CM, TP)			
	DSP (CM, TP)			
	Labview (CM, TP)			
	Architecture des ordinateurs (CM, TD, TP)	1ère année ingénieur 70 étudiants		
	Programmation C (CM, TP)			
2010-2011	Protocole CAN (CM, TP)	3ème année ingénieur 20 étudiants	ESIGETEL	138h
	FPGA (CM)			
	Domotique (CM)			
	Fusion de données (CM)			
	Bus industriels (CM)			
	Flexray (CM)			
	Microcontrôleurs (CM, TP)	2ème année ingénieur 20 étudiants		
	C++ (CM, TP)			
	DSP (CM, TP)			
	Labview (CM, TP)			
	Architecture des	1ère année ingénieur		

	ordinateurs (CM, TD, TP)	70 étudiants		
	CAN et Flexray (CM)	3ème année ingénieur 30 étudiants	EPITA	21h
	Analyse sonore et télémédecine (CM)	Master « Operational and secure systems » 5 étudiants	Télécom SudParis	4.5h
	Analyse sonore et domotique (CM)	3ème année ingénieur 70 étudiants	ESIR Rennes	3h
2009-2010	Microcontrôleurs (CM, TP)	2ème année ingénieur 20 étudiants	ESIGETEL	210 h
	C++ (CM, TP)			
	DSP (CM, TP)			
	Labview (CM, TP)			
	Architecture des ordinateurs (CM, TD, TP)	1ère année ingénieur 70 étudiants		
	Informatique théorique (CM, TD, TP)			
	Théorie du signal (TP)			
	Initiation à l'informatique (CM,TD)	1ère année cycle préparatoire		
	CAN et Flexray (CM)	3ème année ingénieur 30 étudiants	EPITA	21h
	Analyse sonore et télémédecine (CM)	Master « Operational and secure systems » 5 étudiants	Telecom SudParis	4.5h
2005/2006 2006/2007 2007/2008 2008/2009	Protocole CAN (CM, TP)	3ème année ingénieur 30 étudiants	ESIGETEL	210h
	FPGA (CM)			
	Domotique (CM)			
	Biométrie (CM)			
	Microcontrôleurs (CM, TP)	2ème année ingénieur 20 étudiants		
	C++ (CM, TP)			
	DSP (CM, TP)			
	Labview (CM, TP)			
	Architecture des ordinateurs (CM, TD, TP)	1ère année ingénieur 70 étudiants		
	CAN et Flexray (CM)	3ème année ingénieur 30 étudiants	EPITA	21h
2004/2005	Compression audio : les bases de MP3	Master professionnel 30 étudiants	IUP Avignon	4.5h
2003/2004	Circuits électroniques	1ère année 30 étudiants	IUT 1 Grenoble	39h
	Electronique			80h
	Programmation C			96h
2002/2003	Atelier	2ème année ingénieur 30 étudiants	ENSERG	54h
2001/2002	Traitement du signal			18h
2000/2001	Atelier	2ème année ingénieur 30 étudiants	ENSERG	54h

III.5. Responsabilités pédagogiques

Depuis mon recrutement à l'ESIGETEL, j'ai pris la **responsabilité de la Voie d'approfondissement** (option de spécialisation qui est composée de 20-24 étudiants) **« Technologies des systèmes embarqués » en 3ème année** ce qui implique :

- La réactualisation du contenu des enseignements en tenant compte d'évolution du nombre d'heures et évolution de la demande dans le secteur
- La recherche et l'embauche d'enseignants vacataires (20 enseignants vacataires et 4 permanents)
- Le démarchage des entreprises pour la proposition de sujets de projets de fin d'études et sélection des sujets
- L'administration des polycopiés, des enseignants vacataires et du planning

A cette rentrée, j'ai pris la **responsabilité du Domaine d'Application « Télémédecine et Système d'Information de Santé » en 3ème année**, au montage de laquelle j'ai participé. La Voie d'approfondissement « Technologies des systèmes embarqués » a été redéployée en 2ème année.

J'ai aussi réalisé les polycopies des cours que j'enseigne à l'ESIGETEL (une synthèse est présentée dans le tableau ci-dessous).

Matière	Année réalisation	Année étude	Nombre de pages
Architecture des ordinateurs (CM)	2006	1ère année	366
Théorie des graphes (CM, TD)	2010	1ère année	107
Machines à états finis (CM, TD)	2009	1ère année	212
FPGA (CM)	2005	3ème année	110
Microcontrôleur (CM, TP)	2006	2ème année	238
DSP (CM, TP)	2006	2ème année	257
C++ (CM, TP)	2006	2ème année	230
Labview (CM,TP)	2007	2ème année	170
CAN (CM,TP)	2006	3ème année	360
Flexray (CM)	2010	3ème année	58
Fusion de données (CM)	2010	3ème année	54
Domotique (CM)	2006	3ème année	128
Bus industriels – I2C (CM)	2009	3ème année	77

III.6. Développements pédagogiques

En 2010, nous avons mis en place, avec un collègue, deux spécialités **« Télémédecine et Système d'Information de Santé » et « Réseaux de transport intelligents »** qui contribuent en tant que 2 options en 3ème année de la Voie d'approfondissement « Technologies des systèmes embarqués ». Ces spécialités ont été transformées en 2011 en Domaines d'application en 3ème année.

J'ai aussi participé aux côtés des enseignants de l'IUT de Sénart au montage de la licence professionnelle « Systèmes Embarqués pour la télémédecine» qui a été labélisée par le ministère en 2011.

III.7. Collaborations internationales

J'ai été initiateur et porteur de l'accord-cadre ERASMUS entre l'ESIGETEL et l'Université de Craiova, Roumanie, accord signé en 2008. Dans le cadre de cet accord 3 étudiants français sont partis en semestre d'étude et/ou stage et 7 étudiants roumains ont été accueillis à l'école.

Cet accord a permis aussi 2 déplacements d'échange au niveau des enseignants-chercheurs.

III.8. Retombées de la recherche sur l'enseignement

La recherche effectuée sur la télévigilance médicale m'as permis de :

- Encadrer des projets pour le Concours Imagine Cup : 3ème place à la finale française en 2006 et qualification dans la finale française en 2008.
- Encadrer une équipe participant au concours de la bourse STERIA : gagnante de la bourse en 2011.

IV. Synthèse des activités de recherche

Depuis mon recrutement en 2005 en tant que enseignant-chercheur à l'ESIGETEL j'ai développé un axe de recherche des nouvelles technologies de l'information et de communication (NTIC) pour la santé et principalement : la télévigilance médicale. Cette activité a commencé pendant ma thèse de doctorat avec la reconnaissance des sons anormaux et a été élargie à la fusion de données multimodales et à la reconnaissance du locuteur. Cet axe est devenu l'axe prioritaire du laboratoire LRIT de l'ESIGETEL depuis 2009.

IV.1. La télévigilance médicale

Le développement des technologies numériques a permis leur généralisation dans le domaine médical et non seulement pour la transmission d'images et des sons mais aussi de l'information issue de différents capteurs. Dans le domaine médical, contrairement aux autres domaines économiques, le progrès technologique ne génère pas le plus souvent des gains de productivité mais plus de sécurité et de confort pour les patients.

Un autre facteur important c'est l'augmentation de l'âge de la population dans toutes les sociétés autour du globe. En Europe, par exemple, l'espérance de vie pour les hommes est de 71 ans et pour les femmes de 79 ans. En Amérique du Nord cette espérance est de 75 ans pour les hommes et de 81 ans pour les femmes.

D'autre part, les personnes âgées préfèrent préserver leur indépendance et autonomie en vivant le plus longtemps possible à domicile. En même temps, le nombre de spécialistes en médecine suit une tendance inverse et diminue, fait qui a incité le développement des systèmes techniques pour assurer la sécurité. Les personnes âgées vivant à la maison sont dans la plupart des cas isolées avec un risque élevé d'accidents.

En France, 4.5% des hommes et 8.9% des femmes âgées de plus de 65 ans ont un accident. Parmi ces accidents, la partie la plus importante est représentée par les accidents domestiques (~61%) et 54% ont lieu à l'intérieur de la maison. En France, annuellement, 2 millions de personnes âgées font une chute et cela implique 10000 morts. Entre 30% et 55% des chutes sont la cause de contusions et seulement entre 3% et 13% sont la cause de blessures comme : fractures, luxation d'une articulation, …Outre les blessures physiques et les hospitalisations, les chutes sont aussi la cause des chocs psychiques (surtout si la personne ne peut récupérer complètement après la chute).

Pour augmenter la qualité de vie des personnes âgées plusieurs applications ont été développées ces dernières années : télévigilance médicale à la maison pour détecter des éventuelles situations de détresse et la visio-conférence (en qualité élevée) pour permettre la consultation à distance par de spécialistes.

Le terme « télémedecine » a été mentionné dans le dictionnaire de la langue française la première fois en 1980 avec la signification « médecine à distance », plus précisément une partie de la médecine qui utilise les télécommunications pour la transmission des informations médicales (images, rapports, enregistrements,…) en vue du diagnostic à distance, obtenir un 2ème avis, la surveillance continue d'un patient, prendre une décision thérapeutique.

L'intérêt de la télémédecine est loin d'être prouvé et n'est pas sans ouvrir le champ des réflexions, plus particulièrement dans le domaine éthique, légal et économique. Les principales applications de la télémédecine sont :

- **Télédiagnostic** = l'application qui permet à un médecin spécialiste d'analyser un patient à distance et d'avoir accès à son dossier médical. Un cas spécifique est celui du 2ème avis.
- **Téléchirurgie** = un système technique permettant une opération chirurgicale à distance pour des applications spatiales ou militaires. Dans cette catégorie nous avons aussi l'opération à distance d'un système complexe comme un échographe ou la réalité augmentée pour faciliter le travail du chirurgien.
- **Télévigilance médicale** = un système automatique permettant de surveiller des paramètres physiologiques dans la cas d'une maladie chronique ou pour détecter une situation de détresse.
- **Télé-enseignement** = système de visio-conférence permettant aux médecins d'échanger des informations médicales.

Parmi les applications de la télémédecine, le télédiagnostic et la télévigilance médicale sont ceux qui sont le plus étudiées. Le télédiagnostic permet aux médecins spécialistes de consulter les personnes âgées à travers un lien audio-vidéo de qualité pour éviter des déplacements inutiles pour le patient et le médecin. Plusieurs systèmes ont été développés et évalués entre l'hôpital et les maisons de retraites médicalisées ou entre le personnel médical et unité mobile. Le principal défi est d'assurer en continu une qualité audio-vidéo correspondante, la possibilité de transmettre d'autres données médicales (ECG, enregistrements médicaux, …) et la sécurité des données. Pour garantir la qualité de la transmission audio-vidéo des larges bandes sont nécessaires ; couramment des nouveaux algorithmes de compression ont été développés pour éviter la nécessité d'une large bande de transmission.

La télévigilance médicale peut prévenir ou réduire les conséquences des accidents à la maison des personnes âgées ou avec des maladies chroniques. La télévigilance médicale vise la détection automatique des situations de détresses (chute ou malaise) pour permettre aux personnes âgées de vivre en sécurité chez elles. Les systèmes se basent essentiellement sur les technologies des télécommunications (analyse continue des paramètres du patient : respiratoires, cardiaques, …). Cette technique est utilisée dans le développement de l'hospitalisation à domicile (HAD) quand le patient est surveillé médicalement à la maison et plus spécialement dans le cas des personnes âgées. Le HAD évite des hospitalisations inutiles et augmente le confort et la sécurité du patient. Actuellement, les derniers travaux sur la télévigilance médicale visent aussi la prédiction de l'apparition d'une situation de détresse soit à travers une analyse à long terme de l'activité de la personne, soit par la mesure de l'état d'équilibre de la personne.

Plusieurs équipes de recherche ont travaillé sur la proposition d'un système de détection automatique de situations de détresse et surtout la chute en utilisant des capteurs fixes et/ou mobiles. Pour l'instant il n'y a pas de système automatique commercialisé ; il s'agit que de systèmes basés sur un bouton qui permet de générer à la demande du patient une alarme et une conversation téléphonique avec un centre de télésurveillance. Une approche possible est l'étude des rythmes circadiens obtenus en utilisant des capteurs de position (infrarouge, contact de porte, …) [Bellego et al., 2006]. Cette méthode nécessite des bases de données de taille importante et une adaptation au

patient surveillé [Binh et al. , 2008]. Dans d'autres études, l'activité de la personne est surveillée à travers l'utilisation des différents appareils électroménagers (four, réfrigérateur, filtre à café,…) et l'information est transmise en utilisant des liens 3G [Bairacharya et. al. , 2008]. Pour la détection de la chute, plusieurs capteurs mobiles ont été développés. Ils utilisent des accéléromètres [Marschollek et al. , 2008] des capteurs magnétiques [Fleury et al. , 2007] ou la fusion de données avec des capteurs d'une maison intelligente [Bang et al. , 2008].

Plusieurs projets de recherche se sont intéressés à la télévigilance médicale pour les personnes âgées ou des personnes avec des maladies chroniques. C'est le cas du projet TelePat qui a eu comme but la réalisation d'un service de support à distance pour les personnes avec des maladies cardiaques [Lacombe et al. , 2004]. D'autres projets nationaux comme RESIDE-HIS et DESDHIS ont utilisé différentes modalités comme les capteurs infrarouges, capteur mobile à base d'accéléromètre et l'analyse sonore [RI2]. Au niveau européen, plusieurs projets ont investigué le domaine de la fusion des capteurs de la maison intelligente avec la télésurveillance comme le projet SOPRANO qui a comme but la conception d'un système pour l'assistance aux personnes âgées dans la vie courante pour un meilleur confort et sécurité [Wolf et al. , 2008]. En conséquence, les appareils appartenant à l'intelligence ambiante sont connectés en continu à un centre de services et télésurveillance comme dans le projet EMERGE. Ce dernier a comme but l'observation des habitudes à travers une approche holistique pour détecter les anomalies (chute, malaise ou autre urgence) dans le comportement afin de provoquer l'envoi d'un message d'alarme. Le but de SHARE-IT vise à informer et assister l'utilisateur et le soignant à travers un système de surveillance. Le projet se focalise sur la compatibilité avec les technologies existantes et sur la facilité d'intégration dans les systèmes actuels. Le dispositif embarqué, développé dans le cadre du projet CAALYX, orienté sur la surveillance médicale, est capable de détecter des chutes ou accidents en dehors de la maison et de communiquer en temps réel avec le soignant ou avec le service d'urgence. L'information transmise en cas d'urgence inclut la position géographique et une information médicale concernant la personne. Il existe un projet similaire mais limité à la domotique nommé SENSACTION-AAL et qui vise la prévention des accidents grâce au port d'un capteur intelligent.

Nous pouvons aussi citer au niveau national les projets : QuoVADis et Sweet-Home et au niveau européen : CompanionAble qui sont décrits dans la section IV.5.

IV.2. La reconnaissance des sons

L'information extraite de l'environnement sonore de la vie courante est de plus en plus utilisée pour des applications de type télévigilance pour détecter des chutes, les activités de la vie courante ou pour caractériser l'état physique de la personne. L'utilisation du son comme source d'information a l'avantage du fait que les capteurs, en occurrence les microphones, sont pas chères, sont moins intrusives que la vidéo et peuvent être fixés pour éviter le port par le patient. Sinon, le signal sonore présente une redondance de l'information et nécessite des algorithmes spécifiques pour extraire l'information.

La définition du signal et du bruit est spécifique à chaque application ; par exemple pour la reconnaissance de la parole tous les sons de la vie courante sont considérés comme bruits. Parmi les applications utilisant l'information extraite du son nous avons la caractérisation des sons cardiaques [Lima et Barbarosa, 2008] pour détecter des

maladies du cœur ou la caractérisation des sons de ronflement [Ng et Koh, 2008] pour la détection de l'apnée du sommeil. L'avantage de l'utilisation du son pour la détection d'une situation de détresse a l'avantage du fait que le patient n'est obligé de porter un capteur mobile en permanence et aussi de veiller à la charge de sa pile ; le désavantage principal est constitué par l'influence du bruit et dépend des conditions acoustiques [Popescu et al. , 2008], [Litvak et al. , 2008].

Nous avons porté notre intérêt sur l'analyse du flux sonore pour identifier des situations normales ou anormales de la vie courante en reconnaissant autant des sons (approche nouvelle) que des expressions de détresse. La grande majorité des systèmes actuels utilisent seulement la parole comme source d'information en se privant de la richesse des sons de la vie courante. Très peu d'études s'intéressent aux sons de la vie courante comme dans l'article [Moncrieff et al. , 2005] qui utilise le niveau sonore couplé avec l'utilisation des appareils électroménagers pour estimer l'anxiété du patient. Dans [Stagera et al., 2007] quelques sons spécifiques aux appareils électroménagers sont reconnu sur un microcontrôleur en utilisant de la quantification vectorielle ; en analysant à long terme les activités du patient, une situation anormale pourrait être détectée. Dans [Cowling et Sitte, 2002], un système statistique de reconnaissance des sons a été proposé mais avec une évaluation sur un nombre réduit de classes et de fichiers.

Le système a été développé et amélioré dans le cadre des différents projets de recherche auxquels j'ai participé. L'analyse sonore sert à l'identification des activités de la vie courante, aux commandes vocales (domotique) et à l'identification des situations de détresse à travers des sons anormaux (bris de glace, cris, chute d'objet,…) et des expressions de détresse. Nous avons évité la reconnaissance automatique de la parole continue pour préserver l'intimité de la personne.

Le système d'analyse sonore peut utiliser soit une carte son classique et dans ce cas il pourra analyser jusqu'aux deux canaux audio simultanément ou une carte son/acquisition spécifique avec plusieurs entrées. L'implémentation a été faite en temps réel simultanément sur plusieurs canaux pour permettre l'écoute de microphones installés dans chaque pièce.

Le système d'analyse sonore que j'ai développé tout au long de mes activités de recherche, appelé ANASON, est composé de plusieurs modules :

- Un premier module d'analyse continu du signal pour détecter l'apparition d'un signal utile (la plupart du temps impulsionnel). Ce module de détection est basé sur la transformée en ondelettes.
- Une classification son/parole pour permettre un traitement adapté par la suite. Ce module est basé sur des GMMs.
- Dans le cas d'un son, un module de classification des sons parmi des classes prédéfinies sera utilisé. Ce module est basé sur des GMMs
- Dans le cas de la parole le signal sera analysé par moteur de reconnaissance de la parole cherchant à identifier seulement des expressions de détresse.

Détection et extraction des sons utiles. Ce module est appliqué sur tous les canaux sonores simultanément en temps réel. Son rôle est d'envoyer au module de classification son/parole seulement les parties du signal considéré comme utiles pour l'application. Ce module développé pendant ma thèse de doctorat et amélioré par la suite est basé sur la

transformée en ondelettes qui permet une détection seulement sur certains bandes de fréquence mais aussi une bonne résolution temporelle [RI2].

Du point de vue de l'application de télévigilance médicale on considère comme signaux utiles principalement des sons anormaux mais aussi des sons de la vie courante qui peuvent indiquer le bon déroulement d'une journée de la personne. Nous considérons comme bruit tout son sur lequel peut se superposer des sons anormaux : écoulement d'eau, aspirateur, sèche-cheveux, machine à laver, … La principale caractéristique des sons considérés comme utiles est la présence des hautes fréquences et un démarrage avec une énergie forte ; à l'opposé, les bruits sont plutôt stationnaires et plutôt avec un spectre basse fréquence.

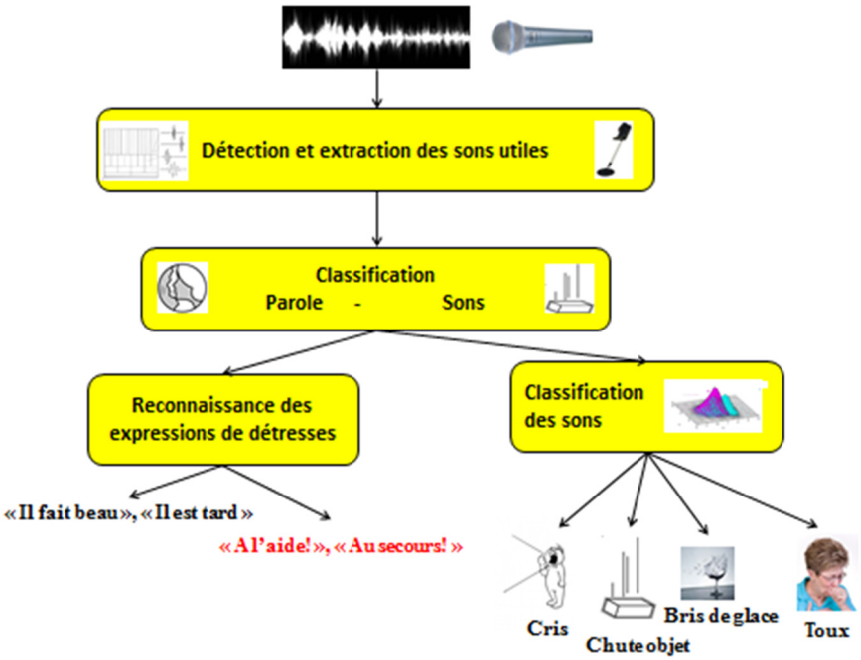


Figure 1 - Architecture du système d'analyse sonore

Tenant compte de ces caractéristiques des signaux à détecter, un algorithme adaptatif par rapport au niveau du bruit stationnaire et capable de détecter, principalement des signaux impulsionnels, a été proposé. Pour l'analyse spectrale du signal, la transformée en ondelettes a été choisie parce qu'elle permet une analyse temps-fréquence avec une meilleur résolution temporelle dans les hautes fréquences qui nous intéressent (Figure 2). Utilisation de cette transformée permet aussi l'utilisation de fenêtres d'analyse de taille relativement importante (réduisant les contraintes temps réel) mais en préservant une bonne résolution temporelle. Le seuillage proposé est de type adaptatif pour permettre de s'en passer du bruit stationnaire ou avec une variation lente. L'adaptation automatique du seuil par rapport aux conditions d'acquisition, a été améliorée par les travaux de Jamal Eddine Rougui, que j'ai encadrés [CI8]. Le module réalise non seulement la détection du début du message mais aussi celle de fin. La détection de fin peut être réglée pour ne pas couper les phrases dans le cas de la parole ou pour segmenter entre les mots.

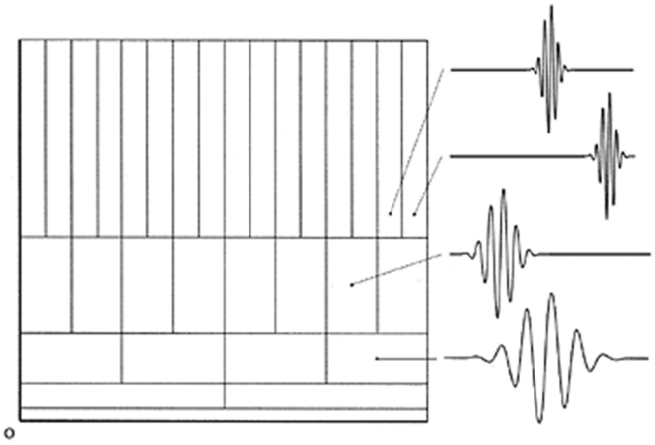


Figure 2 - Décomposition en ondelettes

L'originalité de l'algorithme consiste dans ses performances de détection des signaux utiles même en présence de bruit fort (6% d'erreur à 0 dB de rapport signal sur bruit - RSB) tout en travaillant en temps réel sur plusieurs canaux.

La Classification parole/son et **la classification des sons** sont deux modules qui réalisent une reconnaissance à 2 ou plusieurs classes en se basant principalement, sur un mélange de distributions de Gauss (GMM). L'utilisation des SVM a été et est en cours d'évaluation. La segmentation parole/son/musique représente une application étudiée par plusieurs laboratoires comme première module avant la reconnaissance du locuteur et/ou de la parole. Les solutions proposées se basent essentiellement sur des modèles GMM qui peuvent s'adapter facilement aux signaux à identifier qui n'ont pas de structure spécifique (comme les phonèmes pour la parole). La classification des sons est un domaine nouveau, presque pas étudié au moment du début de ma thèse et surtout du point de vue de l'application de télévigilance médicale. La durée réduite des signaux à reconnaitre et l'impossibilité de les décomposer dans des sous-parties temporelles nous amène logiquement vers les GMM qui ne tiennent pas compte de la structure temporelle des signaux à reconnaitre. Les modèles GMM nécessite une phase d'apprentissage pendant laquelle les paramètres des distributions de Gauss sont évalués. Dans la phase de test, un calcul de vraisemblance d'appartenance de chaque vecteur à chaque classe est effectué.

La première étude a visé le choix des paramètres acoustiques les mieux adaptés par rapport aux signaux à reconnaitre ; cette étude exploratoire a utilisé le critère BIC [Schwarz, 1978], [Roeder and Wasserman, 1997] pour évaluer la pertinence de chaque paramètre. Des paramètres classiques comme : MFCC, LFCC, LPCC, mais aussi des paramètres comme le nombre de passages par zéro (ZCR) ou le barycentre spectrale ont été évalués. Un nouveau type de paramètres acoustiques basés sur la transformée en ondelettes ont été proposés ; cela se sont avérés plus résistants au bruit [RI2]. Ces paramètres appliquent le principe des MFFC sur les différents coefficients de la transformée en ondelettes.

Une nouvelle étude portant sur les paramètres acoustiques en se basant sur l'analyse temporelle et spectrale des signaux est en cours dans le cadre de la thèse de Mohamed Amine Sehili que je co-encadre. Des paramètres en se basant sur l'histogramme des

ensembles, EIH (Ensemble Interval Histogram) [Ghitza, 1994], ont montré une amélioration des performances dans les cas d'un nombre réduit d'échantillons sonores ; ces paramètres ont été étudiés dans le cadre du stage postdoctoral de Jugurta Montalvao [CI4]. EIH utilise une matrice de détecteurs de seuil attachés aux sorties de filtres passe-bande pour générer un histogramme d'inter-détection. EIH est inspiré des systèmes auditifs des animaux et leur mise en œuvre est faite de différentes manières.

Une amélioration de l'algorithme de classification des sons a été développée dans le cadre du stage postdoctoral de Jamal Eddine Rougui [CI8]. Cette approche optimise le nombre de distributions de Gauss pour chaque classe de sons tenant compte du nombre différent d'échantillons sonores disponibles. Le critère utilisé pour l'optimisation est le MDL (Minimum Length Description) [Rissanen, 1989].

Une deuxième approche de reconnaissance basée sur les machines à vecteurs de support (SVM) est en cours d'évaluation [RN3]. Cette méthode de classification, un contre un, se base sur la possibilité de trouver un hyperplan permettant de séparer les deux classes (éventuellement après une transformation vectorielle). Le noyau Gaussien est le choix naturel tenant compte des résultats obtenus avec les GMM. Les premiers résultats de classification sont moins bons que ceux obtenus avec les GMM mais une phase d'optimisation est en cours [CI1].

La **Reconnaissance des expressions de détresses** n'a pas constitué le centre de mes recherches en sachant que les recherches sur la reconnaissance de la parole ou des mots clés sont anciennes. Un module état de l'art basé sur HTK, a été mis en place par Bogdan Ica, co-encadré par moi-même. Ce module permet la classification en français d'un nombre prédéfini des expressions de détresses.

Une collaboration actuelle avec TPT, dans le cadre du projet CompanionAble, nous a permis la participation à l'utilisation d'un moteur de reconnaissance de la parole (Julius) pour la reconnaissance des ordres domotiques et des expressions de détresses en néerlandais et français.

Une attention toute particulière a été portée à la mise en œuvre temps réel pour une évaluation réelle du système. Le système de reconnaissance des sons (ANASON) a été initialement conçu sur un PC et dans un second temps sur un PC embarqué. Les algorithmes ont été optimisés pour permettre le fonctionnement en temps réel simultanément de 8 canaux et avec possibilité de distribution sur plusieurs machines.

La collaboration avec le constructeur de microphones AKG, partenaire du projet CompanionAble, nous amené à évaluer une possible mise en œuvre embarquée sur DSP. Les premières études montrent qu'une partie des modules du système pourrait être mis en place dans le système embarqué qui inclut le microphone. Un projet a été soumis ayant en partie comme tâche la réalisation d'un réseau de microphones avec de l'intelligence distribuée. Sur la même thématique, une étude amont a été effectué par Fréderic Fauberteau sur la possibilité de distribuer les calculs sur plusieurs processeurs en utilisant les temps creux [CI11].

IV.3. La fusion de données multimodales

L'utilisation de l'environnement sonore comme source d'information pour une application de télévigilance médicale est riche mais très dépendante des conditions sonores (présence de bruit, télévision allumé, plusieurs personnes,…) ce qui nous amène à chercher à combiner cette information avec ceux issues d'autres capteurs pour améliorer la robustesse et la flexibilité du système. La robustesse pourrait être améliorée par le rajout de capteurs redondants et la flexibilité par la conception d'un système modulaire capable de fonctionner avec tous ou seulement une partie des capteurs. Les travaux de fusion de données multimodales ont été accomplis à travers la thèse et le stage postdoctoral de Hamid Medjahed et la thèse de doctorat de Paulo Cavalcante. Les résultats présentés sont basés sur la combinaison de capteurs infrarouges et ambiants (développé par INSERM U558 Toulouse), d'un capteur mobile sur la personne (développé par TSP) et le système d'analyse sonore.

Parmi les méthodes de fusion de données nous avons identifiés 2 grandes familles :

- Basés sur les méthodes probabilistes (comme les réseaux Bayésiennes [Cowell et al., 1999] et le raisonnement de décision géométrique comme la distance de Mahalanobis). Ces méthodes ont des performances limitées quand les données sont hétérogènes et insuffisantes pour une modélisation statistique correcte.
- Basés sur des modèles connexionnistes (comme les réseaux de neurones MLP [Dreyfus et al., 2002] et les SVM [Burges et al., 1998]) qui sont très puissantes parce qu'elles peuvent modéliser des données fortement non linéaires mais elles ont des architectures complexes.

En fonction du niveau auquel la fusion de données est réalisée nous avons 3 catégories (Figure 3) :

- **Fusion directe au niveau des sorties des capteurs** ; les données brutes des capteurs sont fusionnées.
- **Fusion au niveau des paramètres** ; cette combinaison permet aux algorithmes de classification de bénéficier en entrée d'un vecteur plus grand.
- **Fusion au niveau des scores de sortie des étages de classification**. Sur chaque capteur toute la chaine du traitement au module de classification est effectuée. La fusion permet d'obtenir une décision plus fiable en se basant sur les scores des sorties mais aussi sur leurs confiances.

Pour l'application de télévigilance étudiée nous avons combiné les trois approches : fusion des données brutes pour les capteurs d'ambiance, fusion de paramètres pour le capteur mobile et fusion de scores pour le système d'analyse sonore.

Dans notre cas, tenant compte du fait que les bases de données sont inexistantes et difficile à enregistrer (difficile à simuler une chute) et parce que les données à traiter sont hybrides nous avons fait le choix, dans une première étape, d'utiliser un système expert basé sur la logique floue.

La logique floue a été introduite par Lotfi A. Zadeh en 1965 [Zadeh L.A., 1978] et représente une extension de la logique classique. Historiquement, elle est liée par rapport au concept de mesure floue introduit par Sugeno [Sugeno M. , 1974]. Des tentatives similaires ont été faites en même temps par Shafer (la théorie de l'évidence

[Shafer, 1979]) et Shackle (la théorie de la surprise [Shackle, 1961]) mais la logique floue a été beaucoup plus étudiée et beaucoup d'applications ont été développées.

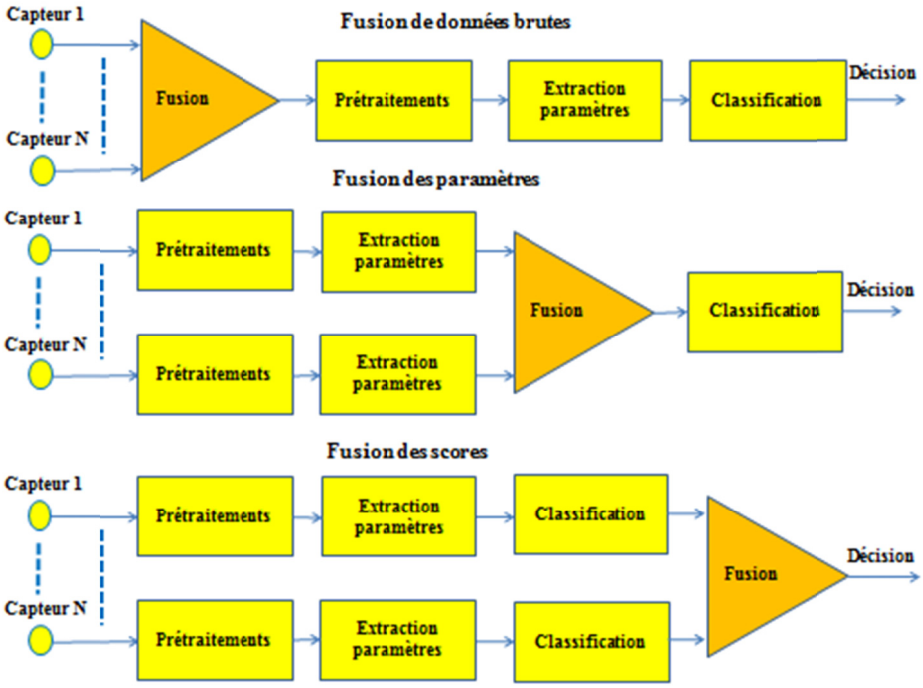


Figure 3 - Niveaux de la fusion de données

L'utilisation de la logique floue peut être faite à 2 niveaux :

- **Représentation de paramètres** : Dans les cas où les données sont incomplètes ou bruités, ou certaines paramètres sont difficiles à mesurer avec précision l'utilisation de la logique floue est naturelle et bien adapté. La représentation floue se base sur des adjectives et adverbes du langage comme : large, petit, moyen, … (exemple [Mandal et al. , 1992]).
- **Représentation des classes** : Si les données d'une classe ne forment pas une partition bien identifiée, une représentation floue qui admet les recouvrements entre classes et mieux adaptée. La plupart des méthodes de reconnaissance basées sur la logique floue sont des extensions des méthodes classiques utilisant les partitions floues comme la méthode c-means [Pedrycz, 1990]. Le partitionnement flou a été introduit par Ruspini en 1969 [Ruspini, 1969].

Les raisons du choix de la logique floue dans le cadre de la plateforme de télévigilance médicale sont :

- Les données à fusionner sont de différentes natures (analogiques, binaires) avec des modes de mise à jour aussi différents : synchrone et périodiques (capteur mobile et capteurs infrarouges) ou asynchrone (le son).
- Enregistrer des bases de données contenant des situations de détresse et très difficile donc l'utilisation des méthodes basés sur l'expérience s'impose. Les données utilisées dans la prise de décision sont imprécises et imparfaits.

La logique floue a été utilisée dans l'aide au diagnostic [Adlassnig et al., 1986], le contrôle des systèmes [Mason et al., 1997], le traitement d'image [Lalande et al., 1997] et la reconnaissance des formes [Zahlmann et al., 1997].

Les étapes d'un système à base de logique floue sont :

1. La « **fuzification** » ou la transformation des variables d'entrées du domaine classique dans des variables de type floues.
2. **L'inférence** de règles appliquées à ces données.
3. La « **défuzification** » ou la transformation en sens inverse, depuis une sortie floue vers une sortie logique classique.

L'étape de « fuzification » consiste dans la transformation d'une variable dans une variable de type floue et son coefficient d'appartenance à travers une fonction d'appartenance [RN4]. En fonction du domaine de variation de la variable d'entrée différents classe floues définis par différentes fonctions d'appartenance peuvent être choisies. Les fonctions d'appartenance les plus utilisées sont : la fonction triangulaire, la fonction trapézoïdale, la Gaussienne, le singleton, sigmoïde, etc.

Dans notre cas, par exemple, la sortie du système d'analyse sonore est représentée par 4 classes d'appartenance : *absence de signal* (singleton), *la classe normale* (trapézoïde, des sons de la vie courante), *la classe des alarmes possibles* (trapézoïde, sons qui pourront indiquer en certaines situations une détresse : bris de verre,…) et *la classe d'alarme* (trapézoïde, sons anormaux : cris, chute d'objets, expressions de détresse) [CI9].

L'étape d'inférence applique les règles de la logique floue aux données d'entrée et réalise une agrégation des résultats. Les règles se basent sur des opérateurs de la logique classique (OU, ET, inverseur) qui ont été redéfinis pour l'application floue. Ces opérateurs s'appliquent sur les coefficients d'appartenance (par exemple le OU devient une somme des coefficients d'appartenance).

Les règles d'inférence sont du type « *SI variable appartient à … ALORS sortie appartient à …*» et ont un poids associé. Il y a plusieurs possibilités pour inférer ces règles, nous pouvons citer les règles de Mamdani et celles de Takagi/Sugeno. Pour une règle de Mamdani le résultat est toujours une classe floue qui est obtenue à travers le produit algébrique et le maximum [Jang et al., 1997]. Dans le modèle Takagi/Sugeno la sortie est la somme pondérée des résultats.

La dernière étape est celle de « défuzzification » qui permet de transformer les résultats de sortie de forme floue dans une valeur classique utilisable pour réaliser une action. Parmi les transformations possibles nous retrouvons : le barycentre de la surface (COA), la bissectrice de la surface (BOA), la distribution de Gauss, la moyenne des maximums (MOM), le plus petit des maximums (SOM), le plus grand maximum (LOM), etc.

Le système de fusion proposé contient deux sorties : une d'alarme et une autre de localisation. Pour la sortie d'alarme nous avons choisie d'utiliser une gaussienne qui nous permet d'obtenir aussi une confiance de notre décision et pour la sortie de localisation des trapèzes.

La sortie d'alarme agrège la sortie du système analyse sonore, celle des capteurs infrarouges qui permettent d'avoir une indication de position couché et celle du capteur mobile. La sortie de localisation utilise principalement les capteurs de présence mais aussi le rapport signal sur bruit des microphones (RSB).

Ce système de fusion de données à base de logique floue a été évalué premièrement sur des scénarios simulés et dans une deuxième étape sur des enregistrements de laboratoire. Les paramètres mesurés ont été :

- **Sensibilité** : le nombre de vrais positifs divisés par la somme des vrais positifs avec les faux négatifs.
- **Spécificité** : le nombre de vrais négatifs divisés par la somme de vrais négatifs et faux positifs.
- **Le taux d'erreur** : le rapport entre le nombre de mauvaises classification et le nombre total d'échantillons.
- **Le taux de bonne classification** : le rapport entre le nombre de bonnes classifications et le nombre total d'échantillons.

La première évaluation a été réalisée sur 100 séquences simulées composé de 70 situations de détresse et 30 situations normales. Nous avons obtenu une sensibilité de 97% et une spécificité de 96%. Le taux d'erreur a été de 3%. Le système a été composé de 10 règles floues.

La deuxième évaluation du système a été réalisé sur 20 scenarios enregistrés dans notre laboratoire parmi lesquels 10 représentent des situations de détresse. Chaque scénario a une durée de 10 minutes. Le taux d'erreur a été de 5%.

L'information issue de ces trois modalités pourrait aussi servir à détecter le type d'activité de la personne et en réalisation une analyse à long terme pouvoir prévoir l'augmentation du risque d'apparition d'une situation anormale. L'analyse des rythmes circadiens de la personne a été étudié et à montrer des bonnes résultats pour les personnes âgées [Virone et al., 2008]. L'originalité de notre approche consiste dans la fusion de plusieurs modalités pour obtenir avec plus de précision les activités de la personne et aussi avoir une capacité à reconnaitre un nombre important d'activités surtout par l'intermédiaire du son. Les premières investigations réalisées dans le cadre du stage postdoctoral de Hamid Medjahed ont montré des résultats encourageants [O1], [CI3]. Les travaux de recherche ont été concentrés dans cette première approche sur l'identification des activités.

Une autre approche basée sur les réseaux d'évidence est en cours d'évaluation dans le cadre de la thèse de Paulo Cavalcante. Un réseau d'évidence a été implémenté et évalué pour la détection de chute en utilisant les trois modalités déjà indiqués. Les réseaux d'évidence se base sur la théorie de Dempster Shafer [Hong et al., 2009] qui est une généralisation de la théorie des probabilités permettant de quantifier l'incertitude. A la place d'utilisé une probabilité d'incertitude c'est un intervalle qui est utilisé. Les limites de cet intervalle sont la *croyance* et la *plausibilité*. Une fonction de masse a été introduite pour représenter la distribution de la croyance pour une grandeur donnée. En caractérisant chaque capteur à travers sa croyance et sa plausibilité et en transmettant ces propriétés à travers un réseau, la sortie de la fusion peut ainsi être déterminée. Les principales opérations au niveau d'un réseau d'évidence sont : la translation et la propagation.

Un réseau d'évidence a été construit pour la détection de chute en se basant sur la sortie du capteur mobile, des capteurs infrarouges et du système d'analyse sonore [O2], [CsA1]. Le réseau a été évalué sur une base de données enregistrée dans notre laboratoire contenant 5 situations normales et 33 chutes parmi lesquels 17 chutes molles

(sans accélération importante). Le taux d'erreur a été de 5% expliqué par le non détection de chutes molles sur les 17. Il faudra tenir compte que les chutes molles ne sont pas détectées par le capteur mobile et donc c'est la fusion des modalités qui a permis la détection de 15 chutes molles sur 17.

A travers les différents stages et la thèse de doctorat qui vient de commencer de Toufik Guettari, la domotique a été approchée comme application secondaire des capteurs de détresse. L'information issue des capteurs choisis pour détecter principalement une situation de détresse peut être aussi utilisée pour apporter du confort à l'utilisateur à travers des commandes domotiques. La parole et le son constituent des interfaces très faciles à utiliser mais dépendantes de la qualité du signal. La combinaison d'autres capteurs pourrait améliorer les performances des systèmes d'analyse sonore et de la parole (la localisation de la personne indique le canal sonore à analyser en priorité). Ces recherches sont menées en étroite collaboration avec la société Legrand.

IV.4. La segmentation et la reconnaissance du locuteur

Les applications de télévigilance médicale qui représentent le point central de mes activités de recherche permettent outre l'identification d'une situation de détresse d'apporter du confort à travers la domotique. L'interaction homme-machine la plus facile est représentée par la parole et les gestes. Je ne me suis pas intéressé à la reconnaissance de la parole elle-même mais aux techniques qui permettent d'identifier les personnes qui parlent et de segmenter le flux de parole. La personne âgée visée par l'application vit seule mais elle pourrait recevoir des invités (la famille ou les intervenants socio-médicaux) et dans ce cas le système de surveillance sonore devra être arrêté si nous ne pouvons pas reconnaitre la voie de la personne à surveiller.

Pendant mon stage postdoctoral au Laboratoire d'Informatique d'Avignon je me suis intéressé aux prétraitements qui pourraient permettre, en utilisant plusieurs canaux simultanés dans l'enregistrement des réunions, à faciliter la segmentation en locuteurs. La technique de segmentation en locuteurs testée a été celle développée dans le cadre du laboratoire et basée sur des modèles de Markov cachés (HMM) évolutifs (nombre d'état évolutif) [Meignier S. et al., 2000]. Cette méthode permet par plusieurs itérations d'évaluer à la fois le nombre de locuteurs et de trouver les frontières de changement de locuteur. La mise en œuvre de l'algorithme a utilisé la librairie ALIZE développée par le LIA [Bonastre et al., 2005].

Une méthode qui évalue en continu le rapport signal sur bruit (RSB) de chaque canal et qui l'utilise une somme pondérée de tous les canaux disponibles a été proposée [CI23]. Le poids correspondant à chaque canal évolue dans le temps en concordance avec le RSB du canal. L'estimation du RSB a été réalisée par deux méthodes : l'utilisation d'un détecteur de parole permettant de séparer le bruit de la parole ou en utilisant une évaluation du spectre du bruit par suivi du minimum statistique [Hirsh, 1993], [Cui et al., 2003]. La méthode proposée a été évaluée dans le cadre de la campagne d'évaluation NIST 2005. La technique de prétraitement a permis une amélioration du RSB des enregistrements et a été comparée avec les résultats de segmentation obtenus sur une simple somme des signaux. Une amélioration de 42% a été observée [CI24].

La reconnaissance du locuteur, principalement pour le contrôle d'accès ou démarrage/arrêt du système a été étudiée à travers les stages de M. Mohamed Chenafa et N.F. Iancu. Pour améliorer les performances des systèmes de reconnaissance du locuteur, une possibilité est de fusionner les diverses informations portées par le signal de parole. Plusieurs études sur la fusion de l'information ont été menées pour améliorer les performances de la RAL [Higgins et al., 2001], [Mami, 2003], [Kinnunen et al., 2004].

Nous avons proposé une nouvelle approche de fusion en utilisant deux types d'informations contenues dans le signal de parole : le locuteur (qui parle ?) et les mots prononcés (ce qui a été dit ?) [O5]. L'objectif de la méthode proposée est d'identifier un couple Locuteur/Mot clé correspondant à un premier signal de test. Cette étape est réalisée en combinant deux systèmes d'identification basés sur le rapport de vraisemblance ; un système d'identification des locuteurs (indépendant du texte) et un système d'identification des mots (indépendant des locuteurs). L'identité du locuteur obtenue par la fusion des deux systèmes d'identification est par la suite vérifiée par un système classique utilisant un deuxième signal de test pour confirmer ou infirmer le locuteur identifié. Dans la pratique, les deux signaux de test peuvent être considérés comme un seul signal segmenté en deux parties automatiquement. La mise en œuvre du système a aussi été réalisée en utilisant la librairie ALIZE [Bonastre et al., 2005].

La reconnaissance du locuteur est basée essentiellement sur des modèles GMM qui permettent de modéliser les caractéristiques de la voix d'une personne en mode dépendant du texte (l'apprentissage et le test sont effectués sur le même texte) ou indépendant du texte (le test s'effectue avec des mots/phrases complètement différents de ceux prononcés durant la phase d'apprentissage). Comme on ne dispose pas pour chaque locuteur de beaucoup d'enregistrements, la création du modèle de chaque locuteur est réalisée en deux étapes : création d'un modèle du monde ou universel en utilisant des enregistrements d'une multitude de locuteurs et l'adaptation de celui-là aux données d'un locuteur en particulier (Figure 4).

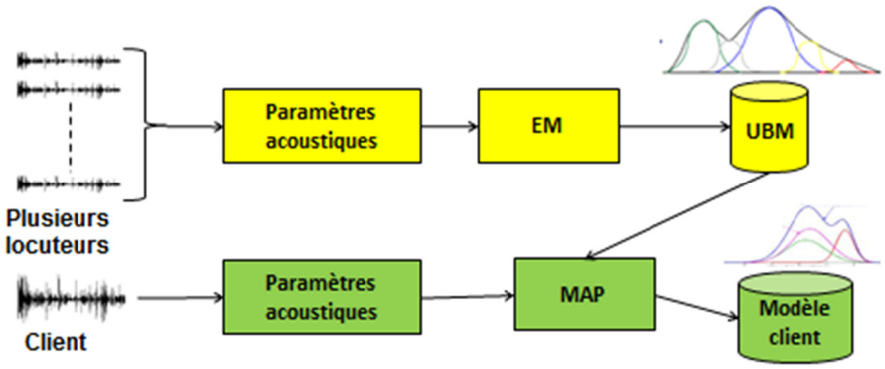


Figure 4 - Apprentissage du modèle du client

L'évaluation d'un système de reconnaissance du locuteur se fait en termes de Taux de Fausses Acceptations, Taux de Fausses Alarmes et Taux d'Egal Erreur. Le Taux d'égale erreur se détermine comme étant la première bissectrice de la courbe ROC (Figure 5).

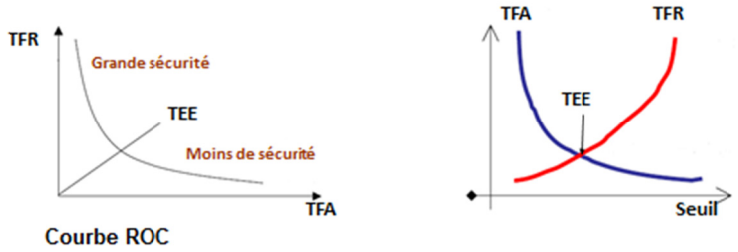


Figure 5 - Courbe ROC

Un système permettant de fusionner la reconnaissance du locuteur avec la reconnaissance de trois mots clés reliés à chaque locuteur a été proposé. La méthode, dans une première étape, combine un système d'identification du locuteur avec un système de reconnaissance des mots isolés. Une fois un locuteur identifié, dans une deuxième étape, on vérifie son identité en utilisant la deuxième partie du texte prononcé. Le système est composé des modules suivants :

- **Segmentation automatique des mots** : extraction automatique des mots.
- **Reconnaissance des mots** : identification des mots clés.
- **Identification du locuteur** : identification du locuteur.
- **Fusion de décision** : choix du locuteur identifié parmi les 3 premiers.
- **Vérification du locuteur** : vérification du locuteur identifié par l'étape précédente.

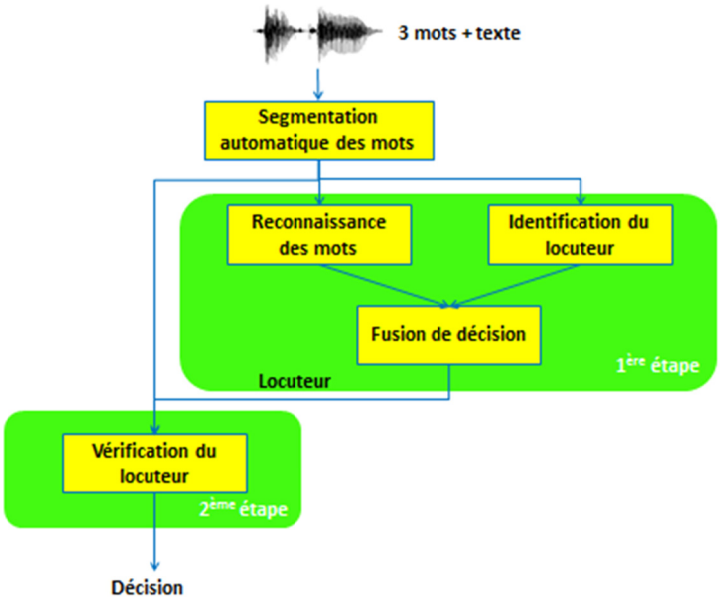


Figure 6 - Système de reconnaissance du locuteur proposé

Le module de **segmentation automatique** est celui de détection utilisé dans la reconnaissance des sons, mais réglé de telle manière qu'il segmente les mots d'une phrase.

Le module d'**identification du locuteur** calcule la vraisemblance entre le premier signal de test (composé de 3 mots) et les modèles des locuteurs clients stockés dans la base de données. Le module est basé sur GMM et il est indépendant du texte. Aucune décision

n'est prise à ce niveau, mais les vraisemblances du signal inconnu par rapport à tous les modèles des clients obtenus sont triées et sauvegardées pour la phase de fusion.

Le même signal, constitué de trois mots, est également utilisé par le module **de reconnaissance des mots**, indépendant du locuteur. Ce système correspond à la fusion de trois sous modules d'identification, un pour chaque mot du premier signal. Les résultats de chaque sous module sont utilisés afin de proposer une ou plusieurs combinaisons de mots de passe possibles. Seuls les cinq premiers résultats de chaque module sont combinés. Plusieurs tests ont été effectués pour déterminer le nombre optimum de mots à utiliser dans la fusion [CI17].

La **fusion de décision** consiste dans le tri des vraisemblances du signal inconnu (composé de 3 mots clés) calculées par rapport aux modèles des locuteurs clients et par rapport aux modèles des mots. Un premier test consiste à comparer le locuteur le plus probable avec les cinq premiers mots clés identifiés. Si son mot de passe est trouvé alors un couple Locuteur/Mot clé correspondant au premier signal de test est identifié. Un second test est effectué en comparant le mot de passe le plus probable avec les cinq premiers locuteurs identifiés, s'il s'agit d'une série de mots clés de l'un d'entre eux, alors un deuxième couple Locuteur/Mot clé est identifié. Dans le cas où deux couples sont identifiés, le couple avec la plus grande vraisemblance (somme des vraisemblances locuteur et mots clés) est retenu. Le système peut rejeter directement un enregistrement s'il n'y a pas de couples identifiés.

Une fois le premier signal de test associé à un locuteur, une **vérification du locuteur** est alors lancée sur le résultat de la fusion. Si la vraisemblance obtenue est inférieure à un seuil, l'identification est confirmée, sinon c'est un rejet et le système considère la personne comme un imposteur. La valeur du seuil est à choisir en fonction du compromis recherché : réduire le nombre de faux rejets ou le nombre de fausses acceptations.

Le système proposé a été évalué sur une base de données enregistrée contenant 58 locuteurs (4.3 h d'enregistrement) qui ont prononcé 20 mots clés, leur nom et leur prénom 10 fois chacun. Les paramètres LFCC ont été utilisés pour l'identification et la vérification du locuteur et les MFCC pour la reconnaissance des mots.

Le système proposé améliore les résultats en termes de taux d'égale erreur par rapport à un système d'identification de l'état de l'art dépendant du texte. Le système de référence sur notre base de données est à 4.76% de taux d'égal erreur et le système proposé à 0.13% en absence de bruit.

Le système a été aussi testé en présence d'un bruit de ventilateur (bruit stationnaire) et en présence d'un bruit de pas (impulsionnel). Pour le bruit de ventilateur le taux de faux rejets reste à 10 % même à un RSB de 10 dB avec un taux de fausses acceptations de 2%. La segmentation automatique n'a pas eu d'erreur parce que l'algorithme a été optimisé pour des bruits stationnaires. Dans le cas du bruit de pas la segmentation introduit 3.5% d'erreur à 20 dB. Dans ce cas le système a un taux de faux rejet de 17% et un taux de fausses acceptations de 1% à 40 dB [RN2].

IV.5. Les projets de recherche

Dans cette section je vais décrire les principaux projets de recherche auxquels j'ai participé et ceux en cours.

IV.5.1. Projet CompanionAble

Le projet CompanionAble (Integrated Cognitive Assistive & Domotic Companion Robotic Systems for Ability & Security) est de type projet Européen FP7 et il se déroule sur 4 ans et demi à partir de janvier 2008.

Les partenaires du projet sont : The University of Reading (Responsable), Technische Universitat Ilmenau, Assistance Publique Hopitaux de Paris, Groupe des Ecoles des Telecommunications, Fundacion Robotiker, Austrian Research Centers GmbH, Legrand France S.A., AKG Acoustics GmbH, ESIEE, ESIGETEL, Université d'Evry-val d'Essone, MetraLabs GmbH Neue Technologien und Systeme, Stichting Smart Homes, Center for Usability Research and Engineering, Universidad da Coruna, Innovation Centre in Housing for Adapted Movement, Bioingenieria Aragonesa, Fundacion Instituto Gerontologico Matia, Verklizan.

L'aide aux personnes âgées est largement reconnue comme impérative, pour vivre à domicile de façon autonome aussi longtemps que possible. Sans l'appui de la stimulation cognitive, l'état des personnes âgées souffrant de la maladie d'Alzheimer débutante ou d'autres maladies cognitives peut se détériorer rapidement et d'agrandir aussi le risque d'exclusion sociale.

CompanionAble fournira la synergie de la robotique avec l'intelligence ambiante et leur intégration sémantique pour fournir l'assistance aux donneurs de soins. Cela permettra de soutenir la stimulation cognitive et la thérapie de gestion du bénéficiaire de soins. C'est par la médiation d'une assistance automatique (robotique et intelligence ambiante) travaillant en collaboration avec le milieu familial que CompanionAble aidera les personnes âgées atteintes de troubles cognitifs.

CompanionAble aborde les questions de l'inclusion sociale et de soins à domicile des personnes souffrant de troubles cognitifs chroniques et plus particulièrement chez les personnes âgées. La méthodologie est basée sur la co-conception et le scénario de validation; les bénéficiaires de soins et leurs proches aidants ainsi que l'ensemble des parties prenantes participeront à la définition des fonctionnalités. Le but est d'assurer de bout en bout la viabilité, la souplesse, la modularité et l'accessibilité ainsi qu'un focus sur l'ensemble des soins, de soutenir l'intégration et de répondre à des questions telles que la préservation de la dignité et de la vie privée ainsi que la sécurité.

CompanionAble évaluera sa solution sur un certain nombre de bancs d'essai européens représentant un utilisateur de base. La collaboration de grands gérontologues, spécialistes des soins aux personnes âgées, institutions, industriels et académiques, y compris la robotique cognitive et l'intelligence ambiante représente un excellent garant de l'expertise de ce projet novateur.

Dans le cadre de ce projet, je suis responsable de la tâche T416 sur la reconnaissance des sons de la vie courante, des expressions de détresses et des commandes vocales qui

a comme partenaires outre l'ESIGETEL, l'Institut Télécom et l'Université de Reading. Mes activités de recherche ont concerné la reconnaissance des sons en utilisant le microphone de type CMT développé pour ce projet par AKG, la communication et l'interaction avec la reconnaissance de la parole, la localisation de la personne à travers des capteurs domotiques mais aussi le son et la vidéo. A travers la thèse Paulo Cavalcante, j'ai aussi participé sur la fusion de données.

IV.5.2. Projet QuoVADis

Le projet QuoVADis est un projet ANR TecSan 2007 qui s'est déroulé de janvier 2008 à juin 2011.

Les Partenaires sont : IBISC (Responsable), Télécom SudParis, APHP (Hôpital Broca), SAMU 92, INSERM U558, ESIEE, ESIGETEL, Robosoft, ASICA

Le projet QuoVADis répond au besoin de compenser les difficultés de communication dues aux pertes dans les capacités cognitives qui génèrent l'isolement social, la dépression, l'insécurité et l'inconfort dans la vie quotidienne. Le système vise d'une part à rétablir le lien affectif avec les proches, aidants et soignants par un système mobile interactif accompagnant la personne en difficulté, et d'autre part à lui permettre de se repérer dans son environnement et de le contrôler. Il a pour but de faciliter la prise en charge des pathologies cognitives (Maladie d'Alzheimer ou apparentées) et d'alléger le fardeau des aidants. L'intérêt de la mobilité introduite par le robot est précisément un accompagnement constant accepté, souhaité et valorisé sur le plan affectif et sécuritaire. Le maintien à domicile de personnes présentant des troubles cognitifs est une alternative à l'hospitalisation qui répond à la tendance actuelle de réduction du nombre de lits hospitaliers. Le projet QuoVADis a pour objectifs de répondre à deux des problèmes posés par le maintien à domicile : la stimulation cognitive et la sécurité médicale des personnes.

Stimulation cognitive du patient. Pour les personnes atteintes de troubles cognitifs (Alzheimer par exemple), les enjeux sont de pouvoir stimuler les capacités cognitives résiduelles en aidant le patient à se repérer dans le temps et l'espace, en le sécurisant en cas d'errance, de désorientation et d'angoisse, et en facilitant sa communication avec l'entourage. L'apport principal visé est l'introduction novatrice dans le domaine de la stimulation cognitive du concept de "robot compagnon" multimédia interactif et mobile permettant de stimuler le patient.

Sécurité du patient. Elle est basée sur un système de télévigilance de patient à domicile qui a pour objectif de permettre à des personnes dépendantes ou fragilisées de rester chez elles et d'être aidées ou "médicalisées" à distance de manière plus réactive en cas d'urgence. Les personnes concernées par cette fonction de sécurité sont les personnes dépendantes mais aussi des personnes atteintes de pathologies cardiaques ou des personnes en phase de convalescence post-clinique nécessitant une surveillance médicale rapprochée.

Dans le cadre de ce projet la plateforme EMUTEM (Environnement Multimodal pour la Télévigilance Médicale) a été développée par Hamid Medjahed. Cette plateforme intègre 3 modalités : le son, les capteurs infrarouges et le capteur mobile et réalise leur

fusion. Elle a été intégrée dans l'architecture globale du projet. La fusion de données est réalisée en utilisant de la logique floue.

IV.5.3. Projet INEASE-CAMED

Le projet INEASE-CAMED est un projet Brancusi (Collaboration France – Roumanie) qui s'est déroulé en 2009-2010.
Les partenaires du projet sont :

- France : ESIGETEL (Responsable), Télécom SudParis
- Roumanie : Université de Craiova (Responsable), Université de Médecine et Pharmacie de Craiova, Université Polytechnique de Bucarest

Le projet INEASE-CAMED (Environnement Intelligent pour l'assistance aux personnes âgées. Diagnostic médical assisté par ordinateur) vise à étudier et développer des systèmes informatiques permettant aux personnes âgées de bénéficier facilement des dispositifs domotiques par l'intermédiaire d'une interface homme-machine ergonomique et simplifiée et de vivre en sécurité chez elles en disposant d'un dispositif de télévigilance médicale. Le deuxième axe du projet concerne le diagnostic médical assisté par ordinateur à distance ou non.

L'espérance de vie en Europe est en constante évolution ce qui implique de plus en plus de personnes âgées qui vivent seules à la maison et qui ont risque élevés d'accident. Actuellement, il y a plusieurs recherches effectuées pour réaliser des systèmes de télévigilance médicale permettant de détecter automatiquement les situations de détresse et d'annoncer les urgences. La plupart des systèmes commercialisés actuellement proposent seulement un bouton portable permettant à la personne d'initier une communication vocale avec le centre de surveillance en cas de détresse. Le principal inconvénient d'un tel système c'est qu'il ne peut pas détecter une situation de détresse si la personne n'est pas en état d'appuyer sur le bouton. Le projet se propose de prolonger les recherches existantes pour proposer un système basé sur l'analyse de l'environnement sonore et aussi sur d'autres capteurs pour la détection automatique d'une situation de détresse. Plus particulièrement, le projet a étudié des méthodes de prétraitement du signal sonore pour une meilleure analyse par le système. Un deuxième axe de la télévigilance auquel le projet s'est intéressé est la détection fiable de chute et la mesure ambulatoire permanente du pouls de la saturation en oxygène. La mesure du pouls et de l'oxymétrie est difficile dans l'ambulatoire à cause des bruits de mouvement du patient et nécessite des techniques spécifiques.

La modélisation et la simulation des organismes vivants s'inscrit dans la tendance moderne d'intégration des connaissances, étant le résultat d'une collaboration interdisciplinaire. L'organisme vivant est un complexe de systèmes qui sont dans un équilibre dynamique en permanence contrôlé par des boucles de réglage. La modélisation mathématique permet de faire ressortir la structure interne et les liens causaux entre les composants. La simulation assure la validation des théories concurrentes, la compréhension des modifications physiologiques et suggère les expériences intéressantes. L'acquisition et le traitement des données présentent des problèmes d'ordre technique parce qu'une grande partie des paramètres du système soit ne sont pas accessibles soit ne peuvent pas être convertis en grandeurs mesurables. Les méthodes FDI basées sur la redondance analytique peuvent être adaptées et appliquées dans l'étude des systèmes physiologiques ; leur avantage consiste dans

l'identification de la structure interne du processus étudié utilisant le traitement avancé des signaux biologiques acquis. Dans le cadre du projet nous avons appliqué des méthodes de détection et localisation analytique des défauts de régulation des systèmes dynamiques régissant des systèmes physiologiques, le but étant d'élaborer et de tester des algorithmes de diagnostic automatique, non invasif, assisté par ordinateur.

Mes activités de recherche sur ce projet ont visé le développement du capteur sonore intelligent mais aussi une collaboration étroite avec la partie roumaine sur la prédiction de glycémie [CI13]. Une analyse temps-fréquence de type Wigner-Ville a été menée sur les données de glycémie enregistrées par l'Université de Médecine de Craiova (Roumanie). Cela nous a permis de pouvoir proposer un système basé sur les réseaux de neurones pour la prédiction de la glycémie [CN7]. Une collaboration sur l'étude des sons du cœur a aussi été initiée et elle est en cours.

IV.5.4. Projet Sweet-Home

Le projet SWEET-HOMe[3] est un projet ANR VERSO 2009 qui a comme but de concevoir un nouveau système de domotique basé sur les technologies sonores. Il est axé sur trois aspects :

- Fournir assistance à travers une interaction homme-machine naturelle (voix et tactile)
- Inclusion sociale
- Assurer la sécurité à travers la détection de situations de détresse.

La personne sera capable de piloter de n'importe quel position dans la maison son environnement, à n'importe quel moment du temps et de la manière la plus naturelle qui soit.

Les partenaires du projet sont : LIG (responsable du projet), ESIGETEL, THEORIS, Visage, Technosens.

Les utilisateurs ciblés sont les personnes âgées fragiles mais encore autonomes. Ce choix est motivé par deux raisons :

- Un système domotique a un coût et il serait plus profitable s'il est utilisé tout au long de la vie qu'uniquement durant la période de la perte d'autonomie.
- Dans le cas de la perte d'autonomie, la personne continuera à utiliser son système seulement avec quelques adaptations ce qui sera plus facile que d'avoir à s'adapter à un nouveau système domotique.

Pour déterminer l'acceptabilité d'un tel système une étude qualitative a été réalisée sur 8 personnes en bonne santé âgées de 71 ans à 88 ans, 7 proches et 3 professionnels [Portet et al., 2011]. L'étude a consisté dans un questionnaire mais aussi avec utilisation d'une maison entièrement équipé en domotique (magicien d'Oz pour la partie sonore). Les quatre aspects les plus importants du projet ont été évalués : la commande vocale, la communication avec l'extérieur, le système domotique et l'agenda électronique. Dans tous les cas, les solutions basées sur la voix ont été mieux acceptées que des solutions plus intrusives comme la vidéo. Pour le respect de l'intimité de la personne il faudra que la solution choisie traite l'information sonore à la volée et ne sauvegarde pas le signal

[3] *sweet-home.imag.fr*

sonore. Par ailleurs, le moteur de la reconnaissance de la parole reconnait seulement un nombre limité de phrases aussi pour préserver l'intimité de la personne.

Le projet veut utiliser des technologies et applications standardisés à la place de tout concevoir. Pour la domotique le standard KNX (KoNneX) qui est standardisé par l'ISO (ISO/IEC 14543) a été retenu. Pour l'inclusion sociale le projet utiliserait des systèmes existants comme e-lio ou Visage. Pour faciliter l'interaction avec l'utilisateur, le système aura trois voies : la commande vocale, tablette internet ou les boutons classiques.

Le système Sweet-Home sera composé, outre les capteurs domotiques communicants sur le bus KNX, de 8 microphones sans fils et un canal sonore venant de la tablette tactile [CI1]. Toutes les sorties de l'analyse sonore et des capteurs domotiques seront traités par un contrôleur intelligent qui selon le cas, enverra des commandes domotiques, initiera des communications avec les proches ou enverra une alarme. Le contrôleur intelligent utilise deux niveaux sémantiques : l'ontologie de niveau bas pour la représentation des capteurs et celle de haut niveau contenant les concepts pour le raisonnement. Cette séparation permettra une adaptation facile du système à chaque type de maison.

Les canaux sonores nécessitent un traitement spécifique pour extraire l'information sonore ou de parole. Le schéma de traitement est composé d'une détection du signal utile suivie d'une classification parole/son et par la suite d'un traitement spécifique (classification des sons ou reconnaissance de la parole).

Sur ce projet, je travaille à travers la thèse de Mohamed Amine Sehili sur la partie détection – classification parole/son – reconnaissance des sons. Nous comparons pour la partie classification l'approche classique GMM avec les SVM. Une participation importante au niveau de l'intégration des modules est également prévue.

V. Conclusions. Perspectives

Mes activités de recherche après l'obtention du diplôme de docteur de l'INPG m'ont permis de continuer dans le cadre du domaine de la télévigilance médicale en approfondissant l'analyse de l'environnement sonore mais aussi en l'élargissant à la fusion de données et à la reconnaissance du locuteur. L'analyse de l'environnement sonore et la fusion de données multimodales sont mes axes majeurs de recherche ; mais autour de ces axes, je me suis aussi orienté vers la reconnaissance du locuteur et la domotique. Ces deux axes originaux, le premier par l'approche de reconnaitre des sons considérés jusqu'à récemment comme bruits et le deuxième par l'application de la fusion à la thématique de la télévigilance médicale.

L'importance de l'extraction de l'information sonore est démontrée par l'intérêt porté à l'heure actuelle par d'autres laboratoires non seulement en France mais aussi à l'étranger (les Etats-Unis et Israël). L'utilisation du son pour l'analyse de l'environnement, dans le cadre des réunions, a été aussi étudiée dans le projet européen CHIL.

Dans le cadre de l'application de télévigilance médicale, les contraintes en termes de fiabilité des systèmes sont très importantes et nécessitent de combiner des capteurs redondants. Le coût et les spécificités des lieux de vie imposent la présence de systèmes adaptables et flexible et limitent le nombre et le choix du type de capteurs. L'absence des bases de données standardisées, comme c'est le cas dans le domaine de la reconnaissance de la parole, et la difficulté d'enregistrer de telles bases, notamment celles devant contenir des situations de détresse, nous a amené dans une première étape à utiliser des systèmes à base de règles comme la logique floue. La localisation de la personne en utilisant l'information existante est aussi une requête constante des services d'urgence pour simplifier leur démarche en cas d'intervention. L'identification des activités de la personne est un axe que j'ai commencé à explorer. Les premiers résultats obtenus démontrent le caractère prometteur de cette approche en termes de prédiction d'une situation anormale.

Tous mes travaux de recherche ont tous été conduits dans le cadre d'une collaboration avec des partenaires académiques et/ou industriels ou dans le cadre d'un projet de recherche (ANR, FP7, FEDER, PHC). L'encadrement des travaux de recherche de doctorants, post-doctorants et masters 2 m'a permis de développer une vision globale de la thématique de recherche, de transmettre le savoir-faire et de gérer les relations humaines. J'ai eu la chance d'encadrer une équipe de recherche (2EC, 2 doctorants et 2 post-doctorants) et depuis 2 ans de prendre la responsabilité du laboratoire de l'ESIGETEL (4 équipes – 10 enseignants-chercheurs). Ces responsabilités m'ont amené à mettre en œuvre mes aptitudes au management et à l'organisation de la recherche. Toutes ces expériences me motivent à présenter mes travaux en vue d'obtenir une HDR qui me permettra de continuer à développer et améliorer ma vision sur le domaine de recherche.

La suite de mes travaux de recherche est toujours envisagée dans le domaine du maintien à domicile des personnes et plus généralement dans les TICs pour la santé. Etant donné les enjeux sociétaux liés aux personnes âgées et leur autonomie, proposer des solutions techniques pour améliorer le confort et la sécurité demeure une priorité de la société, comme le confirment les derniers appels à projets aux niveaux national et

européen. L'analyse sonore apporte une information très riche en termes de contenu linguistique, d'analyse de scène et de localisation. Étant en même temps moins intrusive que l'image, elle se développera encore davantage avec l'apparition des systèmes embarqués, dont la capacité d'autonomie énergétique permet aujourd'hui de fonctionner plus longuement. Malgré cela, l'aspect énergétique demeure encore un frein à leur déploiement. Dans ce contexte, l'architecture que nous proposons, composée de deux étapes détection/classification, ambitionne une réduction de la consommation d'énergie et des ressources. Une première voie investiguée sera celle de l'adaptation et l'optimisation des algorithmes actuels pour le traitement du son sur une plateforme embarquée intégrant le microphone. Dans le même temps, des algorithmes de débruitage, séparation de source et traitement multicanal sont à développer et à adapter pour l'environnement spécifique de l'application de télévigilance.

Comme déjà mentionné, l'analyse du signal sonore apporte de l'information supplémentaire au robot, par rapport aux capteurs usuellement utilisés, tant pour contribuer à une meilleure localisation du robot et de la personne que pour faciliter l'interface homme-robot. Dans ce cadre, nous avons d'ores et déjà engagé des collaborations avec deux partenaires :

- La première avec le laboratoire IBISC explore l'intégration des capteurs sonores pour la localisation du robot dans son environnement. Ce travail couvrira tant les aspects méthodologiques que pratiques puisqu'il est envisagé d'implanter les capteurs sur le robot et d'évaluer la pertinence des approches proposées.

- La seconde avec la Technical University of Munich, s'inscrit dans le cadre d'un projet européen, déjà soumis et portant sur l'interaction homme-robot. Les objectifs du projet sont de développer des techniques adaptatives de classification, capables d'autoapprentissage. Il est aussi envisagé de concevoir et d'équiper une plateforme mobile d'un réseau de microphones distribués entre le robot et l'environnement. Signalons que l'écoute de l'environnement à travers ce réseau de microphones permettra aussi la localisation des sources.

D'un point de vue méthodologique, les perspectives sur la fusion de données pour la télévigilance médicale vont dans le sens de l'utilisation des approches de Dempster-Shaffer pour l'analyse de l'activité de la personne en vue d'une part d'une détection de rupture anormale dans les activités et d'autre part pour la prédiction, à court terme, d'une possible situation inquiétante. Une seconde voie est celle de l'utilisation de la fusion de données sur une seule modalité mais dans un objectif d'une meilleure robustesse : fusionner les sorties de plusieurs types de systèmes de classification de sons.

Les partenariats actuels s'enrichissent rapidement ce qui démontre l'intérêt croissant pour la problématique. Cela est confirmé par les nombreux appels à projets visant le maintien à domicile, la télémédecine et le confort à domicile (« ambient assisting living » - AAL). De plus, des plateformes nationales de télémedecine sont en cours d'installation en plusieurs régions de France ; je suis moi-même sollicité par plusieurs industriels pour du transfert technologique. De même, les hôpitaux et maisons de retraite médicalisées (EHPAD), ont montré leur intérêt pour le test et l'évaluation grandeur nature de nos solutions.

Bien que la télévigilance médicale demeure comme le thème central de nos recherches, une ouverture vers d'autres domaines d'application est déjà envisagée. On pourra citer, à ce titre, la sécurisation des biens et des lieux. Ainsi, des contacts ont été noués avec des industriels pour la sécurisation des entrepôts et la détection anticipée des intrusions.

Au niveau des responsabilités scientifiques, fort des expériences acquises, j'envisage d'assumer la coordination d'un projet de recherche, notamment au niveau européen. Cela me permettra d'une part d'avoir une plus grande maitrise du projet dans sa globalité et d'autre part de coordonner les travaux de plusieurs partenaires. Dans le même temps, la direction d'une plus grande équipe de recherche fait partie de mes perspectives. J'ambitionne aussi de conserver une grande implication au niveau des responsabilités collectives, tant administratives que pédagogiques.

VI. Annexe

VI.1. Références

[Adlassnig et al., 1986] Adlassnig K. P., Fuzzy set theory in medical diagnosis, IEEE Tr. On Syst.,Man, and Cybernetics, March/April 1986,pp. 260–265.

[Bairacharya et al. , 2008] A. Bairacharya, T.J. Gale, C.R. Stack et P. Turner, « 3.5G Based Mobile Remote Monitoring System », Proceedings of EMBC 2008, Vancouver, Canada, Août 2008, pp. 783-786, doi:10.1109/IEMBS.2008.4649269

[Bang et al. , 2008] S. Bang, M. Kim, S.K. Song et S.J. Park, « Toward real time detection of the basic living activity in home using a wearable sensor and smart home sensors », Proceedings of EMBC 2008, Vancouver, Canada, Août 2008, pp. 5200-5203, doi:10.1109/IEMBS.2008.4650386

[Bellego et al., 2006] G.L. Bellego, N. Noury, G. Virone, M. Mousseau et J. Demongeot, *Measurement and model of the activity of a patient in his hospital suite*. IEEE Transactions on TITB, Vol. 10, No. 1, 2006, pp. 92–99

[Binh et al., 2008] X.L. Binh, M. Mascolo, A. Gouin et N. Noury, « Health Smart Home for elders - A tool for automatic recognition of activities of daily living », Proceedings of EMBC 2008, pp. 3316-3319, Vancouver, Canada, August 2008, doi: 10.1109/IEMBS.2008.4649914

[Bonastre et al., 2005] J.-F. Bonastre, F. Wils and S. Meignier, « ALIZE, a free toolkit for speaker recognition », in Proceedings ICASSP '05, Philadelphia, PA, USA, pp. 737 - 740

[Burges et al., 1998] Burges C. J. C., A tutorial on SVM for Pattern Recognition. Data Mining and Knowledge Discovery, volume 2, 1998, pp. 121–167.

[Cowell et al., 1999] Cowell R., Dawid A., Lauritzen S. & Spiegelhalter D., Probabilistic Networks and Expert Systems, 1999, ISBN : 0-387-98767-3.

[Cowling et Sitte, 2002] M. Cowling et R. Sitte, *Analysis of speech recognition techniques for use in a nonspeech sound recognition system,* Digital Signal Processing for Communication Systems, Vol. 703, No. 1, pp. 31-46

[Cui et al., 2003] Cui, X., Bernard, A., Alwan, A.: A noise-robust asr back-end technique based on weighted viterbi recognition. In: Proceedings of Eurospeech 2003, Genova Switzerland (2003)

[Dreyfus et al., 2002] Dreyfus G., Martinez J.M, Samuelides M., Gordon M., Badran F., Thiria S. & Hrault L., Réseaux de neurones. Méthodologie et applications, Eyrolles, 2002.

[Fleury et al. , 2007] A. Fleury, N. Noury et N. Vuillerme, « A Fast Algorithm to Track Changes of Direction of a Person Using Magnetometers », Proceedings of IEEE EMBS 2007, Lyon, France, Août 2007, pp. 2311-2314, doi: 10.1109/IEMBS.2007.4352788

[Ghitza, 1994] Ghitza, O., « Auditory models and human performance in tasks related to speech coding and speech recognition », IEEE Trans. Speech Audio Process., vol.2, no.1, pp.115–132, 1994.

[Higgins et al., 2001] Higgins J. E., Damper R. I., Harris C. J., « Information fusion for subband-HMM speaker recognition », International Joint Conference on Neural Networks, 2001, p. 1504-1509.

[Hirsh, 1993] HIrsh, H.G.: Estimation of noise spectrum and its application to snr-estimation and speech enhancement. Technical report tr-93-012, ICSI, Berkeley, California, USA (1993)

[Hong et al., 2009] Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., Devlin, S.: Evidential fusion of sensor data for activity recognition in smart homes, Pervasive and Mobile Computing, Volume 5, Issue 3, Pervasive Health and Wellness Management, June 2009.

[Jang et al., 1997] Jang J.S.R., Sun C. T. & Mizutani E., Neuro-Fuzzy and Soft Computing :A Computational Approach to Learning and Machine Intelligence. Prentice Hall Upper Saddle River, NJ 1997.

[Kinnunen et al., 2004] Kinnunen T., Hautamäki V., Fränti P., « Fusion of Spectral Feature Sets for Accurate Speaker Identification », Proc. Intl. Conf. Speech and Computer, vol. 1, p. 361-365, 2004.

[Lacombe et al. , 2004] A. Lacombe, J.L. Baldinger, J. Boudy, B; Dorizzi, J.P. Levrey, R. Andreao, C. Perpere, F. Delavault, F. Rocaries et C. Dietrich, *Tele-surveillance System for Patient at Home: the MEDIVILLE system*, Lecture Notes in Computer Science, Springer-Verlag GmbH, Vol. 3118, June 2004, pp 400-407

[Lalande et al.,1997] Lalande A., Legrand L., Walker P. M., Jaulent M. C., Guy F., Cottin Y. & Brunotte F., Automatic detection of cardiac contours on MR images using fuzzy logic and dynamic programming, Proc. AMIA Ann. Fall Symp. 1997, pp. 474–478.

[Lima et Barbarosa, 2008] C.S. Lima et D. Barbosa, « Automatic segmentation of the second cardiac sound by using wavelets and hidden markov models », Proceedings of IEEE EMBC 2008, Vancouver, Canada, Août 2008, pp. 334–337

[Litvak et al. , 2008] D. Litvak, Y. Zigel et I. Gannot, « Fall detection of elderly through floor vibrations and sound », Proceedings of IEEE EMBC 2008, Vancouver, Canada, Août 2008, pp. 4632–4635

[Mami, 2003] Mami Y., Reconnaissance de locuteurs par localisation dans un espace de locuteur de référence, Thèse de doctorat, ENST Paris, France, 2003.

[Mandal et al. , 1992] Mandal D.P., Murthy C. A. & Pal S. K., "Formulation of a multivalued recognition system," IEEE Transactions on Systems, Man, and Cybernetics, 22:607–620 1992.

[Marschollek et al. , 2008] M. Marschollek, K.H. Wolf, M. Gietzelt, G. Nemitz, H. Meyer zu Schwabedissen et R. Haux, « Assessing elderly persons' fall risk using spectral analysis on accelerometric data - a clinical evaluation study », Proceedings of the EMBC 2008, Vancouver, Canada, Août 2008, pp. 3682-3685, doi:10.1109/IEMBS.2008.4650008

[Mason et al.,1997] D. Mason, D. Linkens & N. Edwards, « Self-learning fuzzy logic control in medicine », in Proc. AIME'97, (E. Keravnou et al., eds.), Lecture Notes in Artificial Intelligence 1211, Springer-Verlag, Berlin 1997, pp. 300–303.

[Meignier S. et al., 2000] S. Meignier, J.F. Bonastre, C. Fredouille, T. Merlin, « Evolutive HMMfor speaker tracking system », in Proceedings of ICASSP 2000, Istanbul, Turkey, pp. 1177–1180

[Moncrieff et al. , 2005] S. Moncrieff, S. Venkatesh, G. West et S. Greenhill, « Incorporating contextual audio for an actively anxious smart home », Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference, Melbourne, Australie, Décembre 2005, pp. 373-378, ISBN: 0-7803-9399-6

[Ng et Koh, 2008] A.K. Ng et T.S. Koh, « Using psychoacoustics of snoring sounds to screen for obstructive apnea », Proceedings of IEEE EMBC 2008, Vancouver, Canada, Août 2008, pp. 1647–1650

[Pedrycz, 1990] Pedrycz W., "Fuzzy sets in pattern recognition: methodology and methods," Pattern Recognition, 23(1/2):121-146, 1990.

[Popescu et al. , 2008] M. Popescu, Y. Li, M. Skubic et M. Rantz, « An acoustic fall detector system that uses sound height information to reduce the false alarm rate », Proceedings of IEEE EMBC 2008, Vancouver, Canada, Août 2008, pp. 4628–4631

[Portet et al., 2011] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly – acceptability and objection aspects," Personal and Ubiquitous Computing, in press.

[Rissanen, 1989] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, World Scientific, 1989.

[Roeder and Wasserman, 1997] Roeder, K. andWasserman, L. (1997). Practical bayesian density estimation using mixtures of normals. Journal of the American Statistical Association, 92 :894–902.

[Ruspini, 1969] Ruspini E. H., "A new approach to clustering," Inform, Control, 15(1):22-32, 1969.

[Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6 :461–464.

[Shackle, 1961] Shackle G.L., "Decision, Order and Time in Human Affairs," Cambridge Univ. Press, 1961

[Shafer, 1979] Shafer G., "A Mathematical Theory of Evidence," Princeton Univ. Press 1979.

[Stagera et al., 2007] M. Stagera, P. Lukowiczb et G. Trostera, *Power and accuracy tradeoffs in soundbased context recognition systems*. Pervasive and Mobile Computing, Vol. 3, No. 3, pp. 300–327, ISSN:1574-1192

[Sugeno M., 1974] Sugeno M., Theory of fuzzy integrals and its applications. Thèse de doctorat, Tokyo IT 1974.

[Virone et al., 2008] G. Virone, M. Alwan, S. Dalal, S. Kell, J. A. Stankovic, and R. Felder, "Behavioral Patterns of Older Adults in Assisted Living," IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 3, pp. 387-398, May 2008.

[Wolf et al. , 2008] P. Wolf, A. Schmidt et M. Klein, « SOPRANO - An extensible, open AAL platform for elderly people based on semantical contracts », Proceedings of 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence 2008 (AITAmI'08), Patras, Greece, pp. 225-228

[Zadeh L.A., 1978] Zadeh L.A., Fuzzy sets as a basis for theory of possibility, Fuzzy Set Systems. pp. 3–28, 1978.

[Zahlmann et al., 1997] Zahlmann G., Scherf M. & Wegner A., A neurofuzzy classifier for a knowledge-based glaucoma monitor, Proc. AIME'97, (E. Keravnou et al., eds.), Lecture Notes in Artificial Intelligence 1211, Springer-Verlag, Berlin 1997, pp. 273–284.

VI.2. Liste des publications et communications

La synthèse de mes publications est présentée dans le tableau ci-dessous :

Type	Nombre
Chapitres d'ouvrages	7
Articles de revues internationales avec comité de lecture	3
Articles de revues nationales avec comité de lecture	5
Conférences internationales avec actes et comité de lecture	38
Conférences nationales avec actes et comité de lecture	8
Communications industrielles	3
Communications sans actes	2
Rapports de recherche	17

VI.2.1. Participation à des ouvrages

[O1] H. Medjahed, **D.Istrate**, J. Boudy, J.L. Baldinger, B. Dorizzi, L. Bougueroua, M.A. Dhouib, A Fuzzy Logic Approach for Remote Healthcare Monitoring by Learning and Recognizing Human Activities of Daily Living, accepté dans le livre « Fuzzy Logic », ISBN 979-953-307-578-4

[O2] P.A. Cavalcante, J. Boudy, **D. Istrate**, H. Medjahed, B. Dorizzi, J. C. M. Mota, J.L. Baldinger, T. Guettari, I. Belfeki, Heterogeneous multi-sensor fusion based on an Evidential Network for fall detection, « LNCS 6719 Toward Useful Services for Elderly and People with Disabilities », DOI: 10.1007, ISBN 978-3-642-21535-3_42,pp. 281-285

[O3] **D. Istrate**, J. Boudy, H. Medjahed, J.L. Baldinger, Medical Remote Monitoring using sound environment analysis and wearable sensors, « Recent Advances in Biomedical Engineering », pp.517-532, Octobre 2009, ISBN 978-953-307-013-1.

[O4] **D. Istrate**, M. Vacher and J.-F. Serignat, Embedded Implementation of Distress Situation Identification Through Sound Analysis, « The Journal on Information Technology in Healthcare », 2008, 6(3), pp. 204-211.

[O5] M. Chenafa, **D. Istrate**, V. Vrabie and M. Herbin, Biometric System Based on Voice Recognition using Multiclassifiers, « Lecture Notes in Computer Science 5372 Biometrics and Identity Management », Springer, 2008, ISBN 978-3-540-89990-7, pp.206-215

[O6] M. Vacher, J.-F. Serignat, S. Chaillol, **D. Istrate** and V. Popescu, Speech and Sound Use in a Remote Monitoring System for Health Care, « Lecture Notes in Computer Science, Artificial Intelligence, Text Speech and Dialogue », vol. 4188/2006, 2006, pp. 711-718, ISBN: 978-3-540-39090-9

[O7] **D.Istrate**, E.Castelli, Multichannel Sound Acquisition with Stress Situations Determination for Medical Supervision in a Smart House, « Lecture Notes in Artificial Intelligence », Vol. 2166, 2001, ISBN 3-540-42557-8 Springer-Verlag, pp. 266-272.

VI.2.2. Articles de revues internationales avec comité de lecture

[RI1] **D. Istrate**, M. Vacher and J.-F. Serignat, Embedded Implementation of Distress Situation Identification Through Sound Analysis, The Journal on Information Technology in Healthcare, 2008, 6(3), pp. 204-211.

[RI2] **D. Istrate**, E. Castelli, M. Vacher, L. Besacier and J.-F. Serignat, Medical Telemonitoring System Based on Sound Detection and Classification, IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 2, Avril 2006. **Sélectionné et publié dans Yearbook of Medical Informatics 2007**.

[RI3] **D. Istrate**, H. Medjahed, L. Bougueroua, A. Dhouib, A. Amehraye, J. Boudy et J.-L. Baldinger, Live safely at home using a medical remote monitoring system, Official Journal of the Italian Society of Gerontology and Geriatrics, Vol.23, Suppl. To N.1 Feb 2011.

VI.2.3. Articles de revues nationales avec comité de lecture

[RN1] M. Vacher, **D. Istrate** and J.-F. Serignat, Probabilistic Models for Speech and Sound Analysis, Annals of the University of Craiova, Automation, Computers, Electronics and Mechatronics, vol. 4(31), no. 3, 2007, pp. 129-136

[RN2] **D. Istrate**, N. F. Iancu, M. Chenafa, V. Vrabie, "Fusion de decision pour contrôle d'accès par la voix », Revue Ingénierie des systèmes d'information, Vol.15, N° 2/2010, pages 11-27, ISBN 978-2-7462-2956-3

[RN3] M. A. Sehili, **D. Istrate**, J. Boudy, "Primary investigations of sound recognition for a domotic application using Support Vector", Annals of the University of Craiova, Series: Automation, Computers, Electronics and Mechatronics, Vol.7 (34), N°.2, 2010, pp.61-65, ISSN: 1841-0626

[RN4] H.Medjahed, **D.Istrate**, J.Boudy, F.Steenkeste, B.Dorizzi , "Elderly People Telemonitoring in an Integrated Smart House Environment", Annals of the University of Craiova, Series: Automation, Computers, Electronics and Mechatronics, Vol. 6 (33), N°2, 2009, pp.46-51, ISSN: 1841-0626.

[RN5] **D. Istrate**, M. Vacher, J.-F. Serignat, L. Beasacier, E. Castelli, Système de télésurveillance sonore pour la détection de situation de détresse, ITBM-RBM, Mai 2006, vol. 27, issue 2, pp. 35-45.

VI.2.4. Conférences internationales avec actes et comité de lecture

[CI1] M. Vacher, **D. Istrate**, F. Portet, T. Joubert, T. Chevalier, S. Smidtas, B. Meillon, B. Lecouteux, M. Sehili, P. Chahuara et S. Méniard, « The SWEET-HOME Project : Audio Technology in Smart Homes to improve Well-being and Reliance », Proceedings of IEEE EMBC 2011, 30 Aout – 3 Septembre 2011, Boston, Etats-Unis, pp.5291-5294

[CI2] M.A. Dhouib, L. Bougueroua, **D. Istrate**, M. Pino, C. Bernard, « HoCoS: Home Companion Software. A service oriented solution for elderly home accompanying and remote healthcare monitoring », Proceedings of IEEE EMBC 2011, 30 Aout – 3 Septembre 2011, Boston, Etats-Unis, pp.5343-5346

[CI3] H. Medjahed, **D. Istrate**, J. Boudy, J.L. Baldinger, B. Dorizzi, « A pervasive Multi-sensor Data Fusion for Smart Home Healthcare Monitoring », IEEE Conference on Fuzzy Systems 2011, 27-30 Juin 2011, Taipei, Taiwan, pp. 1466-1473

[CI4] J. Montalvao, **D. Istrate**, J. Boudy and J. Mouba, « Sound Event Detection in Remote Health Care - Small Learning Datasets and Over Constrained Gaussian Mixture Models », IEEE EMBC 2010, 31 Aout – 4 Septembre, Buenos Aires, Argentine, pp.1146 – 1149

[CI5] T. Guettari, P. A. C. Aguilar, J. Boudy, H. Medjahed, **D. Istrate**, J.L. Baldinger, I. Belfeki, M. Opitz, M. Maly-Persy, « Multimodal Localization in the Context of a Medical Telemonitoring System », IEEE EMBC 2010, 31 Aout – 4 Septembre, Buenos Aires, Argentine, pp.3835 – 3838

[CI6] M. Sehili, **D.Istrate**, "First Investigation of Sound Recognition for domotic application using SVM", SINTES 2010, 17-19 Octobre 2010, Sinaia, Roumanie, pp.745-748

[CI7] J.E. Rougui, **D. Istrate**, W. Souidene, M. Opitz et M. Riemann, « Audio based surveillance for cognitive assistance using a CMT microphone within socially assistive technology », IEEE EMBC2009, 2-6 Septembre, Minneapolis, USA, 2009, pp.2547-2550

[CI8] J.E. Rougui, **D. Istrate**, W. Souidene, « Audio Sound Event Identification for distress situations and context awareness », IEEE EMBC2009, 2-6 Septembre, Minneapolis, Etats-Unis, 2009, pp. 3501-3504

[CI9] H. Medjahed, **D. Istrate**, J. Boudy and B. Dorizzi, « Human Activities of Daily Living Recognition Using Fuzzy Logic For Elderly Home Monitoring », IEEE Fuzzy Systems 2009, 20-24 Aout 2009, Jeju Island, Korea, pp.2001-2006

[CI10] H. Medjahed, **D. Istrate**, J. Boudy and B. Dorizzi, « A Fuzzy Logic System for Home Elderly People Monitoring (EMUTEM) », Fuzzy Systems 2009, 23-25 Mars 2009, Prague, République Tchèque, ISBN 978-960-474-066-6, pp. 69-75

[CI11] F. Fauberteau, S. Midonnet and **D. Istrate**, « Power Saving of Real Time Embedded Sensor for Medical Remote Monitoring », ICONS 2009, 1-6 Mars 2009, Cancun, Mexique, pp.63-67, DOI 10.1109/ICONS.2009.47

[CI12] W. Souidene, **D. Istrate**, H. Medjahed, J. Boudy, J.L. Baldinger, I. Belfeki and J.F.Delavaut, « Multi-Modal Platform for In-Home Healthcare Monitoring (EMUTEM) », International Conference on Health Informatics (HEALTHINF 2009), 14-17 Janvier, 2009, Porto, Portugal, pp. 381-386

[CI13] E. Iancu, I. Iancu, **D. Istrate** et M. Mota, « Glucose Level Prediction using Artificial Neural Networks », 9th WSEAS International Conference on Simulation, Modelling and Optimization, 3-5 Septembre, Budapest, Hongrie, ISBN 978-960-474-113-7, pp. 407-412

[CI14] **D. Istrate**, M. Binet, S. Cheng, « Real Time Sound Analysis for Medical Remote Monitoring », IEEE EMBC2008, 20-24 Aout 2008, Vancouver, Canada, pp. 4640-4643.

[CI15] M.Vacher, **D.Istrate**, J.F. Serignat, « Speech and Sound analysis: an Application of Probabilistic Models », SINTES 2007, 18-20 Octobre, Craiova, Roumanie, pp.173-178

[CI16] H.Medjahed, **D.Istrate**, J. Boudy, F.Steenkeste, J.L. Baldinger, I. Belfeki, V. Martins and B.Dorizzi, « A Multimodal Platform for Database Recording and Elderly People Monitoring », BIOSIGNALS 2008, 28-31 janvier 2008, Funchal-Madeira, Portugal, pp.385-392

[CI17] M. Chenafa, **D.Istrate**, V. Vrabie, M. Herbin, « Speaker Recognition using Decision Fusion », BIOSIGNALS 2008, 28-31 janvier 2008, Funchal-Madeira, Portugal, pp.267-272

[CI18] G. Virone, **D. Istrate**, « Integration of an Environmental Sound Module to an Existing In-Home Activity Simulator », IEEE EMBC 2007, ISBN 1-4244-0788-5, August 23-26, Lyon, France, pp. 3810-3814

[CI19] **D. Istrate**, M. Vacher, J.F.Serignat, « Embedded Implementation of Distress Situation Identification through Sound Analysis », the 5th ICICTH Samos : International Conference on Information Communication Technologies in Health, 5-7 Juillet 2007, Samos, Grèce, pp. 226-231

[CI20] **D. Istrate**, M. Vacher, J.F.Serignat, « Generic Implementation of a Distress Sound Extraction System for Elder Care », 28th IEEE EMBS Annual International Conference 2006, New York City, USA, 30 Aout – 3 Septembre, 2006, pp.3309-3312

[CI21] M. Vacher, P. Menendez-Garcia, J.F.Serignat, **D. Istrate**, « First Implementation of a Sound/Speech Remote Monitoring Real-Time System for Home Healthcare », 6th International Conference IEEE Communications 2006, Bucharest (Romania), 14-16 Juin 2006, pp.111-115

[CI22] M. Vacher, **D. Istrate**, J.F. Serignat and N. Gac, « Detection and Speech/Sound Segmentation in a Smart Room Environment », The 3rd Conference on Speech Technology and Human-Computer Dialogue (Sped 2005), Cluj-Napoca, Romania, May 13-14, 2005, pp. 37-48

[CI23] **D. Istrate**, N. Scheffer, C. Fredouille et J-F. Bonastre, « Broadcast News Speaker Tracking for ESTER 2005 Campaign », EUROSPEECH 05, September 2005, Lisboa, Portugal, pp 2445-2448

[CI24] **D. Istrate**, C. Fredouille, S. Meignier, L. Besacier et J. Bonastre, « NIST RT'05S Evaluation: Pre-processing Techniques and Speaker Diarization on Multiple Microphone Meetings », in Proc. MLMI, 2005, pp.428-439.

[CI25] M. Vacher, **D. Istrate** and J. F. Serignat, «Sound Detection Through Transient Models Using Wavelet Coefficient Trees», Complex System Intelligence and Modern Technological Applications, Cherbourg, France, septembre 2004, pp. 367 – 372

[CI26] **D. Istrate**, M. Vacher, J. F. Serignat and E. Castelli, «Multichannel Smart Sound for Perceptive Spaces», Complex System Intelligence and Modern Technological Applications, Cherbourg, France, septembre 2004, pp. 691 – 696

[CI27] **D. Istrate**, M. Vacher, E. Castelli, C. P. Nguyen, «Sound Processing for Health and Smart Home», International Conference On Smart homes and health Telematic (ICOST2004), Singapore, septembre 2004

[CI28] E. Castelli, **D. Istrate**, and C.P. Nguyen, «Sound System Analysis for Health Smart Home», International Conference on Electronics, Informations, and Communications, Hanoi, Vietnam, août 2004

[CI29] M. Vacher, **D. Istrate** and J. F. Serignat, «Sound Detection And Classification Through Transient Models Using Wavelet Coefficient Trees», European Signal Processing Conference, Vienne, Autriche, septembre 2004, pp. 1171 – 1174

[CI30] M. Vacher, **D. Istrate**, L. Besacier, J.F.Serignat,E. Castelli, «Sound Detection and Classification for Medical Telesurvey», IASTED Biomedical Conference, Innsbruck, Autriche, février 2004, pp. 395-399

[CI31] G. Virone, **D. Istrate**, M. Vacher, J. F. Serignat, N. Noury, J. Demongeot, «First Steps in Data Fusion between a Multichannel Audio Acquisition and an Information System for Home Healthcare», IEEE Engineering In Medicine And Biology Society Conference, Cancun, Mexique, 13-15 Septembre 2003, p. 1364-1367

[CI32] **D. Istrate**, G. Virone, M. Vacher, E. Castelli, J. F. Serignat, «Communication Between A Multichannel Audio Acquisition And An Information System In A Health Smart Home For Data Fusion», International Association of Science and TEchnology for Development Conference on Internet and Multimedia Systems and Networks, Honolulu,Hawaii, 13-15 Août 2003

[CI33] M.Vacher, **D.Istrate**, L.Besacier, J.F.Serignat, E.Castelli, «Life Sounds Extraction and Classification in Noisy Environment"», International Association of Science and TEchnology for Development Conference on Signal Image Processing, Honolulu,Hawaii, 13-15 Août 2003.

[CI34] E.Castelli, M.Vacher, **D.Istrate,** L.Besacier, J.F.Serignat, «Habitat Telemonitoring System Based on the Sound Surveillance», International Conference on Information Communication Technologies in Health, Samos, Greece, 13-15 Juillet 2003, pp. 141-146

[CI35] M.Vacher, **D.Istrate**, L.Besacier, J.F.Serignat, E.Castelli, «Smart Audio Sensor for Telemedicine», Smart Objects Conferences sOc'2003, Grenoble, France, 15-17 Mai 2003, pp. 222-225

[CI36] **D.Istrate**, E.Castelli, «Everyday Life Sounds and Speech Analysis for a Medical Telemonitoring System », EUROSPEECH Conference, Aalborg, Danemark, 3-7 septembre 2001, pp. E15 2417-2420

[CI37] E.Castelli, **D.Istrate**, «Multichannel Audio Acquisition for Médical Supervision in an Intelligent Habitat », European Conference on Circuits Theory and Devices, Helsinki, Finlande, 28-31 Août 2001, pp. II-1 II-4

[CI38] E.Castelli, **D.Istrate**, V.Rialle, N.Noury, «Information extraction from speech in stress situation. Application to the Medical Supervision in a Smart House», Conférence ORAGE(ORAlité et GEstualité), Aix-en-Provence, France, 18-22 juin 2001, pp. 362-371

VI.2.5. Conférences nationales avec actes et comité de lecture

[CN1] W. Souidene, **D. Istrate**, H. Medjahed, J. Boudy, J. L. Baldinger, I. Belfeki and J.F. Delavaut, « Une Plateforme Multimodale pour la Télévigilance Médicale », AMINA 2008, 13-15 Novembre 2008, Monastir, Tunisie, pp.456-459

[CN2] T. Guettari, P. A. Cavalcante, J. Boudy, H. Medjahed, **D. Ist**rate, J.-L. Baldinger, I. Belfeki, « Localisation multimodale dans le contexte d'un système de télévigilance médicale », Conférence Handicap 2010, 9-11 Juin 2010, Paris, France

[CN3] J. Montalvão, **D. Istrate**, J. Boudy, « Signal Features for Event Detection in Remote Health Care through Audio Signals », Conférence Handicap 2010, 9-11 Juin 2010, Paris, France

[CN4] J. Montalvão, **D. Istrate**, J. Boudy, « Improved Signal Representation for Event Detection in Remote Health Care Through Masking », CBA 2010 Congresso Brasileiro de Automatica, 12-16 Septembre 2010, Bonito, Brésil, pp.231-234

[CN5] H.Medjahed, **D.Istrate**, J.Boudy, B.Dorizzi, F. Steenkeste, « Environnement multimodal pour la televigilance médicale à domicile EMUTEM », Congrès SFTAG 2009, Troyes, 18-20 novembre 2009.

[CN6] T.Guettari, H. Medjahed, J.Boudy, **D.Istrate**, J.L.Baldinger, I.Belfeki, « Traitement Acoustique pour la Télévigilance Médicale à domicile », Congrès SFTAG, Troyes, 18-20 novembre 2009.

[CN7] E. Iancu, I. Iancu, **D.Istrate**, M.Mota, « Prediction du nivel de glucose en utilisant des réseaux de neuronnes », Congrès SFTAG, Troyes, 18-20 novembre 2009.

[CN8] M. Chenafa, **D. Istrate**, V. Vrabie et W. Souidene, « Reconnaissance automatique du locuteur par fusion de décision », Majestic 2008, 29-31 Octobre, Marseille, France, 2008.

VI.2.6. Communications industriels

[ComI1] **D. Istrate**, « Télésurveillance sonore pour la maison intelligente », Application utilisateur sur le site web National Instruments (http://sine.ni.com/cs/app/doc/p/id/cs-12656)

[ComI2] **D. Istrate**, H. Medjahed, « Environnement Multimodal pour la Télévigilance Médicale à Domicile (EMUTEM) », Application utilisateur sur le site web National Instruments (http://sine.ni.com/cs/app/doc/p/id/cs-12511)

[ComI3] **D. Istrate**, E. Castelli, M. Vacher and J.F. Serignat, « Détection et reconnaissance des sons de la vie courante pour la surveillance médicale », Application utilisateur sur le site web National Instruments (http://sine.ni.com/cs/app/doc/p/id/cs-10061)

VI.2.7. Communications sans acte

[CsA1] P.A. Cavalcante, J. Boudy, **D. Istrate**, H. Medjahed, B. Dorizzi, J. C. M. Mota, J.L. Baldinger, T. Guettari, I. Belfeki, « Heterogeneous multi-sensor fusion based on an Evidential Network for fall detection », International Conference on Wearable Micro and Nano technologies for personalised health (pHealth 2011), 29 Juin -1 Juillet 2011, Lyon, France

[CsA2] H. Medjahed, D.Istrate, « La Fusion de Données Multimodale pour la Télévigilance Médicale à Domicile » à la journée « Avancées en Fusion de données » de GDR ISIS du 11 février 2010.

VI.2.8. Rapports de recherche

[D4.4] D4.4 – « Voice-Commands and Sounds Analysis » - **Projet CompanionAble** - **Propriétaire** ESIGETEL – **Dan Istrate, J.E. Rougui**

[D4.2] D4.2 – « Sensory Systems » - **Projet CompanionAble** - **Propriétaire** ESIGETEL – **Dan Istrate**

[D4.1] D4.1 – « Sensory Systems Specification for both Robot Companion Environment (RCE) and Smart Home (SHE) » – **Projet CompanionAble** - Propriétaire AKG, participation ESIGETEL – **Dan Istrate**

[D4.3] D4.3 – « Human@home tracking and behaviour-emotions modeling" – **Projet CompanionAble** - Propriétaire UIL, participation ESIGETEL – **Dan Istrate et H. Medjahed**

[D4.5] D4.5 – « Human@home tracking and behaviour-emotions modelling II » - **Projet CompanionAble** - Propriétaire UIL, participation ESIGETEL – **Dan Istrate et H. Medjahed**

[D4.21] D4.21 – « Low- and high-level Fusion » - **Projet CompanionAble** - Propriétaire GET, Participation ESIGETEL – **Dan Istrate, H. Medjahed**

[D4.22] D4.22 – « Training and test data specification » - **Projet CompanionAble** - Propriétaire UIL, Participation ESIGETEL – **Dan Istrate**

[D3.2] D3.2 – « Framework Architecture Specification » - **Projet CompanionAble** - Propriétaire GET, Participation ESIGETEL – **Dan Istrate**

[D3.4] D3.4 – « Framework Architecture Specification Update » - **Projet CompanionAble** - Propriétaire GET, Participation ESIGETEL – **Dan Istrate**

[D5.2] D5.2 – « Initial SHE Integration and Conformance Test Report » - **Projet CompanionAble** - Propriétaire AKG, Participation ESIGETEL – **Dan Istrate**

[D5.3] D5.3 – « Draft SHE Technical Handbook » - **Projet CompanionAble** - Propriétaire AIT, Participation ESIGETEL – **Dan Istrate**

[D5.5] D5.5 – « Final SHE Integration and Conformance Test Report » - **Projet CompanionAble** - Propriétaire GET, Participation ESIGETEL – **Dan Istrate**

[D5.6] D5.6 – « Final SHE Technical Handbook and Guide for Trial Site Installation » - **Projet CompanionAble** - Propriétaire AIT, Participation ESIGETEL – **Dan Istrate**

[D11.1] D11.1 – « Partner-specific Exploitation Plans, IPR and Consortium Agreement » - **Projet CompanionAble** - Propriétaire GET, Participation ESIGETEL – **Dan Istrate**

[L4.2] L4.2 – « Système embarqué de surveillance sonore » - **Projet QuoVADis** - **Propriétaire** ESIGETEL – A. Amehraye**, Dan Istrate**, H. Medjahed

[L4.3] L4.3 – « Environnement Multimodal pour la Télévigilance Médicale à Domicile EMUTEM » - Propriétaire GET, **Participation majeure** ESIGETEL - **Dan Istrate**, A. Amehraye, H. Medjahed

[L7.3] L7.3 – « Intégration et validation technique » - Propriétaire GET, Participation ESIGETEL - **Dan Istrate**, H. Medjahed

VI.3. Articles joints

Les 6 publications les plus représentatives :

- D. Istrate, J. Boudy, H. Medjahed, J.L. Baldinger, Medical Remote Monitoring using sound environment analysis and wearable sensors, « Recent Advances in Biomedical Engineering », pp.517-532, Octobre 2009, ISBN 978-953-307-013-1.
- D. Istrate, N. F. Iancu, M. Chenafa, V. Vrabie, "Fusion de decision pour contrôle d'accès par la voix », Revue Ingénierie des systèmes d'information, Vol.15, N° 2/2010, pages 11-27, ISBN 978-2-7462-2956-3
- H. Medjahed, D. Istrate, J. Boudy, J.L. Baldinger, B. Dorizzi, « A pervasive Multi-sensor Data Fusion for Smart Home Healthcare Monitoring », IEEE Conference on Fuzzy Systems 2011, 27-30 Juin 2011, Taipei, Taiwan, pp. 1466-1473
- J. Montalvao, D. Istrate, J. Boudy and J. Mouba, « Sound Event Detection in Remote Health Care - Small Learning Datasets and Over Constrained Gaussian Mixture Models », IEEE EMBC 2010, 31 Aout – 4 Septembre, Buenos Aires, Argentine, pp.1146 – 1149
- J.E. Rougui, D. Istrate, W. Souidene, M. Opitz et M. Riemann, « Audio based surveillance for cognitive assistance using a CMT microphone within socially assistive technology », IEEE EMBC2009, 2-6 Septembre, Minneapolis, USA, 2009, pp.2547-2550
- D. Istrate, M. Vacher and J.-F. Serignat, Embedded Implementation of Distress Situation Identification Through Sound Analysis, The Journal on Information Technology in Healthcare, 2008, 6(3), pp. 204-211.

Medical Remote Monitoring using sound environment analysis and wearable sensors

Dan Istrate[1], Jérôme Boudy[2], Hamid Medjahed[1,2] and Jean Louis Baldinger[2]
[1]*ESIGETEL-LRIT, 1 Rue du Port de Valvins, 77210 Avon France*
[2]*Telecom&Management SudParis, 9 Rue Charles Fourier, 91011 Evry France*

1. Introduction

The developments of technological progress allow the generalization of digital technology in the medicine area, not only the transmission of images, audio streams, but also the information that accompany them. Many medical specialties can take advantage of the opportunity offered by these new communication tools which allow the information share between medical staff. The practice of medicine takes a new meaning by the development and diffusion of Information and Communication Technologies (ICT). In the health field, unlike other economic sectors, the technical progress is not necessarily generating productivity gains but generate more safety and comfort for patients.

Another fact is that the population age increases in all societies throughout the world. In Europe, for example, the life expectancy for men is about 71 years and for women about 79 years. For North America the life expectancy, currently is about 75 for men and 81 for women[i]. Moreover, the elderly prefer to preserve their independence, autonomy and way of life living at home the longest time possible. The number of medical specialists decreases with respect to the increasing number of elderly fact that allowed the development of technological systems to assure the safety (telemedicine applications).

The elderly living at home are in most of the cases (concerning Western and Central Europe and North America) living alone and with an increased risk of accidents. In France, about 4.5 % of men and 8.9% of women aged of 65+ years has an everyday life accident[ii]. Between these everyday life accidents, the most important part is represented by the domestic accidents; about 61% (same source) and 54% of them take place inside the house. In France, annually, 2 millions of elderly falls take place, which represent the source of 10 000 deaths[iii]. Between 30% and 55% of falls cause bruises and only 3% to 13% of falls are the causes of serious injuries such as fractures, dislocation of a joint, or wounds. Apart from physical injury and hospitalization, a fall can cause a shock (especially if the person cannot recover only after the fall). This condition can seriously affect the senior psychology, he might looses

the confidence in his abilities and can result in a limitation of daily activities and, consequently, in a decrease of the life quality.

In order to improve the quality of life of elderly several applications has been developed: home telemonitoring in order to detect distress situations and audio-video transmission in order to allow specialists to diagnose patient at distance.

This chapter describe a medical remote monitoring solution allowing the elderly people to live at home in safety.

2. Telemedecine applications

The term "telemedicine" appears in a dictionary of the French language for the first time in the early 1980's, the prefix "tele" denoting "far away". Thus, telemedicine literally means remote medicine and is described as "part of medicine, which uses telecommunication transmission of medical information (images, reports, records, etc.) in order to obtain remote diagnosis, a specialist opinion, continuous monitoring of a patient, a therapeutic decision."
Using a misnomer, one readily associates the telemedicine to the generic term "health telematics". This term has been defined by the World Health Organization in 1997 and "refers to the activities, services and systems related to health, performed remotely using information technology and communication needs for global promotion of health, care and control of epidemics, management and research for health."

The interest of telemedicine is far from being proved and is not without stimulating reflection, particularly in the areas ethical, legal and economic. The main telemedicine applications are:

- **Telediagnostic** = The application which allow a medical specialist to analyze a patient at distance and to have access to different medical analysis concerning the patient. A specific case can be if a specialist is at the same place with the patient but need a second opinion from another one.
- **Telesurgery** = technical system allowing a surgery at distance for spatial or military applications. Also in this category we can have the distant operation of a complex system like an echograph or the augmented reality in order to help the medicine in the framework of a surgery.
- **Telemonitoring** = an automatic system which survey some physiological parameters in order to monitor a disease evolution and/or to detect a distress situation.
- **Tele-learning** = teleconferencing systems allowing medical staff to exchange on medical information.

Among the main telemedicine applications, telediagnostic and telemonitoring are more investigated solutions. The telediagnostic allows medical specialist to consult the elderly through audio video link in order to avoid unnecessary travel for both patient and medical staff. Several systems were currently developed between hospital and nursing home, or between medical staff and a mobile unit. The main challenges are the audio-video quality,

the possibility to transmit also other medical data (ECG, medical records) and data security. In order to guarantee a good audio-video quality a high bandwidth network is needed.

The medical remote monitoring or telemonitoring can prevent or reduce the consequences of accidents at home for elderly people or chronic disease persons. The increase of aging population in Europe involves more people living alone at home with an increased risk of home accidents or falls. The remote monitoring aims to detect automatically a distress situation (fall or faintness) in order to provide safety living to elderly people.

The medical remote monitoring consists in establishing a remote monitoring system of one or more patients by one or more health professionals (physician, nursing...). This monitoring is mainly based on the use of telecommunication technology (ie the continuous analysis of patient medical parameters of any kind: respiratory, cardiac, and so on...). This technique is used in the development of hospitalizations at home, ie where the patient is medically monitored at home, especially in cases of elderly people. In addition, this method avoids unnecessary hospitalizations, increasing thus the patient comfort and security. The main aim of remote monitoring systems is to detect or to prevent a distress situation using different types of sensors.

In order to improve the quality of life of elderly several research teams have developed a number of systems for medical remote monitoring. These systems are based on the deployment of several sensors in the elderly home in order to detect critical situations. However, there are few reliable systems capable of detecting automatically distress situations using more or less non intrusive sensors. Monitoring the activities of elderly people at home with position sensors allows the detection of a distress situation through the circadian rhythms (Bellego et al., 2006). However, this method involves important data bases and an adaptation to the monitored person (Binh et al., 2008). Other studies monitor the person activity through the use of different household appliances (like oven or refrigerator) (Moncrieff et al., 2005). More and more applications use embedded systems, like smart mobile phones, to process data and to send it trough 3G networks (Bairacharya et al., 2008). In order to detect falls, several wearable sensors was developed using accelerometers (Marschollek et al., 2008), magnetic sensors (Fleury et al., 2007) or data fusion with smart home sensors (Bang et al., 2008).

There are many projects which develop medical remote monitoring system for elderly people or for chronic disease patient like TelePat project[iv] which was aimed at the realization of a service of remote support in residence for people suffering of cardiac pathologies (Lacombe et al., 2004). Other National projects like RESIDE-HIS and DESDHIS[v] have developed different modality to monitor like infra-red sensor, wearable accelerometer sensor and sound analysis. At European level (FP6) several projects has investigated the domain of combination of smart home technologies with remote monitoring like SOPRANO project which aims at the design of a system for the assistance of the old people in the everyday life for a better comfort and safety (Wolf et al., 2008).

Consequently, devices of the ambient intelligence are connected continuously to a center of external services as in the project EMERGE[vi]. This last aims by the behavior observation

through holistic approach at detecting anomalies, an alarm is sent in the case of fall, faintness or another emergency.

Three institutions (TELECOM & Management SudParis, INSERM U558 and ESIGETEL) have already developed a medical remote monitoring modality in order to detect falls or faintness. The TELECOM & Management SudParis has developed a mobile device which detects the falls, measures the person pulse, movement and position and is equipped with panic button (Baldinger et al., 2004). The ESIGETEL has developed a system which can recognize abnormal sounds (screams, object falls, glass breaking, etc.) or distress expressions (Help!, A doctor please! etc.) (Istrate et al., 2008).

Each remote monitoring modality, individually, present cases of missed detections and/or false alarms but the fusion of several modalities can increase the system reliability and allow a fault tolerant system (Virone et al., 2003). These two modalities and others are combined in the framework of CompanionAble project.

3. CompanionAble Project

A larger telemedicine application which includes sound environment analysis and wearable sensor is initiated in the framework of a European project. CompanionAble[1] project (Integrated Cognitive Assistive & Domotic Companion Robotic Systems for Ability & Security) provides the synergy of Robotics and Ambient Intelligence technologies and their semantic integration to provide for a care-giver's assistive environment. CompanionAble project aims at helping the elderly people living semi or independently at home for as long as possible. In fact the CompanionAble project combines a telemonitoring system in order to detect a distress situation, with a cognitive program for MCI patient and with domotic facilities. The telemonitoring system is based on non intrusive sensor like: microphones, infra-red sensors, door contacts, video camera, pills dispenser, water flow sensor; a wearable sensor which can detect a fall and measure the pulse and a robot equipped with video camera, audio sensors and obstacles detectors.

4. Proposed telemonitoring system

Two modalities sound and wearable sensors are presented by following. A multimodal data fusion method is proposed in the next section.

4.1 ANASON

The information from the everyday life sound flow is more and more used in telemedical applications in order to detect falls, to detect daily life activities or to characterize physical status. The use of sound like an information vector has the advantage of simple and cheapest sensors, is not intrusive and can be fixed in the room. Otherwise, the sound signal has important redundancy and need specific methods in order to extract information. The definition of signal and noise is specific for each application; e.g. for speech recognition, all sounds are considered noise. Between numerous sound information extraction applications

[1] *www.companionable.net*

we have the characterization of cardiac sounds (Lima & Barbarosa, 2008) in order to detect cardiac diseases or the snoring sounds (Ng & Koh, 2008) for the sleep apnea identification. Using sound for the fall detection has the advantage that the patient not need to carry a wearable device but less robust in the noise presence and depend from acoustic conditions (Popescu et al., 2008), (Litvak et al., 2008). The combination of several modalities in order to detect distress situation is more robust using the information redundancy.

The sound environment analysis system for remote monitoring capable to identify everyday life normal or abnormal and distress expressions is in continuous evolution in order to increase the reliability in the noise presence. Currently in the framework of the CompanionAble project a coupled smart sensor system with a robot for mild cognitive impairment patients is being developed. The sound modality is used like a simplified patient-system interface and for the distress situation identification. The sound system will participate to the context awareness identification, to the domotic vocal commands and to the distress expressions/sounds recognition. This system can use a classical sound card allowing only one channel monitor or an USB acquisition card allowing a real time multichannel (8 channels) monitoring covering thus all the rooms of an apartment.

Current systems use mainly the speech information from sound environment in order to generate speech command or to analyze the audio scene. Few studies investigate the sound information. The (Moncrieff et al., 2005) uses the sound level coupled with the use of household appliances in order to detect a threshold on patient anxiety. In (Stagera et al., 2007) some household appliances sounds are recognized on an embedded microcontroller using a vectorial quantization. This method was used to analyze the patient activities, a distress situation being possible to be detected through a long time analysis. In (Cowling & Sitte, 2002) a statistical sound recognition system is proposed but the system was tested only on few sound files.

The proposed smart sound sensor (ANASON) analyzes in real time the sound environment using a first module of detection and extraction of useful sound or speech based on the Wavelet Transform (Istrate et al., 2006). The module composition of the smart sound sensor can be observed in the Fig.1. This module is applied on all audio channels simultaneously, in real time. Only extracted sound signals are processed by the next modules. The second module classifies extracted sound event between sound and speech. This module, like the sound identification engine, is based on a GMM (Gaussian Mixture Model) algorithm. If a sound was detected the signal is processed by a sound identification engine and if a speech was detected a speech recognition engine is launched. The speech recognition engine is a classical one aiming at detecting distress expressions like ”Help!” or ”A doctor, please!”.

Signal event detection and extraction. This first module listen continuously the sound environment in order to detect and extract useful sounds or speech. Useful sounds are: glass breaking, box falls, door slap, etc. and sounds like water flow, electric shaver, vacuum cleaner, etc. are considered noise. The sound flow is analyzed through a wavelet based algorithm aiming at sound event detection. This algorithm must be robust to noise like neighbourhood environmental noise, water flow noise, ventilator or electric shaver. Therefore an algorithm based on energy of wavelet coefficients was proposed and

evaluated. This algorithm detects precisely the signal beginning and its end, using properties of wavelet transform even at signal to noise ratio (SNR) of 0 dB. The signals extracted by this module are recorded in a safe communication queue in order to be processed by the second parallel task.

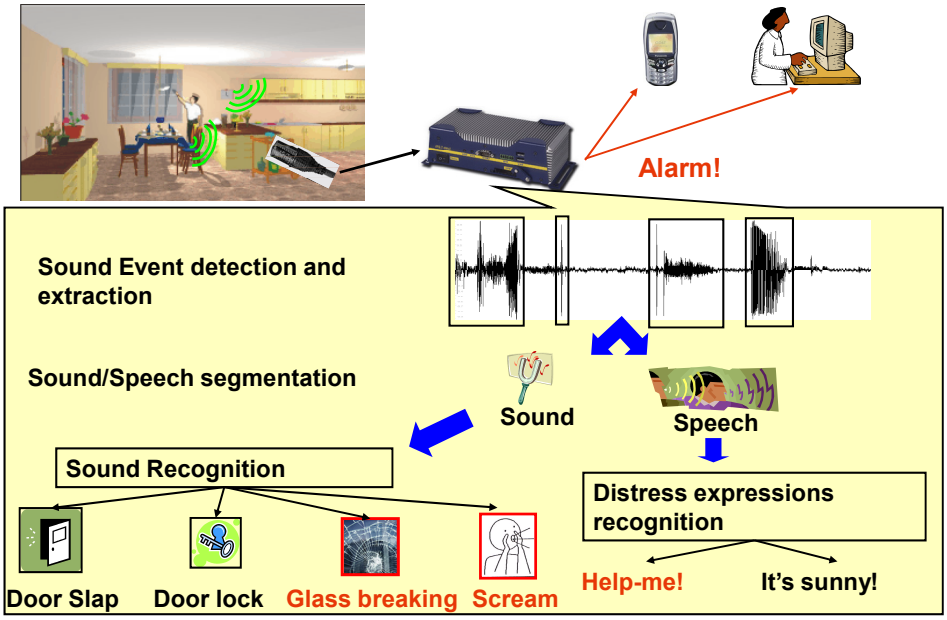


Fig. 1. Sound environment analysis system (ANASON)

Sound/speech segmentation. The second module is a low-stage classification one. It processes the extracted sounds in order to separate the speech signals from the sound ones. The method used by this module is based on Gaussian Mixture Model (GMM). There are other possibilities for signal classification: Hidden Markov Model (HMM), Bayesian method, etc. Even if similar results have been obtained with other methods, their high complexity and high time consumption prevent from real-time implementation.

A preliminary step before signal classification is the extraction of acoustic parameters: LFCC (Linear Frequency Cepstral Coefficients) - 24 filters. The choice of this type of parameters relies on their properties: bank of filters with constant bandwidth, which leads to equal resolution at high frequencies often encountered in life sounds. Other types of acoustical parameters like zero crossing rate, roll-off point, centroid or wavelet transform based was tested with good results.

Sound recognition. This module composes with the previous one the second parallel task and classifies the signal between several predefined sound classes. This module is based,

also, on a GMM algorithm. The 16 MFCC (Mel Frequency Cepstral Coefficients) acoustical parameters have been used coupled with ZCR (Zero crossing rate), Roll-off Point and Centroid. The MFCC parameters are computed from 24 filters. A log-likelihood is computed for the unknown signal according to each predefined sound classes; the sound class with the biggest log likelihood constitute the output of this module.

In the current version, the number of Gaussians is optimized according to data base size which allows having different number of Gaussians for each sound class. Taking into account that for some sounds, especially for abnormal ones, is difficult to record an important number, we have chosen to allow a variation between 4 and 60 Gaussians for the sound models.

Distress expressions recognition. In order to detect distress expressions two possibilities can be considered: the use of a classical speech recognition engine followed by a textual detection of distress expressions or a word spotting system. The first solution has tested with good results through a vocabulary optimization with specific words.

If an alarm situation is identified (the sound or the sentence is classified into an alarm class) this information and the sound signal are sent to the data fusion system. In the case of a typical everyday life sound, only the extracted information (and not the sound) is sending to the data fusion system.

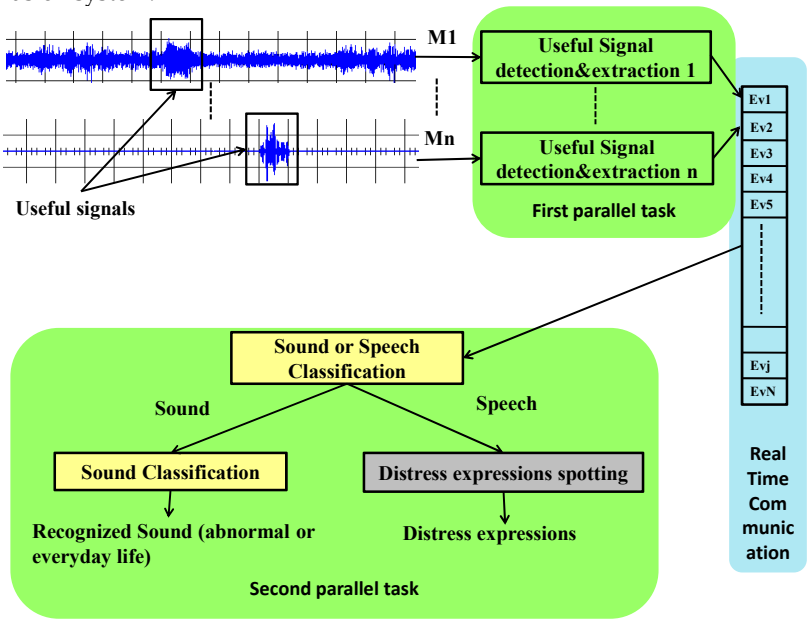


Fig. 2. ANASON real time implementation

ANASON system has been implemented in real time on PC or embedded PC using three parallel tasks (Fig. 2.):

1. Sound Acquisition + Sound Event Detection & Extraction

2. Hierarchical Sound Classification
3. Graphical User Interface and Alarm management

ANASON modality carries out also localization information concerning the microphone which has been used to recognize the abnormal sound or speech and a confidence measure in the output (SNR value).

The speech monitoring allows the system to detect a distress request coming from the patient, if the patient in the distress situation is conscious (the same role that panic button of RFPAT).

Globally, ANASON software has no false alarms and 20 % of missed detections for signals with SNR between 5 and 20 dB (real test conditions). The Useful signal detection and extraction module and the Sound or Speech Classification module work correctly even for signals with a SNR about 10 dB but the sound or speech recognition modules need at least a SNR of 20 dB. We work currently to ameliorate these performances by adding specific filtering and noise adaptation modules.

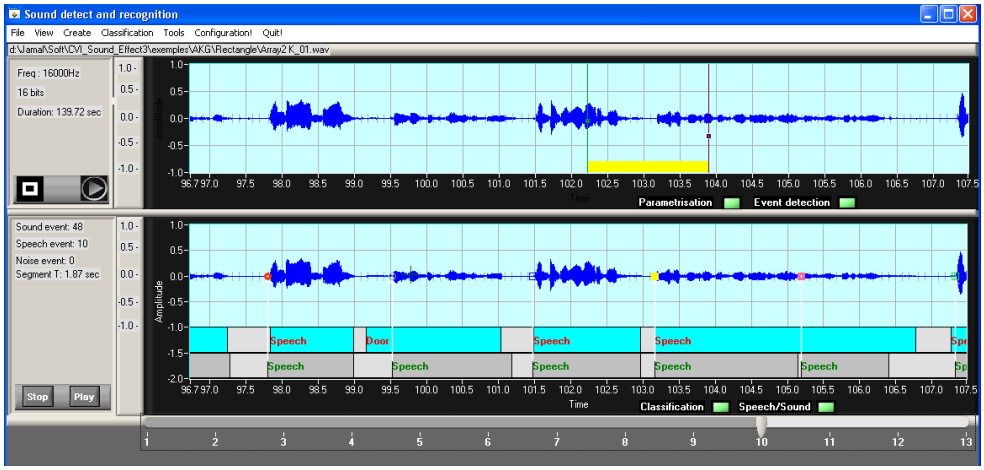


Fig. 3. Example of sound/speech detection and recognition

Fig.3. shown the ANASON algorithm application on a signal recorded in our laboratory. In the second window the blue rectangle represent the automatic output of ANASON and the gray ones the reference labels (manually labels). We can observe some reduced errors on the start/stop time of each event. All detected events were correctly classified.

4.2 RFPAT

The remote monitoring modality RFPAT consists in two fundamental modules (Fig. 2.):

- A mobile terminal (a waist wearable device that the patient or the elderly clips to his belt, for instance, all the time he is at home; it measures the person's vital data and sends it to a reception base station)

- A fixed reception base station (a receiver connected to a personal computer (PC) through a RS232 interface; it receives vital signals from the patient's mobile terminal, analyzes and records them).

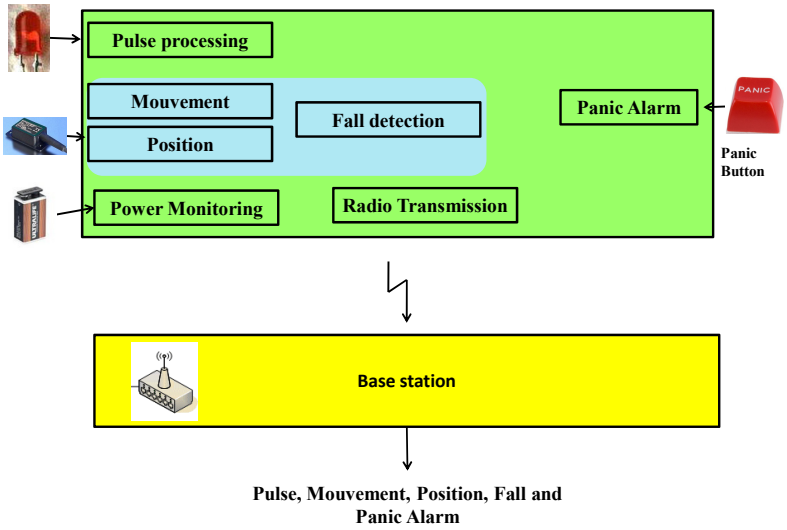


Fig. 4. Wearable device (RFPAT)

All the data gathered from the different RFPAT sensors are processed within the wireless wearable device. To ensure an optimal autonomy for the latter, it was designed using low consumption electronic components. Namely, the circuit architecture is based on different micro-controllers devoted to acquisition, signal processing and emission. Hence, the mobile wearable terminal (Fig. 4.) encapsulates several signal acquisition and processing modules:

- to records pulse rate, actimetric signals (posture, movement) and panic button
- to pre-process the signals in order to reduce the impact of environmental noise or user motion noise.

This latter point is an important issue for in-home healthcare monitoring. In fact, monitoring a person in ambulatory mode is a difficult task to achieve. For the RFPAT system, the noise is filtered in the acquisition stage inside the wearable device using digital noise reduction filters and algorithms. These filters and algorithms were applied respectively to all acquired signals: movement data, posture data and namely the pulse signal (heart rate).

Movement data describes the movement of the monitored person. It gives us information like: "immobile", "normal life movements", "stressed person", etc. Movement data consists also in the percentage of movement, it computes the total duration of the movements of the monitored person for each time slot of 30 seconds (0 to 100% during 30 seconds).

The posture data is information about the person posture: standing up/laying down. The posture data is a quite interesting measurement which gives us useful information about the person's activity.

Thanks to an actimetric system embedded in the portable device, we can detect the situations where the person is approaching the ground very quickly. This information is interpreted as a "fall" when the acceleration goes through a certain threshold in a given situation.

The pulse signal is delivered by a photoplethysmographic sensor connected to the wearable device. After pre-conditioning and algorithmic de-noising it gives us information about the heart rate every 30 seconds.

In the ambulatory mode, the challenging process consists in noise reduction (Baldinger et al., 2004). We afford to reduce the variations of pulse measurement lower than 5% for one minute averaging, which remains in conformity with the recommendations of medical professionals.

Data gathered from the different sensors are transmitted, via an electronic signal conditioner, to low power microcontroller based computing unit, embedded in the mobile terminal.

Currently, a fall-impact detector is added to this system in order to make the detection of falls more specific.

5. EMUTEM platform

A data synchronization and fusion platform, EMUTEM (Multimodal environment for medical remote monitoring), was developed (Medjahed et al., 2009).

In order to maximize correct recognition of the various activities daily live (ADL) like sleeping, cleaning, bathing etc..., and distress situation recognition, data fusion over the different sensors types is studied. The area of data fusion has generated great interest among researchers in several science disciplines and engineering domains. We have identified two major classes of fusion techniques:

- Those that are based on probabilistic models (such as Bayesian reasoning (Cowel et al., 1999) and the geometric decision reasoning like Mahanalobis distance), but their performances are limited when the data are heteregeneous and insufficient for the correct statistical modeling of classes, therefore the model is uncontrollable.
- Those based on connectionist models (such as neural networks MLP (Dreyfus et al., 2002) and SVM (Bourges, 1998)) which are very powerful because they can model the strong nonlinearity of data but with complex architecture, thus lack of intelligibility.

Based on those facts and considering the complexity of the data to process (audio, physiologic and multisensory measurements) plus the lack of training sets that reflect activities of daily living, fuzzy logic has been found useful to be the decision module of the

multimodal ADLs recognition system. Fuzzy logic can gather performance and intelligibility and it deals with imprecision and uncertainty. It has a background application history to clinical problems including use in automated diagnosis (Adlassnig, 1986), control systems (Mason et al., 1997), image processing (Lalande et al., 1997) and pattern recognition (Zahlmann et al., 1997). For medical experts is easier to map their knowledge onto fuzzy relationships than to manipulate complex probabilistic tools.

Everyday life activities in the home split into two categories. Some activities show the motion of the human body and its structure. Examples are walking, running, standing up, setting down, laying and exercising. These activities may be most easily recognized using sensors that are placed on the body (e.g. (Makikawa & Iizumi, 1995)(Himberg et al., 2001)(Lee and Mase, 2002)). A second class of activities is recognized by identifying or looking for patterns in how people move things. In this work we focus on some activities identification belong to these both categories by using fuzzy logic. The use of fuzzy logic is motivated by two main raisons from a global point of view:

- Firstly the characteristic of data to merge which are measurements obtained from different sensors, thus they could be imprecise and imperfect.
- Secondly, the history of fuzzy logic proves that it is used in many cases which are necessary for pattern recognition applications.

5.1 Fuzzy Logic

Fuzzy logic is a powerful framework for performing automated reasoning. It reflects human reasoning based on inaccurate or incomplete data. It uses the concept of partial membership, each element belongs partially or gradually to fuzzy sets that have been already defined. In contrast to classical logic where the membership function $m(x)$ of an element x belonging to a set S could take only two values: $m_S(x) = 1$ if $x \in S$ or $m_S(x) = 0$ if $x \notin S$, Fuzzy logic introduces the concept of membership degree of an element x to a set S and $m_S(x) \in [0, 1]$, here we speak about truth value.

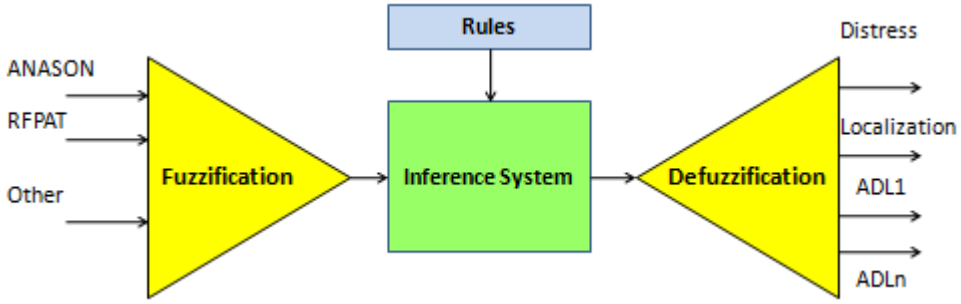


Fig. 5. Fuzzy Logic data fusion

The Fig. 5. shows the main fuzzy inference system steps:

- **Fuzzification**: First step in fuzzy logic is to convert the measured data into a set of fuzzy variables. It is done by giving value (these will be our variables) to each of a membership functions set. Membership functions take different shape: triangular,

trapezoidal, Gaussian, generalized Bell, sigmoidally shaped function, single function etc. The choice of the function shape is iteratively determinate, according to type of data and taking into account the experimental results.

- **Fuzzy rules and inference system**: The fuzzy inference system uses fuzzy equivalents of logical AND, OR and NOT operations to build up fuzzy logic rules. An inference engine operates on rules that are structured in an IF-THEN format. The IF part of the rule is called the antecedent, while the THEN part of the rule is called the consequent. Rules are constructed from linguistic variables. These variables take on the fuzzy values or fuzzy terms that are represented as words and modelled as fuzzy subsets of an appropriate domain. There are several types of fuzzy rules, we mention only the two mains used in our system:
    - Mamdani rules (Jang et al., 1997) which is of the form: If x_1 is S_1 and x_2 is S_2 and...and x_p is S_p Then y_1 is T_1 and y_2 is T_2 and...and y_p is T_p. Where S_i and T_i are fuzzy sets that define the partition space. The conclusion of a Mamdani rule is a fuzzy set. It uses the algebraic product and the maximum as Tnorm and S-norm respectively, but there are many variations by using other operators.
    - Takagi/Sugeno rules (Jang et al., 1997): If x_1 is S_1 and x_2 is S_2 and...and x_p is S_p Then $y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p$. In the Sugeno model the conclusion is numerical. The rules aggregation is in fact the weighted sum of rules outputs.
- **DeFuzzification:** The last step of a fuzzy logic system consists in turning the fuzzy variables generated by the fuzzy logic rules into real value again which can then be used to perform some action. There are different defuzzification methods; in our platform decision module we could use Centroid of area (COA), Bisector of area (BOA), Mean of Maximum (MOM), Smallest of Maximum (SOM) and Largest of Maximum (LOM).

5.2 Fuzzy Logic for medical telemonitoring

The first step for developing this approach is the fuzzification of system outputs and inputs obtained from each sensor and subsystem.

From ANASON subsystem three inputs are built. The first one is the sound environment classification; all sound class and expressions detected are labelled on a numerical scale according to their source. Nine membership functions are set up in this numerical scale according to sound sources as it is in Table 1. N other inputs are associated to each SNR calculated on each microphone (N microphones are used in the current application), and these inputs are split into three fuzzy levels: low, medium and high.

RFPAT produce five inputs: heart rate for which three fuzzy levels are specified normal, low and high; activity which has four fuzzy sets: immobile, rest, normal and agitation; posture is represented by two membership functions standing up/setting down and lying; fall and call have also two fuzzy levels: Fall/Call and No Fall/Call. The defined area of each membership function associated to heart rate or activity is adapted to each monitored elderly person.

The time input has five membership functions morning, noon, afternoon, evening and night which are also adapted to patient habits.

Membership Function	Composition
Human Sound	snoring, yawn, sneezing, cough, cry, scream, laught
Speech	key words and expressions
Multimedia Sounds	TV, radio, computer, music
Door sounds	door claping, door knob, key ring
Water sounds	water flushing, water in washbasin, coffee filter
Ring tone	telephone ring, bell door, alarm, alarm clock
Object sound	chair, table, tear-turn paper, step foot
Machine sounds	coffee machine, dishwasher, electrical shaver, microwave, vaccum cleaner, washing machine, air conditioner
Dishwasher	glass vs glass, glass wood, plastic vs plastic, plastic vs wood, spoon vs table

Table 1. Fuzzy sets defined for the ANASON classification input

The output of the fuzzy logic ADL recognition contains some activities and distress situation identification. They are sleeping, getting up, toileting, bathing, washing hands, washing dishes, doing laundry, cleaning, going out of home, enter home, walking, standing up, setting down, laying, resting, watching TV and talking on telephone. These membership functions are ordered, firstly according to the area where they maybe occur and secondly according to the degree of similarity between them.

The next step of the fuzzy logic approach is the fuzzy inference engine which is formulated by a set of fuzzy IF-THEN rules. This second stage uses domain expert knowledge regarding activities to produce a confidence in the occurrence of an activity. Rules allow the recognition of common performances of an activity, as well as the ability to model special cases. A confidence factor is accorded to each rule and in order to aggregate these rules we have the choice between Mamdani or Sugeno approaches available under the fuzzy logic component. After rules aggregation the defuzzification is performed by the centroid of area for the ADL output.

The proposed method was experimentally achieved on a simulated data in order to demonstrate its effectiveness. The first study was devoted to the evaluation of the system by taking into account rules used in this fuzzy inference system. The used strategy consisted in realizing several tests with different combination rules, and based on obtained results one rule is added to the selected set of rules in order to get the missed detection. With this strategy good results are reached for the ADL output (about 97% of good ADL detection).

6. Conclusions

This chapter has presented the usage of the sound environment information in order to detect a distress situation and the data fusion using Fuzzy Logic between sound extracted information and a wearable sensor. All presented system is the basis of the development of a complex companion system (CompanionAble project). The telemonitoring systems using redundant sensors in order to detect distress situation but also to prevent trough a long time analysis represents a solution to the lack of medical staff. These systems do not replace the care givers but represent only a help for them.

7. References

Adlassnig K. P. (1986). Fuzzy set theory in medical diagnosis. *IEEE Transactions On System, Man and Cybernetics*, Vol. 16, No. 2, pp. 260–265.

Bairacharya A.; Gale T.J.; Stack C.R. & Turner P. (2008). 3.5G Based Mobile Remote Monitoring System, *Proceedings of EMBC 2008,* pp. 783-786, doi: 10.1109/IEMBS.2008.4649269, Vancouver, Canada, August 2008

Baldinger J.L.; Boudy J.; Dorizzi B.; Levrey J.; Andreao R.; Perpre C.; Devault F.; Rocaries F. & Lacombe A. (2004). Telesurveillance system for patient at home: The medeville system, *Proceedings of ICCHP 2004,* pp. 400-407, Paris, France, July 2004

Bang S.; Kim M.; Song S.K. & Park S.J. (2008). Toward real time detection of the basic living activity in home using a wearable sensor and smart home sensors, *Proceedings of EMBC 2008*, pp. 5200-5203, doi: 10.1109/IEMBS.2008.4650386, Vancouver, Canada, August 2008

Bellego G. L.; Noury N.; Virone G.; Mousseau M. & Demongeot J. (2006). Measurement and model of the activity of a patient in his hospital suite. *IEEE Transactions on TITB*, Vol. 10, No. 1, pp. 92–99

Binh X.L.; Mascolo M.; Gouin A. & Noury N. (2008). Health Smart Home for elders - A tool for automatic recognition of activities of daily living, *Proceedings of EMBC 2008,* pp 3316-3319, doi: 10.1109/IEMBS.2008.4649914, Vancouver, Canada, August 2008

Burges C. J. C. (1998). A tutorial on SVM for Pattern Recognition. *Data Mining and Knowledge Discovery*, Vol. 2, No. 2, pp. 121–167.

Cowell R.; Dawid A.; Lauritzen S. & Spiegelhalter D. (1999). *Probabilistic Networks and Expert Systems*, Springer, ISBN: 0-387-98767-3, New York.

Cowling M. & Sitte R. (2002). Analysis of speech recognition techniques for use in a non-speech sound recognition system. *Digital Signal Processing for Communication Systems*, Vol. 703, No. 1, pp. 31-46

Dreyfus G.; Martinez J.M.; Samuelides M.; Gordon M.; Badran F.; Thiria S. & Hrault L. (2002). *Réseaux de neurones. Méthodologie et applications*, Eyrolles, ISBN 2-212-11019-7, France.

Fleury A.; Noury N. & Vuillerme N. (2007). A Fast Algorithm to Track Changes of Direction of a Person Using Magnetometers, *Proceedings of IEEE EMBS 2007*, pp. 2311-2314, doi: 10.1109/IEMBS.2007.4352788, Lyon, France, August 2007

Himberg J.; Mantyjarvi J. & Seppanen T. (2001). Recognizing human motion with multiple acceleration sensors, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 2, No. 2, pp. 747-52

Istrate D.; Castelli E.; Vacher M.; Besacier L. & Serignat J.F. (2006). Information extraction from sound for medical telemonitoring. *IEEE Transactions on TITB*, Vol. 10, No. 4, pp. 264–274

Istrate D.; Binet M. & Cheng C. (2008). Real Time Sound Analysis for Medical Remote Monitoring, *Proceedings of EMBC 2008*, pp. 4640-4643, doi: 10.1109/IEMBS.2008.4650247, Vancouver, Canada, August 2008

Jang J.-S. R.; Sun C. T. & Mizutani E. (1997). *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*, Prentice Hall, ISBN 0132610663, USA

Lacombe A.; Baldinger J.L.; Boudy J.; Dorizzi B.; Levrey J.P.; Andreao R.; Perpere C.; Delavault F.; Rocaries F. & Dietrich C. (2004). Tele-surveillance System for Patient at Home: the MEDIVILLE system, *Lecture Notes in Computer Science*, Springer-Verlag GmbH, Vol. 3118, pp 400-407, June 2004

Lalande A.; Legrand L.; Walker P. M.; Jaulent M. C.; Guy F.; Cottin Y. & Brunotte F. (1997). Automatic detection of cardiac contours on MR images using fuzzy logic and dynamic programming, Proceedings of AMIA'97, pp. 474–478, ISBN 978-3-540-62709-8, Lecture Notes in Artificial Intelligence 1211, Springer-Verlag, Berlin

Lee S.W. & Mase K. (2002). Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, Vol. 1, No. 3, pp. 24-32

Lima C. S. & Barbosa D. (2008). Automatic segmentation of the second cardiac sound by using wavelets and hidden markov models, *Proceedings of IEEE EMBC 20008*, pp. 334–337, Vancouver, Canada, August 2008

Litvak D.; Zigel Y. & Gannot I. (2008). Fall detection of elderly through floor vibrations and sound, *Proceedings of IEEE EMBC 2008*, pp. 4632–4635, Vancouver, Canada, August 2008

Makikawa M. & Iizumi H. (1995). Development of an ambulatory physical activity monitoring device and its application for categorization of actions in daily life. MEDINFO, pp. 747-750

Marschollek M.; Wolf K.H.; Gietzelt M.; Nemitz G.; Meyer zu Schwabedissen H. & Haux R. (2008). Assessing elderly persons' fall risk using spectral analysis on accelerometric data - a clinical evaluation study, *Proceedings of the EMBC 2008*, pp. 3682-3685, doi: 10.1109/IEMBS.2008.4650008, Vancouver, Canada, August 2008

Mason D.;Linkens D. & Edwards N. (1997). Self-learning fuzzy logic control in medicine, *Proceedings of AIME'97*, pp. 300–303, ISBN 978-3-540-62709-8, Lecture Notes in Artificial Intelligence 1211, Springer-Verlag, Berlin

Medjahed H.; Istrate D.; Boudy J. & Dorizzi B. (2009). A Fuzzy Logic System for Home Elderly People Monitoring (EMUTEM), *Proceedings of Fuzzy Systems 2009*, pp. 69-75, ISBN 978-960-474-066-6, Prague, Czech Republic, Mars 2009

Moncrieff S.; Venkatesh S.; West G. & Greenhill S. (2005). Incorporating contextual audio for an actively anxious smart home, *Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference*, pp. 373-378, ISBN: 0-7803-9399-6, Melbourne, Australia, December 2005

Ng A.K. & Koh T.S. (2008). Using psychoacoustics of snoring sounds to screen for obstructive apnea, *Proceedings of IEEE EMBC 2008*, pp. 1647–1650, Vancouver, Canada, August 2008

Popescu M.; Li Y.; Skubic M. & Rantz M. (2008). An acoustic fall detector system that uses sound height information to reduce the false alarm rate, *Proceedings of IEEE EMBC 2008*, pp. 4628–4631, Vancouver, Canada, August 2008

Stagera M.; Lukowiczb P. & Trostera G. (2007). Power and accuracy tradeoffs in sound-based context recognition systems. *Pervasive and Mobile Computing*, Vol. 3, No. 3, pp. 300–327, ISSN:1574-1192

Virone G.; Istrate D.; Vacher M.; Serignat J.F.; Noury N. & Demongeot J. (2003). First Steps in Data Fusion between a Multichannel Audio Acquisition and an Information System for Home Healthcare, *Proceedings of IEEE Engineering In Medicine And Biology Society Conference*, pp. 1364-1367, doi: 10.1109/IEMBS.2003.1279557, Cancun, Mexique, September 2003

Wolf P.; Schmidt A. & Klein M. (2008). SOPRANO - An extensible, open AAL platform for elderly people based on semantical contracts, Proceedings of 3rd Workshop on Artificial Intelligence Techniques for Ambient Intelligence 2008 (AITAmI'08), pp. 225-228, Patras, Greece

Zahlmann G.; Scherf M. & Wegner A. (1997). A neurofuzzy classifier for a knowledge-based glaucoma monitor, *Proceedings of AIME'97*, pp. 273–284, ISBN 978-3-540-62709-8, Lecture Notes in Artificial Intelligence 1211, Springer-Verlag, Berlin

8. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of European Community's Seventh Framework Program (FP7/2007-2011), CompanionAble Project (grant agreement n. 216487).

[i] INSEE. Espérance de vie, taux de mortalité et taux de mortalité infantile dans le monde; Population Reference Bureau of INSEE; 2007; www.insee.fr/fr/themes/tableau.asp?reg_id=98&ref_id=CMPTEF02216; retrieved in November 2008

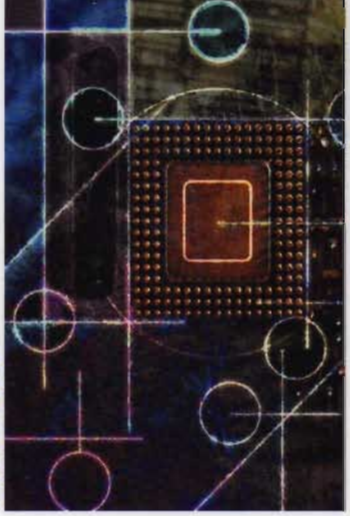
[ii] C. Duval, M.-L. Bouvet and J. Yacoubovitch. Accidents de la vie courante - Données statistiques. Health Ministry France; 2000; http://www.sante.gouv.fr/htm/pointsur/acc_dom/donnees03.htm#22; retrieved on November 2008

[iii] Le Figaro, Accidents domestiques : les personnes âgées très exposées; October 14, 2007; http://www.lefigaro.fr/france/20070604.FIG000000130_accidents_domestiques_les_personnes_agees_tres_exposees.html; retrieved on November 2008

[iv] TelePat project RNTS 2003-2006, http://www.esiee.fr/~research/documents/Index/Projets/Telepat.html; retrieved on November 2008

[v] DESDHIS, ACI Technologies for health 2002/2004

[vi] http://www.emerge-project.eu/; retrieved on November 2008:



Mars-Avril 2010

Ingénierie des systèmes d'information

RSTI série ISI • Volume 15 – n° 2/2010

Traitement de données complexes

sous la direction de

Sébastien Fournier
Aline Cauvin
Nicolas Faessel

Hermes Lavoisier

Revue des Sciences et Technologies de l'Information

série Ingénierie des systèmes d'information • Volume 15 – n° 2/2010

La recherche dans le domaine des systèmes d'information (SI) évolue vers la conception et la réalisation de systèmes ouverts (notamment sur le web) de plus en plus centrés sur les utilisateurs. Cette ouverture fait apparaître de nouveaux verrous technologiques qu'il est nécessaire de lever. Un des principaux verrous, apparu dès le départ, mais encore accentué par l'ouverture relativement récente des systèmes d'information, concerne le traitement de grandes quantités de données qui sont le plus souvent complexes et hétérogènes, comme la parole et l'image. Les articles présentés étudient le traitement de la parole dans le but d'identifier le locuteur ou bien le contenu des propos et la classification de données complexes utilisant des techniques d'apprentissage.

D. Istrate, N. Florin Iancu, M. Chenafa, V. Vrabie
Fusion de décision pour contrôle d'accès par la voix

F. Salam, H. Glotin
Indexation rapide de documents audio par traitement morphologique de la parole

M.-J. Meurs, F. Lefèvre, R. de Mori
Approche bayésienne de la composition sémantique dans les systèmes de dialogue oral

G. Thibault, B. Fertil, J. Sequeira, J.-L. Mari
Indices de formes et de textures

P. Gaillard, M. Aupetit, G. Govaert
Un graphe génératif pour la classification semi-supervisée

ISBN 978-2-7462-2956-3

Hermes Science publications
www.hermes-science.com

Lavoisier
www.Lavoisier.fr

isi.revuesonline.com

Fusion de décision pour contrôle d'accès par la voix

Dan Istrate* — Nicolae Florin Iancu* — Mohamed Chenafa* Valeriu Vrabie**

** ESIGETEL, 1, rue du port de Valvins, F-77210 Avon*

dan.istrate@esigetel.fr

*** CReSTIC, Université de Reims, Moulin de la Housse, F-51687 Reims*

valeriu.vrabie@univ-reims.fr

RÉSUMÉ. *Un système de reconnaissance automatique du locuteur basé sur la fusion de deux systèmes d'identification des locuteurs et un système de vérification de l'identité est proposé. L'utilisateur prononce deux séries de mots-clés. Le premier module basé sur la transformée en ondelettes segmente en temps réel la succession des mots-clés. Les systèmes d'identification basés sur GMM utilisent les mots-clés pour identifier le locuteur (indépendamment du texte) et les mots (indépendamment des locuteurs). En utilisant les valeurs de vraisemblances obtenues par chaque système, deux listes classées de locuteurs et de mots-clés sont obtenues qui par fusion permettent d'identifier un couple Locuteur/Mot-clé. Un système de vérification du locuteur est par la suite appliqué sur le deuxième mot-clé. Une évaluation de chaque module, du système global ainsi que de l'influence du bruit est proposée.*

ABSTRACT. *This paper presents an automatic speaker recognition system based on the fusion of a speaker identification and an identity verification system. Two passwords are required for each user. The proposed system firstly performs a segmentation of the words based on a wavelet transform algorithm. The identification systems are GMM-based and use the first password in order to identify both speaker (text independent) and words (speaker independent). Each system provides an ordered list of likelihoods for both speakers and words which is combinated in order to identify a pair Speaker/Password. An identity verification system is applied on the second password. This paper evaluates each system, the global system and the noise influence.*

MOTS-CLÉS : *reconnaissance du locuteur, reconnaissance des mots, fusion de décision, GMM/UBM, transformée en ondelettes.*

KEYWORDS: *speaker recognition, word recognition, decision fusion, GMM/UBM, wavelet transform.*

DOI:10.3166/ISI.15.2.11-27 © 2010 Lavoisier, Paris

RSTI - ISI – 15/2010. Traitement de données complexes, pages 11 à 27

1. Introduction

L'utilisation de la voix comme modalité biométrique offre l'avantage d'être bien acceptée par les utilisateurs, quelle que soit leur culture. La modalité biométrique vocale n'est pas intrusive, demande un effort réduit de la part de l'utilisateur et a un coût faible. Les performances des systèmes biométriques basés sur la voix sont moins efficaces que les autres modalités biométriques mais représentent une modalité plus acceptable par l'utilisateur et, est la seule qui peut être utilisée par téléphone.

Il est bien connu que les performances des systèmes RAL se dégradent facilement si les conditions d'acquisition diffèrent entre la phase d'apprentissage et de test (bruit ambiant, etc.) Une des méthodes qui peut être utilisée pour améliorer les performances de ces systèmes est de fusionner les diverses informations portées par le signal de parole. Plusieurs études sur la fusion de l'information ont été menées pour améliorer les performances de la RAL (Higgins *et al.*, 2001, Mami, 2003, Kinnunen *et al.*, 2004). Toutefois, les résultats sont moins performants par rapport aux systèmes biométriques utilisant d'autres modalités (empreintes digitales, iris, visage, etc.).

Dans cet article, une nouvelle approche de fusion est proposée en utilisant deux types d'informations contenues dans le signal de parole : le locuteur (qui parle ?) et les mots prononcés (ce qui a été dit ?) (Chenafa *et al.*, 2008). L'objectif de cette méthode est d'identifier un couple *locuteur/mot de passe* correspondant au premier signal de test. Cet étape se fait en combinant deux systèmes d'identification basés sur le rapport de vraisemblance ; un système d'identification des locuteurs (indépendant du texte) et un système d'identification des mots (indépendant des locuteurs). L'identité du locuteur identifié par la fusion des deux systèmes d'identification est par la suite vérifiée par un système de vérification classique utilisant un second signal de test pour confirmer ou infirmer le locuteur identifié. Dans des situations pratiques, les deux signaux de test peuvent être considérés comme un seul signal segmenté en deux parties automatiquement. Le système proposé améliore les résultats en termes de taux d'égal erreur par rapport à un système d'identification de l'état de l'art dépendant du texte. Les différentes expérimentations présentées dans cet article ont été réalisées en utilisant la plate-forme ALIZE développée par le laboratoire LIA (Bonastre *et al.*, 2005).

La section 2 présente brièvement les outils statistiques utilisés (GMM, MFCC, LFCC). La section 3 est dédiée à l'architecture du système, la section 4 illustre les résultats expérimentaux et la dernière section est consacrée à la conclusion et aux perspectives.

2. La reconnaissance du locuteur

La figure 2 montre la structure d'un système de RAL. Ce système fonctionne en deux phases (apprentissage et reconnaissance) et peut être utilisé pour les deux modes : identification et vérification.

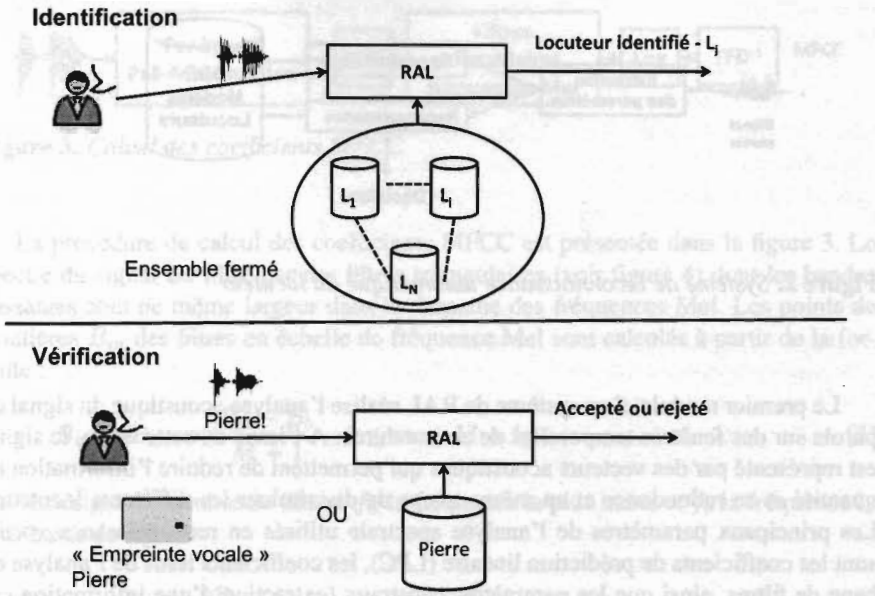


Figure 1. *Vérification et identification du locuteur*

Dans les systèmes d'identification, un locuteur inconnu est comparé au N locuteurs connus stockés dans la base de données et la meilleure vraisemblance représente la décision de la reconnaissance. C'est une reconnaissance en ensemble fermé, le système proposant un locuteur parmi ceux de la base. Le rejet d'un locuteur inconnu est très difficile à mettre en place parce que cela nécessite l'introduction d'un seuil de rejet ; dans ce cas on parle d'un ensemble ouvert.

Dans les systèmes de vérification, une identité est proclamée par un locuteur et le système compare sa voix au modèle correspondant à l'identité proclamée. Si la vraisemblance dépasse un seuil prédéfini, le locuteur est accepté, sinon, il est rejeté (figure 2).

Pour chaque système, deux modes peuvent être distingués : dépendant du texte et indépendant du texte. Les systèmes dépendant du texte réalisent les modèles des locuteurs sur le même texte que celui qui est prononcé en phase de reconnaissance par rapport à un système indépendant du texte qui ne connaît pas le texte prononcé par la personne à reconnaître.

La phase d'apprentissage du RAL consiste dans la création d'un modèle représentant les caractéristiques de la voix du locuteur qui seront utilisées par la suite dans la phase de test pour rendre une décision sur l'identité du locuteur inconnu.

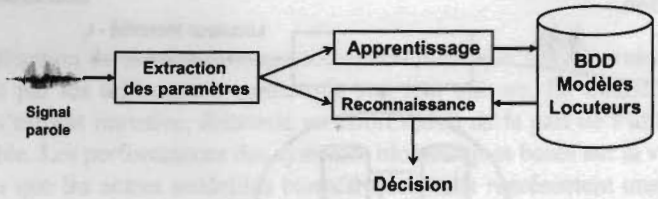


Figure 2. *Système de reconnaissance automatique du locuteur*

Le premier module d'un système de RAL réalise l'analyse acoustique du signal de parole sur des fenêtres temporelles de courte durée. A l'issue de cette étape, le signal est représenté par des vecteurs acoustiques qui permettent de réduire l'information en quantité et en redondance et en même temps de discriminer les différents locuteurs. Les principaux paramètres de l'analyse spectrale utilisés en reconnaissance vocale sont les coefficients de prédiction linéaire (LPC), les coefficients issus de l'analyse en banc de filtres, ainsi que les paramètres cepstraux (extraction d'une information caractéristique à partir du spectre du signal) (Furui, 1981, Hermansky, 1990). Les caractéristiques utilisées dans les systèmes actuels de reconnaissance du locuteur sont soit les paramètres cepstraux linéaires LFCC (*Linear Frequency Ceptrum Coefficients*), soit les paramètres cepstraux de type MFCC (*Mel Frequency Ceptrum Coefficients*), éventuellement complétés par la variation temporelle à court terme de ces mêmes paramètres (première et deuxième dérivée, Δ et $\Delta\Delta$).

La modélisation du locuteur peut être réalisée soit en utilisant des techniques statistiques : mélange de distributions de Gauss (GMM), chaînes de Markov cachées (HMM), soit des techniques connexionnistes comme les réseaux de neurones.

2.1. *MFCC (Mel frequency cepstral coefficients)*

Les coefficients MFCC sont des coefficients cepstraux très souvent utilisés en reconnaissance automatique de la parole. Le calcul des paramètres MFCC utilise une échelle fréquentielle non linéaire qui tient compte des particularités de l'oreille humaine (Furui, 1981).

L'échelle de fréquence Mel (B) est définie par :

$$B(f) = 2595 \log\left(1 + \frac{f}{700}\right), \qquad [1]$$

où f représente la fréquence en Hz et $B(f)$ la fréquence suivant l'échelle de fréquence Mel.

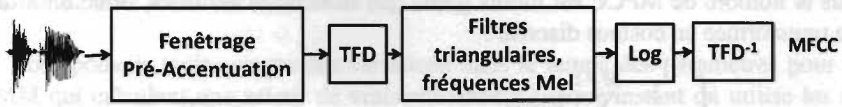


Figure 3. *Calcul des coefficients MFCC*

La procédure de calcul des coefficients MFCC est présentée dans la figure 3. Le spectre du signal est filtré par des filtres triangulaires (voir figure 4) dont les bandes passantes sont de même largeur dans le domaine des fréquences Mel. Les points de frontières B_m des filtres en échelle de fréquence Mel sont calculés à partir de la formule :

$$B_m = B_1 + m\frac{B_h - B_b}{M+1} \quad 0 \leq m \leq M+1, \qquad [2]$$

où M désigne le nombre de filtres, f_h la fréquence la plus haute et f_b la fréquence la plus basse du signal.

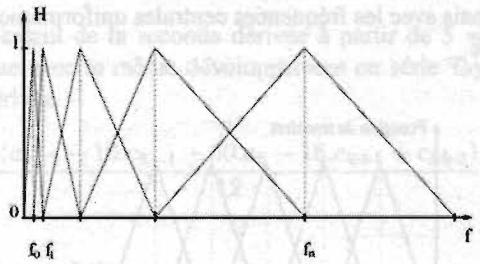


Figure 4. *Filtres en fréquences Mel*

Dans le domaine fréquentiel, les points f_m discrets correspondants sont calculés d'après :

$$f_m = \left(\frac{N}{F_s}\right) B^{-1}\left(B_b + m\frac{B_h - B_b}{M+1}\right), \qquad [3]$$

où $B^{-1}(i)$ désigne la fréquence correspondante à la fréquence i de l'échelle Mel, $B_i^{-1} = 700(10^{\frac{i}{2595}} - 1)$

Les coefficients cepstraux de fréquence en échelle Mel (MFCC) peuvent être obtenus par une transformée de Fourier inverse à partir des coefficients en sortie des filtres.

Mais le nombre de MFCC est moins grand que le nombre de filtres, donc on utilise une transformée en cosinus discrète :

$$c(n) = \begin{cases} \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} E(m) & , n = 0 \\ \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} E(m) \cos\left(\frac{\pi n(m+\frac{1}{2})}{M}\right) & , 0 \leq n < M \end{cases} \quad [4]$$

Finalement nous obtenons une suite de vecteurs acoustiques qui représentent spectralement le signal temporel en entrée.

2.2. *LFCC (Linear frequency cepstral coefficients)*

Les coefficients LFCC sont calculés de la même manière que les MFCC, mais à la différence que les fréquences des filtres sont uniformément réparties sur l'échelle linéaire des fréquences et non plus sur une échelle Mel. Les filtres utilisés sont aussi des filtres triangulaires mais avec les fréquences centrales uniformément reparties comme la figure 5 le présente.

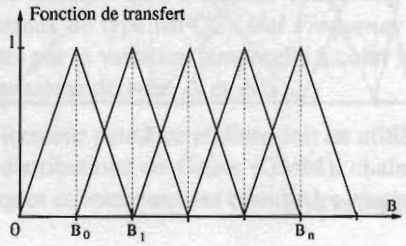


Figure 5. *Filtres uniformes pour LFCC*

2.3. *Énergie*

Comme paramètre acoustique, on utilise l'énergie logarithmique du signal qui est définie comme suit :

$$E = \ln\left(\sum_{i=0}^{N} s_i^2\right) \quad [5]$$

où N est le nombre d'échantillons du signal, s_i l'échantillon i du signal.

2.4. *La dérivée première et seconde ($\Delta, \Delta\Delta$)*

Pour pouvoir tenir compte des variations dans le temps des paramètres pour les GMM qui calculent une valeur de vraisemblance à chaque instant on utilise les dérivées de ceux-ci. La dérivé d'un paramètre acoustique est la mesure de sa variation en temps. Comme la fonction de variation des paramètres acoustiques est inconnue et seulement des valeurs à des instants précis sont connues, le calcul de la première dérivée et de la deuxième dérivée se fait par une approximation.

Comme la fonction de variation des paramètres acoustiques est connue seulement en des instants précis, le calcul de la dérivée première doit être fait par approximation (Press *et al.*, 2002). Ceci nécessite de connaître au moins 2 valeurs de la fonction autour du point concerné, les calculs se simplifient en utilisent 5 valeurs régulièrement espacées. Les 2 valeurs précédant la valeur courante c_k sont c_{k-2} et c_{k-1}, les 2 valeurs suivantes c_{k+1} et c_{k+2}. La formule d'approximation de la dérivée première [6] s'obtient simplement à partir de la décomposition en série de Taylor de la fonction.

$$\Delta c_k = \frac{-(c_{k+2} - c_{k-2}) + 8.(c_{k+1} - c_{k-1})}{12} \quad [6]$$

La formule de calcul de la seconde dérivée à partir de 5 valeurs régulièrement espacées est obtenue avec le même développement en série Taylor que celui utilisé pour la première dérivée.

$$\Delta\Delta c_k = \frac{-(c_{k-2} - 16.c_{k-1} + 30.c_k - 16.c_{k+1} + c_{k+2})}{12} \quad [7]$$

2.5. *Reconnaissance statistique*

Il existe deux familles d'algorithmes en RAL : les méthodes déterministes (comparaison dynamique et quantification vectorielle) et statistiques (modèles à mélange de distributions de Gauss (GMM) et modèles de Markov cachés (HMM), les dernières étant les plus utilisées dans ce domaine. Nous avons choisi d'utiliser un système à base de GMM utilisant un modèle d'adaptation globale appelé modèle du monde UBM (*Universal Background Model*) pour la classification des locuteurs. Ce choix est basé sur plusieurs facteurs : la modélisation par GMM/UBM est très flexible par rapport au type du signal et les GMM offrent un bon compromis entre les performances et la complexité du calcul.

Les modèles GMM consistent à supposer que la distribution des données peut être décrite comme une somme pondérée de densités de Gauss multidimensionnelles (Reynolds, 1995). Cette modélisation est classique dans le domaine de la reconnaissance des formes car elle correspond à une situation où les données appartiennent à un ensemble de classes distinctes, avec une probabilité d'appartenance propre à chaque classe. Le cas particulier considéré ici est celui où dans chaque classe les données suivent une loi gaussienne. Ce choix tient essentiellement au fait que la loi de Gauss

appartient à une famille de distributions dites exponentielles pour lesquelles le problème de l'identification des composantes du mélange se trouve simplifié (Bimbot *et al.*, 2004). La densité de probabilité gaussienne a la forme de l'équation :

$$f_m(\vec{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det(C_m)}} e^{\left[-\frac{1}{2}(\vec{x}-\vec{\mu_m})^t C_m^{-1}(\vec{x}-\vec{\mu_m})\right]}, \quad [8]$$

où :

- $\vec{x}$ est le vecteur des distributions à modéliser
- $\vec{\mu_m}$ est le vecteur moyen du vecteur $\vec{x}$
- C_m la matrice de covariance du vecteur $\vec{x}$
- d est la dimension du vecteur $\vec{x}$
- C^{-1} est l'inverse de la matrice C et $.^t$ dénote la transposée.

Lors de la phase d'apprentissage, tous les vecteurs acoustiques d'apprentissage sont utilisés pour déterminer le poids correspondant à chacune des M gaussiennes, le vecteur acoustique moyen et la matrice de covariance de chacune des gaussiennes. Pour chacune des gaussiennes ($1 \leq m \leq M$) de la classe ω_k, les paramètres suivants sont ceux qui caractérisent le modèle GMM de la classe :

- le nombre de paramètres acoustiques utilisés d (toujours le même)
- les poids de chaque gaussienne $\pi_{k,m}$ qui respecte la condition : $\sum_{m=1}^{M} \pi_{k,m} = 1$
- les vecteurs moyens $\mu_{k,m}$
- les matrices de covariance $C_{k,m}$

L'apprentissage a pour but d'estimer les paramètres des gaussiennes qui composent le modèle à partir des vecteurs acoustiques. L'apprentissage d'une classe se décompose en deux étapes successives : tout d'abord l'obtention de valeurs approximatives des paramètres des gaussiennes de la classe, ensuite l'optimisation des valeurs de ces paramètres par un algorithme de type EM (*Expectation Maximisation*). L'algorithme EM fait intervenir des variables latentes que l'on ne peut observer directement. Dans notre cas, chaque vecteur $\vec{x}$ est décrit non seulement par les d paramètres acoustiques (valeurs mesurées) mais aussi par le sous-ensemble S_i (défini par un centroïde) auquel il se rattache. Il va maximiser la vraisemblance de façon itérative, mais le vecteur $\vec{x}$ sera maintenant rattaché aux M sous-ensembles S_i avec une probabilité particulière, sans que l'on puisse déterminer à quel sous-ensemble S_i il appartient réellement. C'est ce paramètre que l'on qualifie de donnée *cachée* ou *latente*.

Pendant la phase de classification, on doit déterminer la classe ω_l la plus probable à partir du calcul de la vraisemblance, pour le vecteur acoustique $\vec{x}$ obtenu à l'instant t, et pour chacune des classes de sons ω_k ($1 \leq k \leq K$) :

$$p(\vec{x} \mid \omega_k) = \sum_{m=1}^{M} \pi_{k,m} \cdot \frac{1}{(2\pi)^{\frac{d}{2}} |C_{k,m}|^{\frac{1}{2}}} \cdot e^{\left[-\frac{1}{2} (\vec{x} - \mu_{k,m})^t \cdot C_{k,m}^{-1} \cdot (\vec{x} - \mu_{k,m}) \right]} \quad [9]$$

Un signal à tester est transformé dans une suite de vecteurs acoustiques $X = [\vec{x}_1, \vec{x}_2, \ldots]$ qui ont d paramètres acoustiques. Il appartiendra avec le maximum de vraisemblance à la classe ω_l pour laquelle $p(X \mid \omega_l)$ est maximale, conformément à l'équation :

$$p(X \mid \omega_l) = \max_{k=1}^{K} \left(p(X \mid \omega_k) \right) \quad [10]$$

Dans le cas présent comme le nombre d'enregistrements disponibles pour chaque locuteur à modéliser (et donc à reconnaître) est faible, nous utilisons une méthode basée sur un modèle du monde UBM. En effet les caractéristiques de la voix d'un nombre important de locuteurs sont modélisées sous la forme du modèle UBM en utilisant l'algorithme EM. La création des modèles des clients est faite en 2 étapes : apprentissage d'un GMM par EM sur les données spécifiques du locuteur suivi d'une adaptation MAP du modèle UBM par celui-ci (figure 6). L'adaptation MAP concerne la moyenne et/ou la covariance des distributions de Gauss des modèles des clients et du modèle du monde (UBM). Plus généralement, seule l'adaptation de la moyenne est effectuée ; la moyenne résultante est une somme entre la moyenne obtenue par EM sur les données du client multiplié par une constante α ($\alpha < 1$) et la moyenne de l'UBM multipliée par $1 - \alpha$.

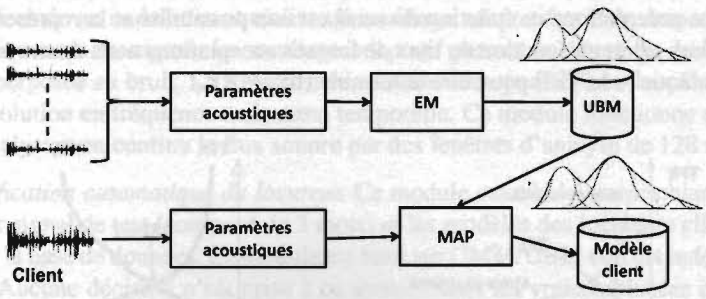


Figure 6. *Apprentissage d'un modèle client*

2.6. *Évaluation des performances des systèmes RAL*

Avant de définir les taux qui permettent d'évaluer un système de reconnaissance en général nous devons définir les deux types d'erreurs d'un tel système (figure 7) :

Faux rejet = Le client est rejeté alors que l'identité proposée est la sienne (« Joe prétend être Joe mais le système d'authentification le rejette »)

Fausse acceptation = Le client est accepté alors que l'identité proposée n'est pas la sienne (« Jane prétend être Joe mais le système d'authentification l'accepte »)

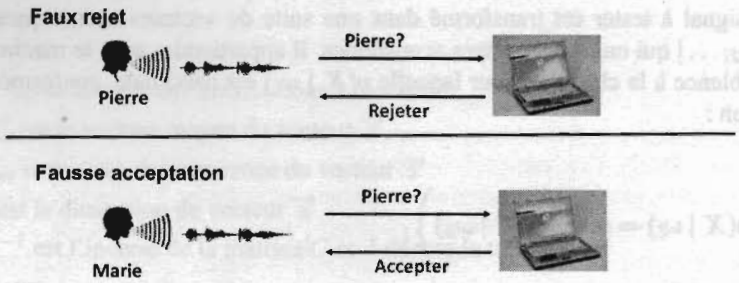


Figure 7. *Définition faux rejet/fausse acceptation*

Trois taux d'erreurs peuvent être définis comme suit :

Taux de faux rejets (TFR) = Pourcentage des clients présentés devant le système qui sont rejetés par erreur

Taux de fausses acceptations (TFA) = Pourcentage des imposteurs acceptés par le système

Taux d'égal erreur (TEE) = Le taux de faux rejets quand celui-ci et celui des fausses acceptations sont égaux.

Pour évaluer un système de RAL la représentation des courbes des faux rejets et de fausses acceptations en fonction du seuil est une possibilité et la représentation du taux de faux rejets en fonction du taux de fausses acceptations nous donne une courbe ROC sur laquelle le TEE peux être déterminé (figure 8).

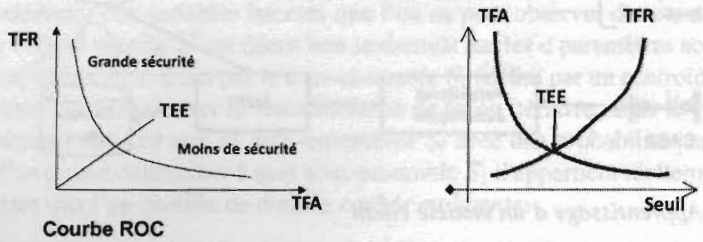


Figure 8. *Courbe ROC*

3. Système proposé

Dans cet article une nouvelle méthode de RAL basée sur la fusion de décision est proposée, l'architecture globale du système étant illustrée dans la figure 9. Cette méthode dans une première étape combine un système d'identification du locuteur avec un système de reconnaissance des mots isolés. Une fois un locuteur identifié, dans une deuxième étape on vérifie son identité en utilisant la deuxième partie du texte prononcé. Chaque module composant le système est décrit par la suite.

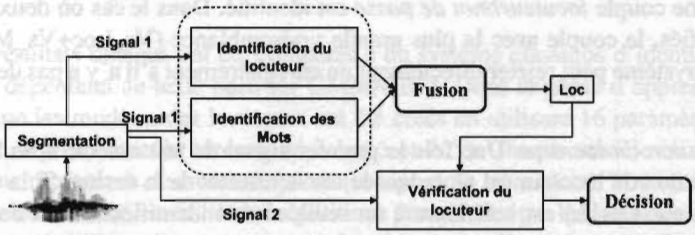


Figure 9. *Architecture globale du système*

Segmentation automatique. Nous supposons que le microphone du système est ouvert en permanence donc nous allons utiliser un système capable à détecter l'apparition du signal de parole et de le segmenter dans des composants. Nous avons utilisé l'algorithme de segmentation automatique déjà développé pour la reconnaissance des sons et des expressions de détresse (Istrate *et al.,* 2006). Ce système est basé sur une transformée en ondelettes et réalise un seuillage adaptatif principalement sur les hautes fréquences du signal. Le seuillage adaptatif permet la détection de la parole même superposée au bruit. La transformée en ondelettes permet à la fois d'avoir une bonne résolution en fréquence mais aussi temporelle. Ce module fonctionne en temps réel en analysant en continu le flux sonore par des fenêtres d'analyse de 128 ms.

Identification automatique du locuteur. Ce module calcule la vraisemblance entre le premier signal de test (composé de 3 mots) et les modèles des locuteurs clients stockés dans la base de données. Le module est basé sur GMM/UBM et il est indépendant du texte. Aucune décision n'est prise à ce niveau, mais les vraisemblances du signal inconnu par rapport à tous les modèles des clients obtenus sont triées et sauvegardées pour la phase de fusion.

Identification automatique des mots isolés. Le même signal, constitué de trois mots, est également utilisé par le module d'identification des mots, indépendant du locuteur. Ce système correspond à la fusion de trois sous-modules d'identification, un pour chaque mot du premier signal. Les résultats de chaque sous-module sont utilisés afin de proposer une ou plusieurs combinaisons de mots de passe possibles. Seuls

les cinq premiers résultats de chaque module sont combinés. Plusieurs tests on été effectués pour déterminer le nombre optimum de mots à utiliser dans la fusion.

Fusion de décision. Après le tri des vraisemblances du signal inconnu (composé de 3 mots de passe) calculées par rapport aux modèles des locuteurs clients et par rapport aux modèles des mots, un premier test consiste à comparer le locuteur le plus probable avec les cinq premiers mots de passe identifiés. Si son mot de passe est trouvé alors un couple *locuteur/mot de passe* correspondant au premier signal de test est identifié. Un second test est effectué en comparant le mot de passe le plus probable avec les cinq premiers locuteurs identifiés, s'il s'agit d'une série de mots-clés d'un d'entre eux alors un deuxième couple *locuteur/mot de passe* est identifié. Dans le cas où deux couples sont identifiés, le couple avec la plus grande vraisemblance (Vs_Loc+Vs_MDP) est retenu. Le système peut rejeter directement un enregistrement s'il n'y a pas de couples identifiés.

Vérification du locuteur. Une fois le premier signal de test associé à un locuteur, une vérification du locuteur est alors lancée sur le résultat de la fusion. Si la vraisemblance obtenue (Vs_L) est inférieure à un seuil, alors l'identification est confirmée, sinon c'est un rejet et le système considère la personne comme un imposteur. La valeur du seuil est à choisir en fonction du compromis recherché : réduire le nombre de faux rejets ou le nombre de fausses acceptations.

4. Expérimentations

4.1. *Base de données*

Nous avons enregistré un corpus de parole spécifique à notre application de contrôle d'accès. Le corpus contient les enregistrements de 20 mots isolés répétés 5 fois par chaque locuteur et le nom et le prénom de chaque locuteur répétés 10 fois. Le nombre des locuteurs est de 58 dont 53 hommes et 5 femmes (au total 4, 28 heures d'enregistrement). Les fichiers sont stockés dans un format « .wav » avec une fréquence d'échantillonnage de $f_s = 44.1$ kHz.

4.2. *Paramétrisation*

La modélisation a été réalisée en utilisant des paramètres MFCC pour la modélisation des mots de passe et des paramètres LFCC pour les locuteurs. Les filtres uniformes des paramètres acoustiques LFCC permettent une caractérisation meilleure du spectre complet de la voix des locuteurs. Toutes les 8 ms le signal est caractérisé par un vecteur acoustique composé de 16 coefficients cepstraux suivi de l'énergie et de leurs dérivées (Δ, $\Delta\Delta$).

4.3. *Modèle du monde (UBM)*

Pour l'adaptation des modèles des locuteurs, nous avons testé différentes tailles de l'UBM (nombre de distributions de Gauss) : 64, 128, 256 et 512. Il est à noter que l'UBM est créé en utilisant tous les enregistrements de la base de données. Le meilleur compromis performance/temps de calcul a été obtenu en utilisant 128 gaussiennes pour le modèle UBM.

4.4. *Système de référence*

Les résultats obtenus ont été comparés à un système classique d'identification du locuteur dépendant du texte basé sur GMM/UBM. Dans la phase d'apprentissage de ce système les modèles des locuteurs ont été créés en utilisant 16 paramètres LFCC, l'énergie, la première et la seconde dérivée (Δ et $\Delta\Delta$). Chaque modèle est créé en utilisant trois répétitions des mots de passe (composé de trois mots). Toutefois, la phase de reconnaissance utilise toutes les répétitions prononcées par les locuteurs imposteurs et les deux répétitions des mots de passe prononcées par les locuteurs clients.

4.5. *Données d'apprentissage et de test*

Pour les deux systèmes d'identification (locuteur et mots de passe), le premier signal est composé de trois mots combinés d'une façon unique de l'ensemble des 21 mots possibles. Ce signal est utilisé pour la phase d'apprentissage et de test. Comme un deuxième signal est utilisé par le système de vérification, nous avons testé deux types de vérifications : dépendante du texte utilisant le nom prénom de l'utilisateur comme deuxième signal et indépendante du texte utilisant un mot parmi les 21 disponibles comme deuxième signal. La base de données a été divisée en deux groupes : 49 clients et 9 imposteurs. Afin d'évaluer le système proposé, le nombre des tests positifs et négatifs est identique.

– Le système d'identification du locuteur (indépendant du texte) utilise 3 enregistrements de 17 mots des 49 clients pour la phase d'apprentissage ($\approx$ 29 minutes). Pour la phase de test, le système utilise 2 enregistrements des 20 mots de 49 locuteurs clients et 5 enregistrements des 20 mots de 8 locuteurs imposteurs (792 essais).

– Le système d'identification des mots (indépendant du locuteur) utilise 3 enregistrements de 49 locuteurs clients pour la phase d'apprentissage ($\approx$ 29 minutes). Pour la phase de test, le système utilise 2 enregistrements des 49 clients et tous les enregistrements des imposteurs (792 essais).

– Le système de vérification utilise 8 enregistrements de la deuxième série de mots-clés de chaque client pour la phase d'apprentissage ($\approx$ 7 minutes) et 2 enregistrements des 49 locuteurs clients ainsi que tous les enregistrements des 8 locuteurs imposteurs pour la phase de test.

– Le système de référence utilise pour la phase d'apprentissage 3 enregistrements de 3 mots des 49 locuteurs clients (≈ 8 minutes). Pour la phase de test, nous avons utilisé 2 enregistrements de 3 mots (utilisés en apprentissage) des 49 locuteurs clients et 3 enregistrements des 8 locuteurs imposteurs (576 essais)

4.6. *Résultats et discussion*

4.6.1. *Evaluation sans bruit et avec une segmentation idéale*

Le tableau 4.6.1 présente les performances de chaque étape du système comparé au système de référence en termes de taux d'égale erreur (EER).

Systèmes	Paramètres	EER (%)	
Système de Référence indépendant du texte	16 LFCC+Énergie+ΔΔ	4,76 %	
Fusion entre le système d'identification des locuteurs et des mots	Locuteur : 16 LFCC+Énergie+ΔΔ Mots : 16 MFCC+Énergie+ΔΔ	0,38 %	
Vérification du locuteur après fusion	16 LFCC+Énergie+ΔΔ	**Dep.**	**Indep.**
		0,13 %	**0,26 %**

Tableau 1. *Performances des différents systèmes*

La première étape du système proposé (fusion) améliore le taux d'égale erreur en passant de 4,76 % du système de référence à 0,38 %. La deuxième étape du système (vérification après fusion) améliore les résultats de 31 % par rapport à la première étape en mode indépendant du texte (EER de 0,26 %) et de 34 % en mode dépendant du texte (EER de 0,13 %).

4.6.2. *Évaluation en présence du bruit*

Le système global a été testé en présence du bruit en utilisant une segmentation automatique. Deux types de bruit ont été investigués : ventilateur (bruit stationnaire) et pas (bruit impulsionnel). Les fichiers contenant la suite des trois mots de passe et du nom et prénom de la personne ont été bruités par une addition pondérée pour obtenir le rapport signal bruit (RSB) désiré. Trois RSB ont été évalués : 10, 20 et 40 dB.

Pour obtenir le RSB désiré, l'énergie du bruit doit être adaptée. Les différentes étapes de génération des fichiers de tests sont alors :

– Le calcul de l'énergie moyenne par échantillon du signal s_i avec :

$$(E_{\text{signal utile}})_{dB} = 10 \cdot \log\left(\frac{1}{N}\sum_{i=0}^{N-1} s_i^2\right) \quad [11]$$

– La détermination du niveau de l'énergie moyenne du bruit nécessaire pour obtenir le rapport signal sur bruit (RSB) désiré sachant que :

$$E_{\text{bruit nécessaire}} = 10^{\frac{\left(E_{\text{signal utile}}\right)_{dB} - RSB}{10}} \quad [12]$$

– Le calcul de l'énergie moyenne par échantillon du bruit b_i suivant la formule :

$$E_{\text{bruit}} = \frac{1}{N} \sum_{i=0}^{N-1} b_i^2 \quad [13]$$

– Le calcul du coefficient de multiplication de chaque échantillon de bruit b_i en vue d'obtenir le RSB désiré :

$$\text{Coeff} = \sqrt{\frac{E_{\text{bruit nécessaire}}}{E_{\text{bruit}}}} \quad [14]$$

Si la phase de segmentation a indiqué moins ou plus de 4 segments (3 mots de passe et le nom-prénom) nous avons considéré comme erreur de segmentation et le test a été arrêté.

Les erreurs du système global sont présentées dans le tableau 2. Le taux d'erreur de segmentation est calculé comme étant le rapport entre le nombre de fichiers pour lesquels le nombre de segments est $\neq 4$ et le nombre total de fichiers. Nous observons que la segmentation automatique fonctionne très bien en présence du bruit stationnaire parce que le seuil s'adapte aux variations lentes du bruit. Pour le bruit de pas qui est impulsionnel, les erreurs de segmentation apparaissent en-dessous de 20 dB de RSB. Ces taux d'erreurs ne tiennent pas compte du fait que la segmentation des mots peut être imparfaite (une partie du mot étant coupée).

Bruit	RSB	Erreur segmentation	Fusion		Global	
			TFR	TFA	TFR	TFA
Ventilateur	40 dB	0 %	9.3 %	9.8 %	11.9 %	1.9 %
	20 dB	0 %	5.7 %	5.9 %	5.7 %	2.1 %
	10 dB	0 %	9.9 %	10.3 %	10.7 %	2.1 %
Pas	40 dB	0 %	16.1 %	8.9 %	17 %	1 %
	20 dB	3.5 %	76.4 %	0 %	76.4 %	0 %

Tableau 2. *Performances du système en présence du bruit*

Les performances du système de reconnaissance du locuteur sont présentées sous forme du taux de faux rejets (TFR) et du taux de fausses acceptations (TFA). Dans le tableau 2 les performances après la fusion des deux systèmes d'identification et les performances globales sont présentées. Nous observons que l'utilisation du système de vérification améliore considérablement le taux de fausses acceptations. La présence

du bruit et l'utilisation de la segmentation automatique qui introduit des erreurs (partie des mots coupés ou rajout des zones de signal contenant seulement le bruit) réduit les performances du système par rapport aux conditions idéales. Le système proposé a des performances acceptables en présence du bruit stationnaire de ventilateur jusqu'au RSB de 10 dB. La présence du bruit de pas qui est impulsionnel génère autant d'erreurs de segmentation que d'erreurs de reconnaissance.

5. Application

Le système proposé de reconnaissance du locuteur a été conçu pour une application de type contrôle d'accès dans un bâtiment ou dans une salle (par exemple la salle des systèmes informatiques). L'utilisateur n'a pas à appuyer sur un bouton avant de prononcer son mot de passe parce que le système a été conçu pour écouter en permanence l'environnement sonore. Les bruits qu'on peut rencontrer dans un bâtiment sont : des pas, des claquements de porte, de la parole, ventilateur/climatisation, etc. Ces bruits peuvent être divisés dans deux types : stationnaires (ventilateur, climatisation) et impulsionnels (pas, claquement de porte). Le système proposé a été testé avec un bruit de chaque type.

La segmentation automatique du flux de parole est effectuée en temps réel et la partie de reconnaissance du locuteur en temps différé mais en parallèle avec la segmentation elle-même. Cela veut dire qu'après la réception des trois premiers mots la première phase du système d'identification est lancé en même temps que le système enregistre la dernière composante du signal nécessaire pour la vérification. En conclusion, l'attente de l'utilisateur est approximativement égale au temps de vérification de la dernière étape qui se trouve en-dessous d'une seconde sur processeur actuel.

6. Conclusion et perspectives

Dans cette étude, nous avons présenté plusieurs expériences pour améliorer les performances des systèmes de reconnaissance automatique du locuteur. Dans un premier temps une expérience sur la fusion entre l'identification du locuteur et l'identification des mots prononcés est proposée. Nous montrons que le fait de modéliser l'ensemble des mots prononcés par le locuteur apporte des améliorations significatives par rapport au système de référence.

La deuxième expérience propose une vérification automatique du locuteur en prenant en entrée le résultat de la première expérience (locuteur identifié). Le but ici est de confirmer ou infirmer le résultat obtenu par la fusion (identification du locuteur et du mot prononcé). La deuxième expérience nous a permis de réduire le nombre d'imposteurs acceptés par le système et d'améliorer les résultats de la fusion. Ainsi le système global a de meilleures performances que le système de l'état de l'art (système de référence). Nous avons évalué le système en présence de deux types de bruit : un

impulsionnel et un autre stationnaire. Le système garde de bonnes performances en présence du bruit de ventilateur.

Nos études prochaines visent à rendre le système plus robuste par rapport aux bruits impulsionnels par une adaptation des modèles et par une amélioration de l'algorithme de segmentation automatique. Une implémentation temps réel du système proposé est en cours.

7. Bibliographie

Bimbot F., Bonastre J.-F., Fredouille C., Gravier G., Chagnolleau I.-., Meignier S., Merlin T., Garciya J., Delacrétaz D.-., Reynolds D., « A tutorial on text-independent speaker verification », *EURASIP Journal on Applied Signal Processing*, vol. 4, p. 430-451, 2004.

Bonastre J.-F., Wils F., Meignier S., « Alize, a free toolkit for speaker recognition », *Proc. Intl. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, vol. 1, p. 737-740, 2005.

Chenafa M., Istrate D., Vrabie V., Souidene W., « Reconnaissance automatique du locuteur par fusion de décision », *Actes Majestic 2008*, Éditions Hermès, Paris, Marseille, 29-31 October, 2008.

Furui S., « Cepstral Analysis Techniques for Automatic Speaker Verification », *IEEE Trans. Acoust. , Speech, Signal Processing*, vol. 29, p. 254-272, 1981.

Hermansky H., « Perceptual linear predictive (PLP) analysis of speech », *J. Acoust. Soc. Am.*, vol. 87, p. 1738-1752, 1990.

Higgins J. E., Damper R. I., Harris C. J., « Information fusion for subband-HMM speaker recognition », *International Joint Conference on Neural Networks*, p. 1504-1509, 2001.

Istrate D., Castelli E., Vacher M., Besacier L., Serignat J., « Information Extraction From Sound for Medical Telemonitoring », *IEEE Transactions on TITB*, vol. 10, p. 264-274, April, 2006.

Kinnunen T., Hautamäki V., Fränti P., « Fusion of Spectral Feature Sets for Accurate Speaker Identification », *Proc. Intl. Conf. Speech and Computer*, vol. 1, p. 361-365, 2004.

Mami Y., Reconnaissance de locuteurs par localisation dans un espace de locuteur de reference, Thèse de doctorat, ENST Paris, France, 2003.

Press W. H., Flannery B. P., Teukolsky S. A., Vetterling W. T., *Numerical Recipes in C ; The Art of scientific Computing ;The second Edition*, ISBN 0-521-43108-5, Cambridge University Press, 2002.

Reynolds D., « Speaker identification and verification using Gaussian mixture speaker models », *Speech Communication*, vol. 17, p. 91-108, 1995.

2011 IEEE International Conference on Fuzzy Systems
June 27-30, 2011, Taipei, Taiwan

A Pervasive Multi-sensor Data Fusion for Smart Home Healthcare Monitoring

Hamid Medjahed
Dan Istrate
ESIGETEL, LRIT
Avon-Fontainebleau, France
Email: {hamid.medjahed,dan.istrate}@esigetel.fr

Jerome Boudy
Jean-Louis Baldinger
Bernadette Dorizzi
Telecom SudParis, EPH, Evry, France
Email: {jerome.boudy,jean-louis.baldinger}@it-sudparis.eu

***Abstract*—Today elderly people are the fastest growing segment of the population in developed countries, and they desire to live as independently as possible. But independent lifestyles come with risks and challenges. Medical in-home telemonitoring (and, more generally, telemedicine) is a solution to deal with these challenges and to ensure that elderly people can live safely and independently in their own homes for as long as possible. In this context we propose an automatic in-home healthcare monitoring system for several uses and to meet the needs identified above. The proposed telemonitoring system is a multimodal platform with several sensors that can be installed at home and enables us to have a full and tightly controlled universe of data sets. It integrates elderly physiological and behavioral data, the acoustical environment of the elderly, environmental conditions and medical knowledge. Each modality is processed and analyzed by specific algorithms. A data fusion approach based on fuzzy logic with a set of rules directed by medical recommendations, is used to fuse the various subsystem outputs. This multimodal fusion increases the reliability of the whole system by detecting several distress situations. In fact this fusion approach takes into account temporary sensor malfunction and increases the system reliability and the robustness in the case of environmental disturbances or material limits (Battery, RF range, etc.). The Fuzzy logic fusion methods brings high flexibility to the telemonitoring platform especially in combining modalities or adding other sensors. The proposed telemonitoring system will ensure pervasive in-home health monitoring for elderly people.**

Keywords: fuzzy logic, fuzzy control, multimodal data fusion, telemedicine, healthcare telemonitoring.

I. Introduction

France's population continues to age significantly and a recent study carried out by the French national institute of statistic and economic studies (INSEE) shows a new distribution of age classes in France. In fact, almost one in three people will be over 60 years in 2050, against one in five in 2005, and France will have over 10 million of people over 75 years and over 4 million of people over 85 years. Therefore, the aging population and the increase in life expectancy have led to new models of aging where technology can play a role in monitoring the quality of life, by detecting or even predicting adverse events and hence supporting independence. Automatic in-home telemonitoring of distress situations has been a common focus in Gerontechnology [1] because medical telemonitoring at home, is an interesting solution compared to health facility institutions for the elderly, since it offers medical surveillance in a familiar atmosphere for the patient.

To address these issues, researchers are developing technologies to enhance a resident's safety and monitor health conditions using sensors and other devices. Numerous projects are carried out in the world especially in Europe, Asia and North America, on the home healthcare telemonitoring topic. They aim for example to define a generic architecture for such telemonitoring systems [2], to conduct experiment of a remote monitoring system on a specific category of patients [3] (Insufficient cardiac heart, asthma, diabets, patients with Alzheimer's disease, or cognitive impairement, etc.), or to build smart apartments [4], sensors and alarm systems adapted to the healthcare telemonitoring requirements [5].

However, little research exists to motivate and guide such technology. For example, most monitoring systems use some form of learning method to discriminate between different types of normal and abnormal events. These algorithms require large amounts of training data that can be dificult to obtain, especially data describing abnormal events that are by definition scarce occurrences. The most crucial issue for all these systems is the lack of experimental data and information representing many situations and several person profiles. Most of these systems also take into account only one modality, like medical sensors (Blood pressure, pulse, oxymeter) or localization sensors (Infrared or contacts) to survey patient. But among established medical remote systems, there are few commercial solutions and business models.

In this paper we present the use of the multimodal system called EMUTEM (Environnement Multimodal pour la Télévigilance Médicale) [6] for distress situations detection. The proposed system was evaluated on two data bases: one recorded by our self and another one recorded in an experimental house by elderly people. The platform developed within this framework research manages a system consisting of:

- A set of microphones placed in all rooms of the elderly's house, that allow remote monitoring of the acoustical environment of the elderly (Anason [7]).
- A wearable device called RFpat [8] that can measure physiological data like ambulatory pulse heart rate, detect posture (standing/ sitting and laying), fall of the equipped person and his activity rate.

978-1-4244-7316-8/11/$26.00 ©2011 IEEE

- A set of infrared sensors that detect the person's presence in a portion of a given home part, his posture and also his movement (Gardien [8]). Plus a set of domotic sensors like contact sensors, temperature sensors and several other domotic sensors for environment conditions monitoring.

Subsystem data streams have been separately processed with suitable algorithms for abnormal situations detection. In order to maximize correct classification performance between normal and distress situations, data fusion over the different sensors types is studied.

The originality of this research is the combination of various modalities in the home, about its inhabitant and their surroundings. The new multimodal data fusion approach based on fuzzy logic allows high flexibility for the EMUTEM platform especially in combining modalities or adding other sensors. It constitutes an interesting benefit and impact for the elderly person sufering from loneliness. This work complements the stationary smart home environment in bringing to bear its capability for integrative continuous observation and detection of critical situations.

II. The Telemonitoring System

We define a smart environment as one with the ability to adapt the environment to the inhabitants and meet the goals of comfort and efficiency. In order to achieve these goals, our first aim was focused on providing such as environment. We consider our system as an intelligent agent, which perceives the state of the environment by using sensors and acts consequently with device controllers.

A. Environment Sensing and Data Collection

The experimental area is a surface of $20m^2$ in our laboratory which is arranged in two rooms with a technical area in order to evaluate and to supervise experiments. The hardware framework is reported in Figure 1. It integrates smart sensors (Infrared, change state sensors, audio, physiological,) linked to a PC. Microphones for audio monitoring are linked to the PC through an external sound card (in order to allow good signal to noise ratio independent from the PC), and can be interpreted

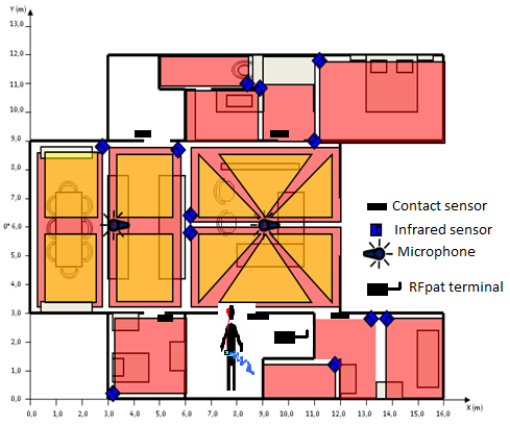


Fig. 1. In-home sensor disposal.

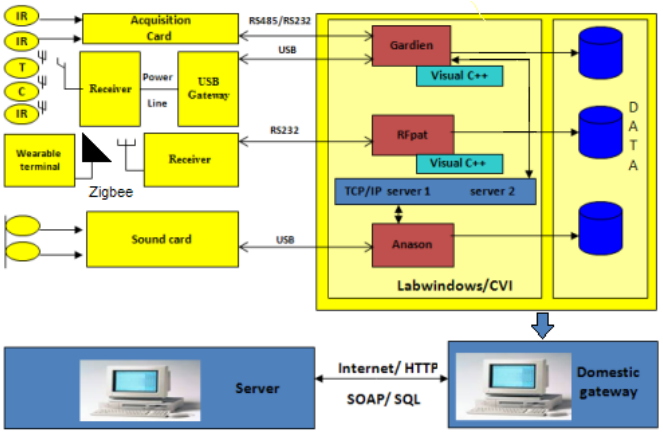


Fig. 2. EMUTEM software architecture.

as a single smart audio sensor achieved by Anason software. Infrared sensors are fixed on specific places of the house in walls and ceiling. They are in permanent communication via radio frequency communication, with one receiver, which is connected to the USB port of the PC. Change state sensors transmit also information to this receiver through radio frequency communication. The powerline is used to get back data from the receiver for software processing. The weareble device RFpat is caried by the elderly and continuously monitors his physiological data and emergency calls. It transmits these data to an indoor reception base station via ZigBee.

Figure 2 shows the software architecture of the multimodal platform EMUTEM. It provides a general user interface which encapsulates the Anason subsystem. It is implemented under LabWindows/CVI software and communicates with RFpat and Gardien sub-systems by a client-server model using TCP/IP and appropriate application protocols. Gardien is implemented in C++ and recovers data every 500 ms. RFpat is also implemented in C++ and receives data from a receiver every 30 s. The use of the inter-module communication through TCP/IP sockets allows each module (subsystem) to be run on a different computer, and to synchronize each telemonitoring modality channel.

The user can interact with the system via internet navigator and supervises the different applications. This feedback provides a significant help to the system manipulation and the system flexibility obtained through TCP/IP sockets communication allows adding other potential sensors such as a heart monitoring sensors (ECG). Data acquired from sensors are stored in the local computer directly as text files assigned to each modality. Data also could be exchanged using http or ftp protocol via web services technologies SOAP, and saved in a dedicated server.

III. Distress Situation Detection

Detecting and gathering data about the elderly person and his environment is the first step and one of the most fundamental tasks in building intelligent telemonitoring systems. With the increasing intelligence expectation, using multiple sensors

is the only way to obtain the required breadth of information, and fusing the outputs from these multiple sensors is often the only way to obtain the required depth of information when a single sensing modality is inadequate [9]. However, in our telemonitoring context different sensors use different physical principles, cover different information space, and generating data in different formats at different sampling rates. The obtained data have different resolution, accuracy, and reliability properties.

Based on those effects, the key to produce the required detection is to use the right method that properly fuses the provided data from various sources. This is what multimodal data fusion stands for. Therefore, typical multisensors data fusion methods are analyzed in this section, in seeking for a most generalizable and adaptable method.

A. Adapted Approaches for Multi-sensors Data Fusion

To select a suitable method for EMUTEM's multimodal data fusion module, we have probabilistic approaches whose performance can be reconsidered in our application for many reasons. The classical inference method quantitatively compares the probability that an observation can be attributed to a given assumed hypothesis. But it has the following major disadvantages [10], (1) difficulty in obtaining the density functions that describe observations used to classify the object, (2) complexities that arise when multivariate data are encountered, (3) its capability to assess only two hypotheses at a time, and (4) its inability to take direct advantage of a priori likelihood probabilities.

The Bayesian inference method [11] also has some weaknesses that prevent it from being used in our multimodal data fusion module. The key limits are: (1) difficulty in defining a priori probabilities, (2) complexities when there are multiple potential hypotheses and multiple conditionally dependent events, (3) mutual exclusivity required for competing hypotheses, and (4) inability to account for general uncertainty and to represent imprecision.

Even if Dempster-Shafer methods [12] use a general level of uncertainty, they cannot be the main data fusion method for two reasons: the difficulty to estimate mass function and their restrict domain of application.

The neural networks method [13] is not very well adapted to EMUTEM's Data fusion module because of drawbacks. First, the mapping mechanism is not well understood even if the network can provide the desired behavior. Second, the neural network method is, generally speaking, not suitable to work in a dynamic sensor configuration environment, because each sensor needs a unique input node and each possible sensor-set configuration needs to be specifically trained. Third, the complex architecture of neural networks prevents experts adding their knowledge easily.

SVM methods [14], despite their transit in the characteristics space which is disconnected from any physical reality, could fulfill the requirement of intelligibility because only support vectors are important in identifying margins between classes. However, it is necessary that boundaries between classes are rendered intelligible by a graphical way in the space of inputs. This vision must take into account an input space of any size even if greater than 3. In this case, the SVM identifies a large majority of learning examples as support examples. It means that an analyst should remember too many relevant individuals for the construction of boundaries between classes and this is impossible.

The fuzzy logic method [15] is the proposed way to meet these challenges of this multimodal data fusion application. According to the nature of data to process in EMUTEM platform, fuzzy logic is the well adapted approach for the telemonitoring decision. It deals with inaccuracy and uncertainty. It allows a great flexibility to combine several sensors.

IV. Implementation

A. Fuzzy Classifier

The main advantages of using fuzzy logic system are the simplicity of the approach and the capacity of dealing with the complex data acquired from the subsystems discribed previously in the second section. Fuzzy set theory offers a convenient way to do all possible combinations with these data. Fuzzy set theory is used in this system to determine the most likely distress situations that might occur for elderly persons in their home. The data fusion is carried out at three different levels: for sound/speech environement at the decision level, for smart home sensors system at the input data level and for the wearable physiological sensor at the representation level.

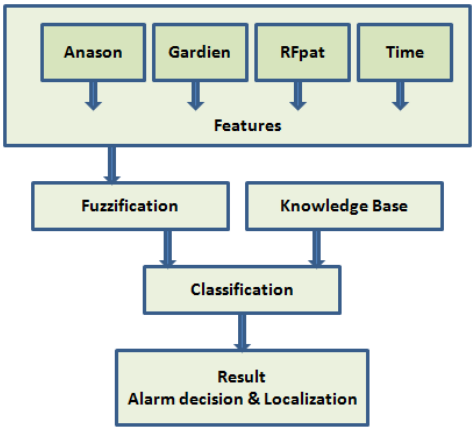


Fig. 3. Structure of the fuzzy classifier.

B. Input and Output fuzzification

The first step for implementing the fuzzy logic multimodal data fusion approach is the fuzzification of outputs and inputs of the fuzzy inference system (FIS) obtained from each subsystem.

From Anason subsystem three inputs are built. The first one is the sound environment classification. Sound classes are labeled on a numerical scale according to their alarm level. Four membership functions are set up in this numerical scale according to the following fuzzy levels: no signal, normal, possible alarm and alarm as it is shown in figure 4. The

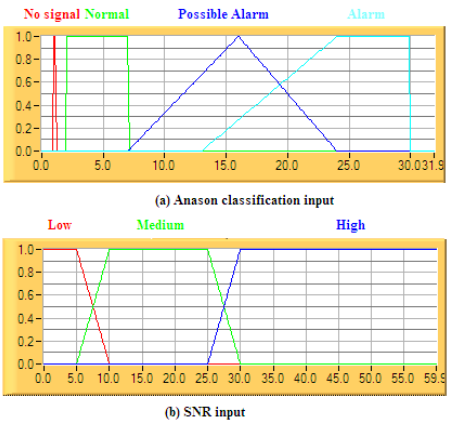


Fig. 4. Fuzzy sets defined for sound environment inputs.

second input is associated with the SNR (Signal-to-noise ratio) calculated on each microphone (two microphones are used in the current application), and this input is split into three fuzzy levels: low, medium and high.

RFpat provides physiological data to EMUTEM platform. RFpat produce five inputs:

- Heart rate is fuzzified with three fuzzy levels: normal, possible alarm and alarm.
- Activity has four fuzzy sets: immobile, rest, normal and agitation. The trapezoid function displays all these linguistic variables.
- Posture is represented by two membership functions standing up / seating down and lying.
- Fall and call have also two fuzzy levels: Fall/Call and No Fall/Call and a singleton function is associated to these linguistic variable.

Parameters of membership functions associated to the heart rate and the activity, are adjustable according to the monitored elderly person. An automatic procedure to adapt these intervals based on 30 minutes recording was proposed.

For each infrared sensor a counter of motion detection with

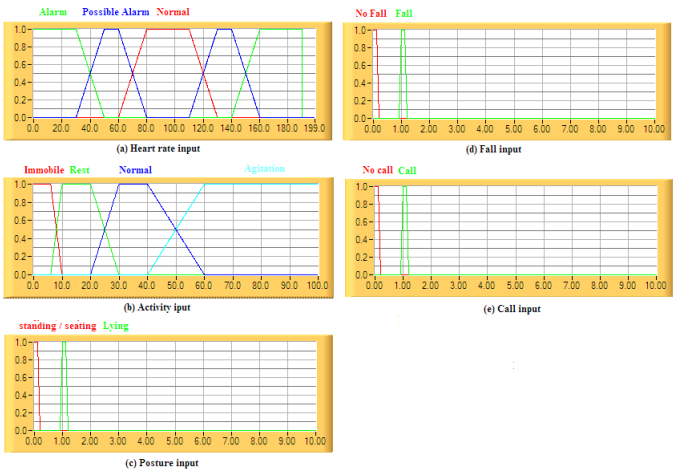


Fig. 5. Fuzzy sets defined for Vital data inputs.

three fuzzy levels (low, medium, high) is associated. It is reset every 5 seconds. A global counter for all infrared sensors with three fuzzy membership functions (low, medium, high) is also used and is reset every 60 seconds. A trapezoid membership function is chosen to characterize these fuzzy sets.

A singleton membership function is assigned to each change state sensor with two linguistic variables: On and Off.

The last input which is the time. It has two membership functions: day and night. These two membership functions are also adaptable to each patient habits. Trapezoid functions are used to divide the time input.

In order to reach the objective of the EMUTEM platform which is the identification of distress situations of an elderly person at home, two outputs are associated to the fuzzy inference engine of EMUTEM platform.

The first one is called Alarm with two linguistic variables normal and alarm.

To refine the decision of the EMUTEM platform a second output is added to its fuzzy inference system component. This second output is named Localization. It is a very important information for the diagnostic. The identification of the position of the monitored person during a distress situation is helpful knowledge for medical diagnosis.

Two membership function models are selected: Gaussian functions are chosen for the alarm outputs; Trapezoid functions for the localization output where the classical areas of a house are its fuzzy levels or linguistic variables.

C. Fuzzy Rules Aggregation and Defuzzification

The EMUTEM fuzzy inference engine is formulated by two groups of fuzzy IF-THEN rules. One group controls the output variable localization according to values of the input variables issued from infrared sensors and the SNR of each microphone. The other group controls the output linguistic variable alarm according to all inputs. These fuzzy rules are decided through experimentation and according to some expert knowledge.

A confidence factor is assigned for each rule. Outputs used in a rule are multiplied by the confidential factor issued from each subsystem. Thus the rule's output value depends on the reliability of each subsystem and the sureness of rules. To aggregate these rules we have chosen the Mamdani model [16] instead of the Takagi Sugeno. These two models are available under the EMUTEM fuzzy logic component. Mamdani model offers us a good way of modeling the normal and distress situations, because these two classes don't form a clean partition but a fuzzy one.

After rules aggregation the defuzzification is performed by the smallest value of maximum method for the alarm output in order to obtain also a confidence level of each alarm's decision, and the centroid of area for the localization output.

Each subsystem specifies the situation of the elderly person and the degree of anxiety. When a decision is very complicated, that is, there are many decisions; the fuzzy method is especially useful. It is also easy to check, modify, and add/delete every fuzzy variable for better automated analysis. To summarize the proposed data fusion method, the proposed fuzzy destress

situation classifier is comprised of two major function blocks, decisions of each subsystems and fuzzy classifier as shown in Figure 3. Decisions of subsystems are utilized as inputs for the classification using the fuzzy classifier. The derived parameters will be exported to the latter for the classification. The Mamdani fuzzy method is used. Fuzzy logic if-then rules are formed by applying fuzzy operations to these membership functions for given inputs and outputs. The resulting output membership functions are added together using desired weights yielding a sort of probability function. The resulting output areas can then be used to estimate the expected value of the output variable by using the specified defuzzification method. In general, any fuzzy classifier has to undergo iterative adjustment in terms of fuzzy variables, including the choice of membership functions, and the definition of rules in the knowledge base.

D. Software Architecture

Figure 6 provides a synoptic block-diagram scheme of the software architecture of the EMUTEM system; it is implemented under Labwindows CVI and C++ software. It is developed in a form of a design component.

We can distinguish three main components, the acquisition module, the synchronization module and the fuzzy inference component. It can run off-line by reading data from a data base or online by processing in real time data acquired via the acquisition module.

To avoid the loss of data, a real time module with two multithreading tasks is integrated in the synchronization component. The EMUTEM system is now synchronized on the smart home sensors (Gardien) subsystem because it has the lowest sampling rate (2 Hz) and periodicity. The data from others modalities are memorized and used several time in order to have the same sampling rate (RFPAT data is used 60 times).

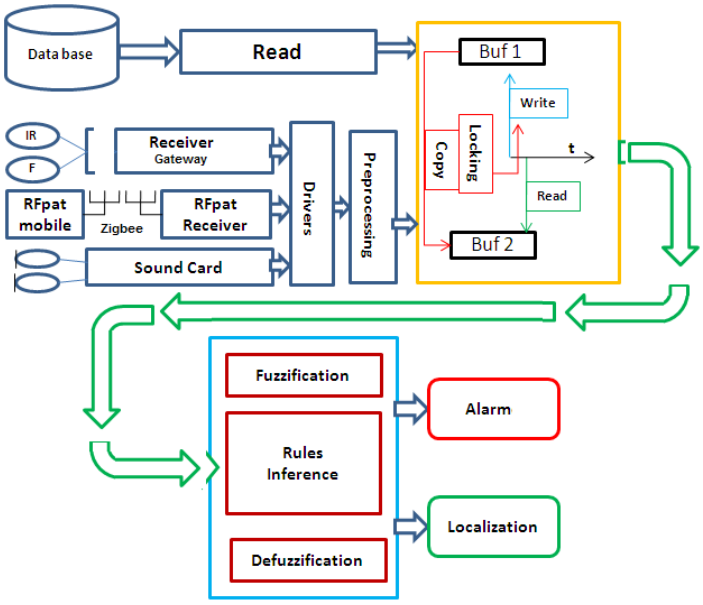


Fig. 6. EMUTEM software components design .

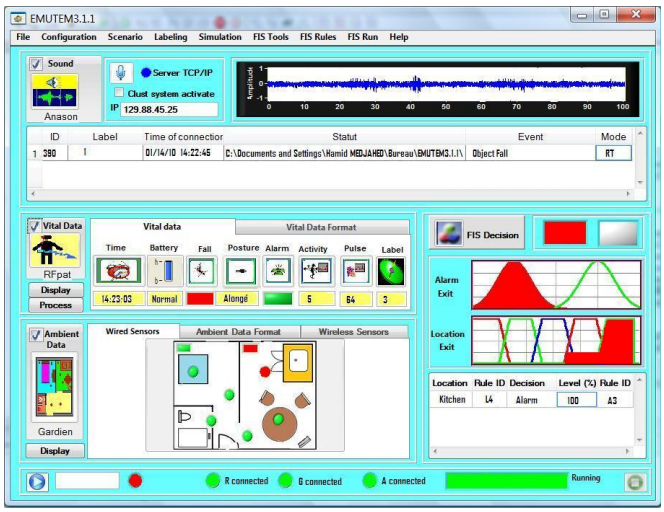


Fig. 7. Graphical Interfaces for Intelligibility.

This multithreading technique allows a maximum delay for an asynchronous data (Sound or alarm from RFPAT) of 0.5 s.

We have developed Fuzzy tools for the data fusion step. These tools allow the easy configuration of input/output intervals of Fuzzification, the writing of fuzzy rules and the configuration of the defuzzification method. It is also possible to add others modalities to this fuzzy inference system which make the EMUTEM platform flexible. Two outputs are associated to the fuzzy inference system, Alarm for distress situation detection and Localization for elderly person position detection.

E. Graphical Interfaces for Intelligibility

To offer enhanced intelligibility for the EMUTEM platform, a graphical user interface (GUI) was developed for this task. It facilitates the various configurations required for the functioning of the platform. It is very useful for users.

Figure 7 shows this general graphical interface. It is possible to build up membership functions of inputs and outputs and displaying them under this graphical interface.

We could also write rules via this graphical interface. It is also possible to write rules to text file by using a specific language, that we have developed, understandable by our system.

These Graphical interfaces provide EMUTEM with a useful simplicity for users and with a flexibility that allows adding other modalities. They allow experts to add their knowledge in a user friendly way.

V. The In Home Monitoring Database

In order to test and to validate the EMUTEM platform, an environment for acquiring and recording a multimodal database (HOMECAD [8]) is integrated under this platform. To record this multimodal database, users can interpret elderly activities by following a reference scenario which summarizes the everyday life of elderly persons. These scenarios are divided into two categories: either a critical scenario with one

or more distress events, or a normal one without any distress event. To define these scenarios a study is performed and they were instigated under CompanionAble European project. In this project, some elderly living alone were followed up by a co-worker team, in order to summarize and to describe their daily routine. The recorded multimodal database gathers physiological data, environment sounds and others different information gathered by ubiquitous sensors associated to Gardien subsystem.

First recordings were performed in our laboratory. Participants in this study were 10 adult volunteers who were recruited from the university and from the community. Participants ranged in age from 25 to 50 years. Participants were asked to perform several activities according to the used reference scenario. As each scenario lasts 10 minutes, this task corresponds to 200 minutes of recorded data.

The second stage of recording occurred in Broca Hospital in Paris under the framework of CompanionAble project. Ten adult volunteers from the community were enrolled to participate in this recording stage. The participants profile varied, ranged in age from 65 to 75 years, and the sample was 60% female and 40% male. The last stage of recording was also performed under the CompanionAble project and was performed in SmartHomes at Eindhoven. Fifteen adult volunteers were involved in this recording. For both stages of recording, a scenario lasts 15 minutes and each participant performs three different scenarios.

An additional process of simulation is also integrated in our platform as a way to overcome the lack of experimental data and the difficulty of recording some medical data such as the cardiac frequency during distress situations.

Taking into account the multimodality character of the data, a multidimensional indexing process is used in order to obtain a full description of data sets. In order to index our multimodal database, we have retained the SAM standard indexing file [17] generally used for speech database description. It indicates information about the file and describes it by delimiting the useful part for further analyzing and processing.

VI. Experimental Results and Validation

The realization of an experimental process requires the use of appropriate metrics for evaluating the performance of the platform by comparing the system's results to expected results. It is useful to describe some parameters or metrics that allows an objective evaluation of the results. To perform this experimental process we have selected the following metrics:

- *Sensitivity (Se)*: Identify patterns of real abnormal situations as distress ones.
- *Specificity (Sp)*: Don't identify normal situations as distress situations.
- *Error rate (Err)*: It is the ratio between the number of the misclassified samples and the total number of the samples.
- *Perfect classification(Pc)*: It is the ratio between the number of the correct classified samples and the total number of the samples.

		EMUTEM Classification	
		Distress sequence	Normal sequence
Simulated sequences	Distress sequence	68	2
	Normal sequence	1	29

TABLE I
Classification results for Alarm output with simulated data by using 10 rules

Indices of sensitivity (Se), specificity (Sp), error rate (Err) and perfect classification (Pc) are calculated from rates of true/false positive/negative, marked respectively with these symbols TP, FP, TN, and FN.

$$Se(\%) = \frac{TP}{TP+FN} \times 100 \quad (1)$$

$$Sp(\%) = \frac{TN}{TN+FP} \times 100 \quad (2)$$

$$Err(\%) = \frac{FN+FP}{TN+FN+FP+TP} \times 100 \quad (3)$$

$$Pc(\%) = \frac{TN+TP}{TN+FN+FP+TP} \times 100 \quad (4)$$

The just exposed metrics of the statistic data are very important to estimate the classification accuracy.

In order to demonstrate the effectiveness of this software, firstly we started by using simulated data in order to validate each rule. This first step of simulation gave very promising results for the alarm generation and localization without any false decision for each rule.

After that 100 sequences of simulation are used to test EMUTEM, where 70 sequences represent distress situation and 30 sequences represente normal situation, because we wanted especially to test the distress part that is more difficult to record in a real situation.

In order to evaluate the classification accuracy the confusion matrix has been calculated for this simulation. Table I displays the obtained results with 10 rules.

Sensitivity Se	97%
Specificity Sp	96%
Error rate Err	3%
Perfect classification Pc	97%

TABLE II
Performance indices for Alarm output obtained with simulated data by using 10 rules

From table I we can deduce some indices of performance which are displayed in table II. The obtained results of EMUTEM's performance are good and they demonstrate the reliability of the EMUTEM platform. Even if we have 3% of misclassified sequences, this error rate could be overcome by adding to the fuzzy inference system the right rules that take into account the misclassified situations, and also by associating to each rules the right weight.

For the localization output, also we have obtained about 98% of good localization.

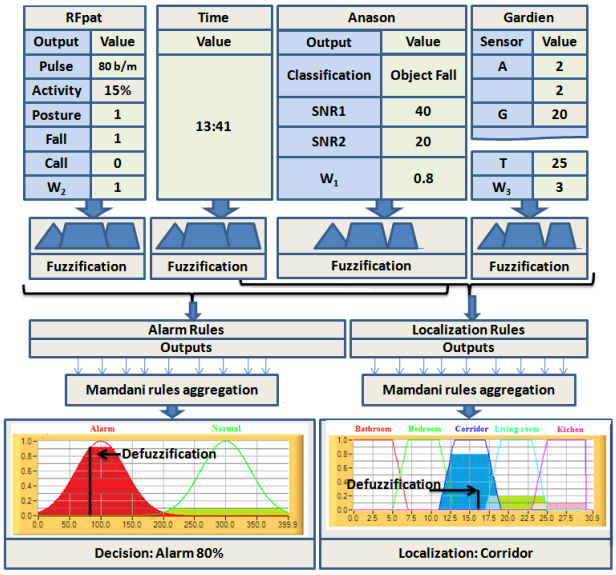


Fig. 8. Results for a stream of data

Figure 8 shows results for a steam of data and summarizes the process of the fuzzy inference system. The first step of this process is to determine the output of each membership function assigned to each input. Then the outputs of the rules are calculated by applying the fuzzy operators (AND and OR) in the antecedent part of all rules, using the T-norm and T-conform operators respectively and the implication from the antecedent to the consequent, using T-norm operator. After that the aggregation of the consequents across the rules of each output is performed by using the Mamdani model, it is the unification process of the all rule outputs. Finally the process of defuzzification is done by extracting out one crisp value as the output, the smallest value of maximum method is chosen for the alarm output and the centroid of area for the localization output.

The EMUTEM platform was tested with the HOMECAD database described in the previous section. This first study is devoted to the evaluation of the system by taking into account rules used in this fuzzy inference system. The strategy consisted in realizing several tests with different combination rules, and based on the obtained results, one rule is added to the selected set of rules, or removed from this selected set of rules in order to get the missed detection. Based on the obtained results some weights of rules are also changed. With this strategy good results are reached for the alarm output with 10 rules and 16 rules for the localization outputs. The data fusion classification using the fuzzy logic approach has given better performance results when compared to analytical method and using artificial neural networks (ANN) for classification (Table III). More over problems associated with conventional neural network architectures such as learning rate limitation and difficulty in selecting the optimal number of hidden units are eliminated.

With this strategy good results are reached for the alarm output with 10 rules and 16 rules for the localization outputs,

	Fuzzy classifier (10 rules)	ANN (3 hidden layers)
Se	95%	87%
Sp	93%	85%
Err	5%	12%
Pc	95%	88%

TABLE III
PERFORMANCE INDICES FOR ALARM OUTPUT OBTAINED WITH FUZZY CLASSIFIER AND ANN METHOD

about 95% of good Alarm detection and 97% for good Localization. The rate of misclassification for the alarm output corresponds to situation that are not detectible by the sensors used by EMUTEM and also the difficulty to find the right rule to overcome these situations. For the localization output the error rate could be justified by the fact that we use an area where an apartment is simulated thus the calibration of infrared sensors is very hard.

These first results encourage us to perform further tests in real time in order to have an effective evaluation of the EMUTEM platform.

VII. Conclusion

In this work, we focused on the area of automatic home healthcare telemonitoring, in which health information is automatically collected with the help of sensors and processed by special algorithms and fused in order to make good decisions about elderly persons living alone at home. The proposed multimodal data fusion method based on fuzzy logic represents a fast and easy tool for the interpretation of the fuzzy decision process even for very high dimensional input spaces and allows fast detection of errors. The impact of the input features plays an important role on the final decision process. The algorithm supports the implementation of the expert's knowledge and optimized the system easily. It is possible to increase the performance further by adding more related input variables and with more data to enrich the knowledge in the rules.

The EMUTEM platform which encloses this architecture is implemented and validated by simulation and by using real data. The experimental results were accurate and robust. The main advantage of the presented method consists of the low-computational expenses inherited from the characteristics of fuzzy systems. This approach allows the easiest combination between data and adding other sensors.

The fuzzy logic decision module reinforces the secure detection of older person distress events and his localization. This constitutes a great asset for the EMUTEM system, by offering the possibility in the near future, to implement a very intelligent remote monitoring system in care receiver houses, thus to build very reliable smart houses.

Acknowledgment

The authors would like to thank the contribution of European Community's Seventh Framework Programme (FP7/2007-2011), CompanionAble Project (grant agreement n.216487).

REFERENCES

[1] H. Bouma, *Gerontechnology: A framework on technology and aging, In H. Bouma and J. Graafmans, editors, Gerontechnology*, ISO Press. pages 1-6, Amsterdam, 1993.

[2] D. Doermann and D. Mihalcik, *A system approach to achieving carernet, an integrated and intelligent telecare system*, IEEE Trans Biomed Eng, 2:1-9, 1998.

[3] N. Noury, P. Barralon, G. Virone, P. Boissy, M. Hamel, and P. Rumeau, *A smart sensor based on rules and its evaluation in daily routines*, in Proc of the IEEE-EMBC, pages 3286-3289, Cancun, Mexico, september 2003.

[4] G. Elger and B. Furugren, *"smartbo",an ICT an computer-based demonstration home for disabled people*, in Proc. of the 3rd TIDE Congress : Technology for Inclusive Design and Equality Improving the Quality of Life for the European Citizen, Helsinki, Finland, 1998.

[5] G.A.W. West, S. Greenhill and S. Venkatesh, *A probabilistic approach to the anxious home for activity monitoring*, in Proc. 29th Annual International Computer Software and Applications Conference: COMPSAC, pages 335-340, Edinburgh, Scotland 2005.

[6] H. Medjahed, D. Istrate, J. Boudy and B. Dorizzi, *Human Activities of Daily Living Recognition Using Fuzzy Logic For Elderly Home Monitoring*, IEEE Fuzzy Systems (FUZ-IEEE) 2009, 20-24 Aout 2009, Jeju Island, Korea, pp.2001-2006.

[7] J.E. Rougui, D. Istrate, W. Souidene, *Audio Sound Event Identification for distress situations and context awareness*, EMBC2009, September 2-6, Minneapolis, USA, 2009, pp. 3501-3504.

[8] H. Medjahed, D. Istrate, J. Boudy, F. Steenkeste, J.L. Baldinger, I. Belfeki, V. Martins and B. Dorizzi, *A Multimodal Platform for Database Recording and Elderly People Monitoring*, BIOSIGNALS 2008, 28-31 janvier 2008, Funchal-Madeira, Portugal, pp.385-392.

[9] S. Stillman and I. Essa. *Toward reliable multimodal sensing in aware environments*, Perceptual User Interfaces (PUI 2001) Workshop (held in conjunction with ACM UIST 2001 Conference), Orlando, Florida, November 15-16, 2001.

[10] Lawrence A. Klein. *Fuzzy set theory in medical diagnosis*, IEEE Tr. On Syst.,Man, and Cybernetics, pages 260-265, March/April 1986.

[11] S.J Russell and P. Norvig, *Probabilistic reasoning in artificial intelligence: a modern approach*, Upper Saddle River 2nd, pages 492-494, N.J Prentice Hall, 2003.

[12] G. Shafer, *A mathematical theory of evidence*. Princeton University Press, Princeton, NJ, 1976.

[13] C. Bishop. *Neural networks for pattern recognition*, Oxford, 1995.

[14] C.J.C. Burges. *A tutorial on SVM for pattern recognition*, Data Mining and Knowledge Discovery, Volume 2, pages 121-167, 1998.

[15] L. Zadeh, *Fuzzy sets. in Fuzzy Models for Pattern Recognition: Methods That Search for Structures in Data*, edited by J. Bezdek and S. Pal, IEEE Press, pages 35-45, New York 1992.

[16] J.S.R. Jang, C.T Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing : A Computational Approach to Learning and Machine Intelligence*, Prentice Hall Upper Saddle River, NJ 1997.

[17] D. Well, J. Barry, W. Grice, M. Fourcin, and A. Gibbon, *Esprit project 2589-multilingual speech input/output assessmen, methodology and standardization, In final report*, Technical Report SAM-UCLG004, University College London, 2004.

32nd Annual International Conference of the IEEE EMBS
Buenos Aires, Argentina, August 31 - September 4, 2010

Sound Event Detection in Remote Health Care – Small Learning Datasets and Over Constrained Gaussian Mixture Models

Jugurta Montalvão, Dan Istrate, Jerôme Boudy and Joan Mouba

***Abstract*— The use of Gaussian Mixture Models (GMM), adapted through the Expectation Minimization (EM) algorithm, is not rare in Audio Analysis for Surveillance Applications and Environmental sound recognition. Their use is founded on the good qualities of GMM models when aimed at approximating Probability Density Functions (PDF) of random variables. But in some cases, where models are to be adapted from small sample sets instead of large but generic databases, a problem of balance between model complexity and sample size may play an important role. From this perspective, we show, through simple sound classification experiments, that constrained GMM, with fewer degrees of freedom, as compared to GMM with full covariance matrices, provide better classification performances. Moreover, pushing this argument even further, we also show that a Parzen model can do even better than usual GMM.**

I. INTRODUCTION

Acoustic Event Detection and Classification is a recent sub-area of computational auditory scene analysis [1] where particular attention has been paid to automatic surveillance systems [2], [3]. More specifically, the use of audio sensors in monitoring applications has proven to be particularly useful for the detection of distress situation events, mainly when the person suffers from cognitive illness. The efficient detection and recognition of the distress situation is one part of the socially assistive robotics technology [4] aimed at providing affordable personalized cognitive assistance.

In recent works, it has been shown that automatic detection of relevant events for remote healthcare can be done in a rather conventional way, through the analysis of short segments (50ms) of digitalized signals from microphones strategically placed into rooms (e.g. places in the house of a elderly person under medical care). These short segments of sounds are then processed and features are extracted, much like what is done in speech or speaker recognition. Indeed, features such as Mel Frequency Cepstral Coefficients (MFCC) [5] and Matching Pursuit (MP) [6], along with Gaussian Mixture Models (GMM), are not rarely deployed for this kind of task.

J. Montalvão is with Faculty of Electrical Engineering, University of Sergipe, São Cristóvão, CEP. 49100-000, Brasil, *jmontalvao@ufs.br*

D. Istrate and J. Mouba are with the ESIGETEL school, 1, Rue du Port de Valvins, 77210 Fontainebleau-Avon, France, *dan.istrate@esigetel.fr*

J. Boudy is with the Télécom SudParis, 9 Rue Charles Fourier, 91011, Evry, France, *Jerome.Boudy@it-sudparis.eu*

Signals to be detected in healthcare scenarios show high variability from one instance to another, even for supposedly equivalent acoustic sources (intra-class variability). For instance, one can easily notice, through simple experiments, that door clapping sounds strongly depend on the door size, on the material the door is made of, and even on the room acoustics. This high variability explains indeed why recognition rates rapidly fall with increasing number of classes, as discussed in [6], and it rises a relevant question concerning adaptation of general classifiers to specific scenarios. More precisely, once a classifier was trained to recognize some classes of relevant sounds, one straightforward approach to adapt this classifier to a specific environment (e.g. a given apartment or house) is the adjustment of the universal class models to the specificities of the new environment, through some few new sound recordings locally acquired. But for very irregular classes of sounds, where new instances (new recordings) may strongly deviate from previously learnt universal model, this adaptation may be equivalent to obtaining a new model, instead of an incremental adaptation. In such cases, usual probabilistic models based on GMM, whose mixture parameters are found through the well-known Expectation-Maximization (EM) algorithm [7], demand a certain amount of new training signals to properly work. The acquisition of new training samples *in loco*, for model adjustment, may become cumbersome.

By contrast, if a classifier is able enough to properly learn a model from a few samples (recordings) per class, then the need for a universal model may be dropped in favor of full model learning *in loco*, from few recordings made in each new environment.

Probabilistic models are typically based on Probability Density Function (PDF) estimation from limited data sets, which is a classical problem in pattern recognition [8]. From this perspective, in this work, we focus on the problem of how to obtain useful GMM based PDF approximations, even when datasets are too small.

Our approach is greatly simplified if we define model regularization in a wide point of view, from which Parzen models with Gaussian kernels are regarded as over-regularized GMM.

II. THEORETICAL BACKGROUND

a) PDF ESTIMATION AND MODEL REGULARIZATION: PDF estimation from limited data sets is a classical problem in pattern recognition for which many approximated solutions are presented in literature

978-1-4244-4124-2/10/$25.00 ©2010 IEEE

[8], [10]. Probably the most widely used PDF model is GMM, along with EM algorithm for parameter adaptation (learning). It is worth noting that, though the EM is not the fastest algorithm for GMM optimization [9], it is usually simpler to apply, which can partially explain its widespread popularity in many application fields. However, in addition to its possibly poor convergence rate (depending on the data distribution and the initial estimates of its parameters), it also presents the following drawbacks [10]:Its likelihood-based criterion presents a multitude of useless global maxima; Convergence to parameter values associated with singularities is more likely to occur with small data sets, and when centers are not well separated. Indeed, it is well-known that likelihood is often unrepresentative in high dimensional problems, which can be true in some low-dimensional problems as well [11].

In order to cope with these drawbacks, model regularization is the usual solution, through which, the searching throughout the parameter space is constrained. Therefore, as far as regularization approaches lead to parametric constraints, we propose a wide point of view from which any reduction imposed to the mixture freedom degree is regarded as a kind of model regularization. Accordingly, regularization strategies can be roughly split into four categories, namely: (I) The most usual approach to regularization is based on the addition of a term to the unconstrained criterion function, which expresses constraints or desirable properties of solutions. (II) For models obtained via clustering-like algorithms (including the EM, which can be loosely seen as a soft clustering algorithm [10]), a straightforward regularization approach is that of averaging estimates from many independent initializations. (III) For GMM, regularization can be easily obtained by imposing constraints on the mixture component parameters (e.g. by imposing constraints or lower bounds on the covariance matrix).(IV) Conexionist models (e.g. artificial neural networks) can also be regularized, or partially regularized by pruning , though it is not always explicitly referred to as a regularization procedure.

In [12], for instance, two approaches to GMM regularization are compared: one based on averaging (II), and the other based on an explicit regularization term (I). Both provided improved models (if compared to the unconstrained one), with similar performances.

Thanks to this wide regularization concept, the nonparametric Parzen method [8], [10] can loosely be regarded as a mixture model based method with strongly-constrained mixture components (III). Thanks to this strong constraint on the Gaussian placement, the Parzen approach gives an instantaneous PDF approximation (no iterations) but, in spite of its simplicity, it is known that, under some constraints on the Parzen window width parameter, the convergence of the estimated PDF with the actual one is guaranteed, when the number of samples tends to infinity [10]. In other words, many small isotropic (radial basis) Gaussian kernels, with identical dispersion, can virtually approximate any PDF shape.

Although EM and Parzen approaches come from different paradigms - namely, parametric and nonparametric PDF estimation, respectively - they share a striking structural similarity, whenever the Parzen method is based on Gaussian kernels. In both cases, the actual PDF is approximated by a GMM. However, GMM provided by the Parzen method are intrinsically regularized, for kernel centers cannot move (structural regularization - IV) and identical radial dispersions are imposed on all kernels (parametric regularization - III). In this paper, we propose a useful point of view from which both kinds of PDF estimates – i.e. GMM learnt via EM and Parzen – are seen as GMM, with different levels of regularization. More precisely, starting from GMM with unconstrained covariance matrices (full covariance matrices), we can obtain several levels of parametric regularization, through the replacement of full covariance (Level 0) matrices with: **Level 1**: one diagonal covariance matrix for each Gaussian in the Mixture; **Level 2**: one scalar covariance matrix for each Gaussian in the Mixture; **Level 3**: the same scalar covariance matrix for all Gaussian in the Mixture; On this third level of parametric model regularization (III), we impose identical and isotropic Gaussian kernels throughout the mixture. Structurally, we are very close to the Parzen model with Gaussian kernels. In fact, the only remaining difference is that Gaussian centers cannot move during adaptation/learning of the Parzen model. **Level 4**: similar to Level 3, but Gaussian centers are not allowed to move during model adaptation/learning (i.e. the Parzen model).

PDF estimation in all proposed levels are given by:

$$f_X(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i g(\mathbf{x}|\mathbf{c}_i, \mathbf{R}_i) \quad (1)$$

where X stands for the multivariate random source to be modeled, $\mathbf{\Theta} = [\alpha_1, \ldots, \alpha_M, \mathbf{c}_1, \ldots, \mathbf{c}_M, \mathbf{R}_1, \ldots, \mathbf{R}_M]$ stands for the mixture parameter vector, and $g(\mathbf{x}|\mathbf{c}_i, \mathbf{R}_i)$ corresponds to the i-th Gaussian kernel of the mixture, with mean vector and covariance matrix given by $\mathbf{c}_i$ and $\mathbf{R}_i$, respectively. We further impose that $0 \leq \alpha_i \leq 1$ and $\sum_{i=1}^{M} \alpha_i = 1$.

This parametric model includes the Parzen model with Gaussian kernels, whenever the following restrictions on the parameter vector are imposed:

$$\mathbf{\Theta} = [\alpha_i = 1/M, \mathbf{c}_i = \mathbf{x}_i, \mathbf{R}_i = \sigma^2 \mathbf{I}] \quad (2)$$

where $i = 1, \ldots, M$.

These restrictions lead to a GMM equivalent to that obtained by the nonparametric Parzen method, where each Gaussian kernel center, $\mathbf{c}_i$, is directly given by a sample vector. Applying these restrictions to Equation 1

yields:

$$f_X(\mathbf{x}) = (1/M)\sum_{i=1}^{M} g(\mathbf{x}|\mathbf{x}_i, \sigma^2\mathbf{I}) \quad (3)$$

Consequently, as we can observe in Equation 2, under such strong constraint, the only free parameter in the model is σ, the Gaussian radial dispersion.

This is a single scalar parameter, and optimizing $\mathbf{\Theta}$ through likelihood maximization, in this case, is equivalent to find the value of σ that maximizes likelihood, which can be easily done by simple exhaustive one-dimensional search, through cross-validation approach [10]. By contrast, free parameters in Equation 1, corresponding to conventional GMM, are adapted through EM, in this work.

b) SIGNAL SEGMENTATION AND SHORT-TIME ANALYSIS: Raw signals are represented by samples, $s(n) \in \Re$, where $n \in \aleph$. In this work, samples are regularly taken at 16KHz. We assume that each raw signal, corresponding to each recorded file in our database, contains at least one relevant event corresponding to one of those sound sources, arbitrarily limited here to 4 or 6 (see Section III).

Therefore, in this work, we use both the whole sound file (Scheme I) and a single segmented sound from each file (Scheme II) we first use an algorithm to detect a single event and crop the corresponding subset of samples, $s_s(k)$, where $k \in \{k_{begin}, ..., k_{end}\}$. This segmentation task is done here by a very simple algorithm, based on power measurement.

Afterwards, segmented intervals of sound, $s_s(k)$, for each sound file in the database, are short-time analyzed. That is to say that windows of 500 consecutive samples (approx. 31 ms at 16KHz) are taken as signal vectors to be projected in a new space of reduced dimension. In other words, the frame-by-frame analysis corresponds to an MFCC projection of short-time overlapping windows of 500 samples on 24D vectors of coefficients.

c) GMM with optimized number of Gaussians: The Scheme I uses as input the whole useful signal frames with smal features extraction (MFCC). The proposed technique, provide for each sound type a representative GMM model with a different order according to the spectral signature of the event corresponding and the training data duration. In this stage the classification module use a Cluster software package to automatically estimate the parameters of GMM from sample data [13]. The clustering procedure applies the EM algorithm together with an agglomerative clustering strategy to estimate the number of clusters which best explains the data. The estimation is based on the Rissenen order identification criteria known as minimum description length (MDL). This is equivalent to maximum-likelihood (ML) estimation when the number of clusters is used, but in addition it allows the number of clusters to be accurately estimated.

TABLE I
AVERAGE CLASSIFICATION ERROR RATIO, TRAINING WITH 7 RANDOMLY FILES, TESTED WITH OTHER FILES, SCHEME I AND II

Mixture Model	Av. error ratio (%) Scheme I	Av. error ratio (%) Scheme II
GMM, full	27.37%	48.76%
GMM, diag.	19.19%	42.00%
GMM, scalar	16.50%	21.60%
GMM, single scalar	16.50%	20.75%

III. DATASET, EXPERIMENTAL APPROACHES AND RESULTS

In this Section, we present experimental results obtained with a subsets of the sound database gathered in the framework of the (European) CompanionAble Project (http://www.companionable.net/).

d) Dataset: All files were recorded at a sampling rate of 16KHz, and only a single channel (monaural sound) of each recording is used in this work. The database subset used in this work contains only 6 classes, namely: door clapping (574 files), glass breaking (88 files), steps (22 files), screaming (73 files), cough (41 files), metal object falls (12 files).

e) Experimental Approach: Only 7 files, from each class, are arbitrarily chosen to model training. They are then processed (MFCC), producing 24D vectors of coefficients, $\mathbf{x}$, seen as instances of C multivariate random variables, $X_1, \ldots, X_C$, corresponding to C classes of sound. Model learning, in this work, corresponds to the use these instances to estimate the underlying PDF associated to each random variable, $f_{X_1}(\mathbf{x}), \ldots, f_{X_C}(\mathbf{x})$.

In both cases, with conventional GMM or Parzen models, any new sound is classified by comparing the averaged likelihood of each model for a given set of patterns (extracted from a recorded sound). More precisely, as far as we do not accept a no-classification result (reject class), we just take the class associated to the highest averaged likelihood as the recognized class.

f) Comparison results between Scheme I and II: We have evaluated the Scheme I and II on 4 sound classes using 7 files for training. In the Table I we can constate that the GMM algorithm using MDL in order to optimize the Gaussian number obtain some better results. Therefore for the two schems the level 2 and 3 constraints allow the best performances in the conditions of reduced number of training files.

g) More results with Scheme II: Five experiments were carried out in each GMM regularization level, from unconstrained GMM (Level 0) to over-constrained GMM (Level 4 - Parzen models). These experiments were designed to highlight the impact of constraint/regularization, in an increasing way, of GMM on performance assessment. Concerning GMM structure, the number of Gaussians is arbitrarily fixed to 8 (note that, unlike Scheme I, the number of Gaussian is the

TABLE II

AVERAGE CLASSIFICATION ERROR RATIO, TRAINING WITH 5 RANDOMLY CHOSEN FILES, AND TESTED WITH OTHER 7 FILES

Mixture Model	av. error ratio (%)	95% conf. interval
GMM, full	77.4%	±2.3%
GMM, diag.	66.7%	±1.0%
GMM, scalar	34.3%	±8.7%
GMM, single scalar	32.8%	±9.0%
Parzen models	16.7%	±6.1%

TABLE III

NUMBER OF FREE PARAMETERS PER MODEL, IN 24D SPACE

Mixture Model	parameters to be adapted
GMM, full cov. mat.	$625M$
GMM, diagonal cov. mat.	$49M$
GMM, scalar cov. mat.	$26M$
GMM, single scalar cov. mat.	$25M + 1$
Parzen models	1

same for all classes). Another important implementation aspect is that Gaussian centers, in EM algorithm, are initialized with points taken at random from the training set.

In each experiment, 7 files are randomly chosen, without reposition, in each class as a learning dataset, whereas the test dataset is formed by 7 files per class randomly chosen from the remaining files. Averaged error ratios are thus estimated 5 times per each class, and those independent error estimates are used to provide the confidence intervals presented in Table II.

It is clear that increased regularization improves classification performance, and we believe that the huge amount of free parameters in usual GMM (i.e. with full or diagonal covariance matrices), as compared to the limited amount of data for model training, mainly explains the performance gain of more constrained models. To further highlight the decreasing degree of freedom in each model, in Table III, we explicitly present their respective number of parameters to be adapted, per level of parametric and structural regularization.

IV. CONCLUSIONS

In this preliminary work, we present evidences that traditional GMM adapted with EM algorithms may not be a suitable PDF model to be rained with a small amount of training samples. Though it was presented through experiments with a reduced number of classes, we may easily recognize that it comes from a wider and quite older discussion concerning PDF estimation in pattern recognition domain, not always taken into account in practical applications. Here, we compared GMM with 4 levels of parametric and structural wide-sense regularization (as proposed in Section II-.0.a), from GMM with full covariance matrices to Parzen model with Gaussian kernels (seen here as an over-constrained GMM). By comparing performances with these models, we gave one illustration, through simple experiments, that even if both GMM and Parzen models are theoretically able to converge to the true PDF to be estimated, under training data "shortage", they provide remarkably different error ratios. What we observe through our experiments is that, with a reduced number of instances for the model training, the path taken by the Parzen model seems to be more performing, in terms of classification.

Thus, what we claim here is that it is a clear matter of model regularization: the more regularized, the better, if the number of training patterns are too limited, and we highlight that training data "shortage" is indeed a frequent condition met in healthcare applications, since one needs to train specific sound models for each new environment to be monitored (e.g. care receiver's house, flat). Moreover, combined to incremental training strategies, this approach can offer a good and fast existing sound models adaptation for a given environment presenting some time variabilities.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the Brazilian Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNP), as well as the contribution of the European Community's Seventh Framework Programme (FP7/2007-2011), Companion-Able Project (grant agreement n. 216487).

REFERENCES

[1] G. Valenzise, L. Gerosa, M. Tagliasacchi, F. Antonacci, A. Sarti, "Advanced Video and Signal Based Surveillance", *in AVSS 2007*, vol. 2, issue 5-7, 2007, pp 21-26.

[2] D. Wang, G. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms and Application*, Wiley-IEEE Press, 2006.

[3] C. Zieger, M. Omologo, "Acoustic event classification using a distributed microphone network with a GMM/SVM combined algorithm", *in Interspeech*, September 2008, pp 115-118.

[4] J.L. Rouas, J. Louradour, S. Ambellouis, "Audio Events Detection in Public Transport Vehicle", *in Proc. of the 9th International, IEEEConference on Intelligent Transportation System*, Sept. 2006, pp 733-738.

[5] D. Istrate, M. Binet, S. Cheng„ "Real time sound analysis for medical remote monitoring", *in Proc. IEEE EMBC*, Vancouver, Canada, Aug. 2008, pp 4640-4643.

[6] S. Chu, S. Narayanan, C.-C. J. Kuo, Environmental sound recognition with time-frequency audio features, *Trans. Audio, Speech and Lang. Proc.*, vol. 17, no 6, 2009, pp 1142-1158.

[7] A. Dempster, N. Laird, D. Rubin, Maximum likelihood estimation from incomplete data using the em algorithm, *J Royal Stat Soc.*, vol. 39, 1977, pp 1-38.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Inc., 2nd edition, 2001.

[9] D. M. Titterington, A. F. M. Smith, U. E. Makov, *Maximum likelihood estimation from incomplete data using the EM algorithm*, John Wiley & Sons, Inc., 1985.

[10] A. Webb, *Statistical Pattern Recognition*, Wiley, 2nd edition, 2002.

[11] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.

[12] D. Ormoneit and V. Tresp. Improved gaussian mixture density estimates using bayesian penalty terms and network averaging, *The MIT Press, Advances in Neural Information Processing Systems*, 1996, pp 542-548.

[13] Bouman, *Cluster: An unsupervised algorithm for modeling Gaussian mixtures*, available from http://www.ece.purdue.edu/ bouman, April, 1997.

31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Audio Based Surveillance for Cognitive Assistance Using a CMT Microphone within Socially Assistive Technology

J.E. Rougui, D. Istrate, W. Souidene, M. Opitz and M. Riemann

Abstract — **This work proposes a system for Acoustic Event Detection and Classification (*AEDC*) using enhanced audio signal provided by a CMT (Coincidence Microphone Technology) microphone. The CMT microphone through signal processing algorithm provides an enhanced signal in several azimuths with a step of 15°. The AEC module exploits this technology to increase classification performance. The automatic detection system based on DWT uses an adaptive threshold for a different energy level and sampling rate quality. The classification system is based on an unsupervised order estimation of Gaussian mixture model adapted to the variability of sound event acoustic information and the representation cost.**

I. Introduction

Audio based surveillance systems stem from the field of automatic audio classification and matching. Traditional tasks in this area are speech/music segmentation and classification or audio retrieval. More recently, specific algorithms covering the detection of particular classes of events for multimedia-based surveillance have been developed.

Acoustic Event Detection and Classification is a recent sub-area of computational auditory scene analysis [1] where particular attention has been paid to automatic surveillance systems [2], [3], [4]. In particular, the use of audio sensors in surveillance and monitoring applications has proven to be particularly useful for the detection of distress situation events, chiefly when the patients suffer from cognitive illness. The recent research work in medicine has concluded that some patients with mild cognitive impairment will develop Alzheimer in the future. The efficient detection and recognition of the distress situation is one part of the socially assistive robotics technology [5] aimed at providing affordable personalized cognitive assistance.

This work deals with the classification of speech and non-speech events, where the considered non-speech events are typical sounds that may occur in everyday life. In practice some of the sound events may be considered as a noise of everyday life which can perturb the recognition task.

The proposed implementation is based on a hierarchical approach that has also been employed in [6]. We propose a specific system able to detect a speech utterance used as input for distress expression recognition system or/and dialogue system. The use of an acoustic system for tracking and recognition remains most useful compared to video surveillance, especially in a home environment. Mainly we consider the human solo sounds as a vital signals like "Snore, Cough, Cry,…,etc.".

We extend the previous work from using an omnidirectional microphone-based, firstly, to exploit the acoustic diversity observed by a set of CMT microphones-based placed far from each other and, secondly, to decrease the mismatch that can be caused by several factors. The aim is to select a useful signal component out of several events occurring at the same time. The CMT microphone localizes the sound event and can provide an enhanced signal if two sound sources are presented at the same time. The main goal is to develop a system that is robust to the presence of noise that might be generated for example by the hairdryer, vacuum cleaner or water flushing.

This research is being conducted under the European Project CompanionAble[1] an internationally active group dedicated to carrying out leading-edge research in computer vision and signal processing for man-machine communication, including patient home-care, gesture-based interaction, biometry, video surveillance.

II. Context and Goals

The proposed audio based surveillance system is developed in the framework of CompanionAble project with the three goals: patient security, domotic application and context awareness.

In order to assure theses goals the global system is designed to use a multiple microphones in each area depending on the room dimensions and properties. The larger room will be equipped with one or two CMT microphones which allow sound localization, however the other rooms will contain omnidirectional microphones. Fig.1 presents the sound processing architecture.

The analysis system consists of the two modules that allow the localization of useful event audio segment. The identification of the event given by the audio segment is carried out on 24 channels generated by a process provided by the CMT microphone. However, the segmentation module is carried out only on the omnidirectional signal. In the case of simultaneous detections the low level data fusion chooses signals based on the signal-to-noise ratio (SNR). The detection module associated with the CMT microphone

J.E. Rougui, D. Istrate and W. Souidene, are with LRIT-ESIGETEL, 1, Rue du Port de Valvins, 77210 Fontainebleau-Avon Cedex, France, {jamal.rougui,dan.istrate,wided.souidene}@esigetel.fr.
M. Opitz and M. Riemann are with AKG Acoustics GmbH Lemböckgasse 21-25 A-1230 Vienna, Austria. martin.opitz@harman.com, marco.riemann@harman.com.

[1] www.companionable.net

communicates with localization algorithm in order to enhance the signal in the useful direction.

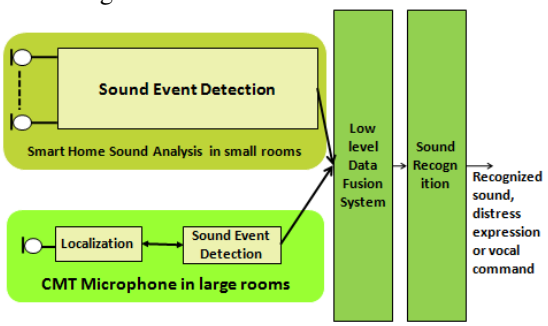


Fig.1 – Sound Processing Architecture

III. CMT AND LOCALIZATION

In order to increase system robustness for all possible locations of the acoustic source, a CMT microphone network is adopted here. At least one microphone per room is used in order to ensure a good spatial coverage

The CMT microphone consists of one pressure transducer and three first order pressure gradient electret transducers, each with a diaphragm, with each pressure gradient transducer having a first sound inlet opening, which leads to the front of the diaphragm, and a second sound inlet opening, which leads to the back of the diaphragm. Both sound inlets are on the same side of the disc shaped pressure gradient transducers.

The three pressure gradient transducers lie all in one plane. Their respective main directions – the directions of their maximum sensitivity – are lying in the same plane and are inclined relative to each other by 120 degrees. The acoustical centers of all 4 microphones are lying close together within a sphere with few millimeters radius.

In the further context we will refer to azimuthal detection of the direction of sound incidence only, as this is the most important localization information in the context of CompanionAble.

IV. SOUND DETECTION AND CLASSIFICATION

The sound flow provided by the CMT microphone is analyzed trough a hierarchical approach that involves firstly a useful signal detection followed by an event classification.

The first sound analysis module is the event detection module which is an important step before the event classification, especially when the events detection occurs in a variable noise of the home environment.

The signal classification starts with a sound/speech identification followed by a classification adapted to the identified signal. If the label was speech, a speech recognition engine is used and if a sound was identified a sound classes recognition system is launched. In this paper we are focusing on the sound identification.

A. *Sound event segmentation*

The audio segmentation must be able to detect a short event like an impulsive signal. Ideally, the segmentation module must be robust against a low signal energy due to a distant acquisition and different acquisition qualities. The classic techniques of event detection are based on the signal energy threshold or on other statistical features threshold [6],[7].

The wavelet based event detection algorithm proposed in [8] was adopted in this work. This algorithm is based on DWT (Discrete Wavelet Transform) using Daubechies wavelets with 6 vanishing moments. An adaptive threshold, depending on average and standard deviation of the energy is applied on the high frequency wavelet transform coefficients.

B. *Unsupervised Gaussian mixture modeling*

The extracted signal is analyzed by a hierarchical classification system. Firstly a classification between vocal and non vocal is carried out. In the case of non vocal signal a new classification between some everyday life sounds and noises is started. The sound classes were defined by CompanionAble consortium in order to allow the distress situation detection but also to help context awareness identification. Each classification module is based on GMM with an optimized number of Gaussian mixture [11]. Fig. 2 presents the hierarchical signal classification and the detected sound classes.

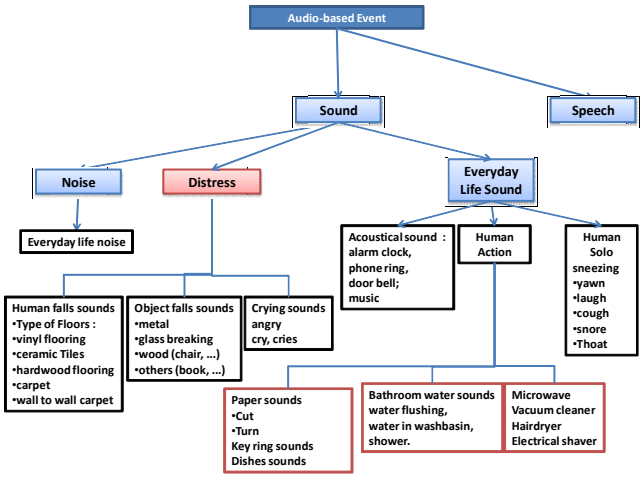


Fig. 2 - Hierarchical Sound event for smart home application

C. *Coupling CMT with sound analysis module*

The sound source localization algorithm of CMT allows listening in 24 directions (15° angular resolution in the horizontal plane). The signal coming from the omnidirectional microphone which contains all information is analyzed by the sound segmentation module. The start and stop information for each detected signal is used in order to segment the 24 azimuthal files. As shown in Fig.3 the processed files given to each azimuth have the same content with a different SNR. The low level data fusion (Fig. 1) is composed by a matching algorithm between all extracted signals in order to choose that one, which is best suited for the classification. In fact the classification is carried out on all segments and the output is a matrix composed by the most probable classification hypothesis for each segment on each azimuth coupled with its likelihood ($ClassHyph_{i,j}$ $ML_{i,j}$). For each detected segment the classification

hypothesis with the Maximum likelihood is considered like the identified signal.

In the next section we compare the results obtained with the omnidirectional microphone, being part of the CMT microphone, with those results obtained on enhanced localized signals.

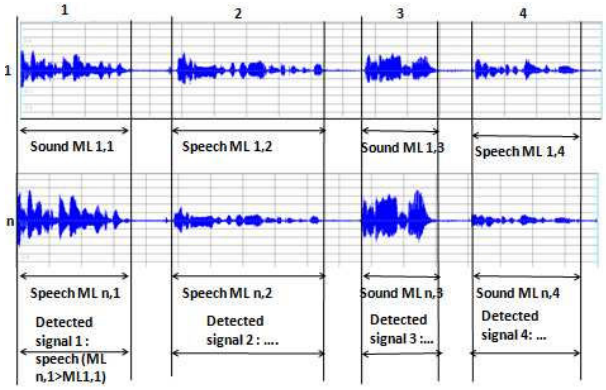


Fig.3 - Real time analysis of audio files processed-based CMT microphone by each azimuth in case of everyday life situations, discussion between 2 persons.

V. EVALUATION

A. Training Corpus for localization and classification

In order to use the localization ability of the CMT microphone, a database with training data has to be recorded in advance. For this purpose, the CMT microphone was placed on an acoustic boundary layer as is also foreseen in the actual application. The impulse responses of all 4 single transducers of the CMT microphone were measured for 24 directions in the horizontal plane corresponding to an azimuth distance of 15 degrees between the single measurements. For the measurements an AKG proprietary PC based measurement system was used. For the measurements, a Tannoy loudspeaker Sytem600 emitting a periodic noise signal with low crest factor was used. Applying the DFT (Discrete Fourier Transform), the corresponding transfer functions were determined and the results were stored in the database of the training corpus. The influences of the measurement loudspeaker, the amplifier and A/D-D/A-converters were determined by a reference measurement with a ½" calibrated measurement microphone and were removed from the CMT microphone data.

The sound classification module has currently 24 sound classes trained on 108' of signal and 5 noises of everyday life (Vacuum cleaner, Water flushing, Dishwasher, HairDrayer, RadioTv) trained on 18' of signal. The classes referring to sound and speech for the first classification level were trained on all existing sounds and on 38' of speech respectively.

B. Test Corpus

In order to evaluate our CMT based sound analysis approach we have recorded 20 scenarios in ESIGETEL laboratory using two CMT microphones. The sound signals were acquired with a RME DSP Multiface II card at 44.1 kHz sampling rate. Calibration of recording level was done using a Tannoy Precision 6D loudspeaker generating white noise with 70dB$_{SPL}$ linear weighting.

The recordings were made in two different rooms: one with a rectangular shape (Fig. 4) and another one with a triangular shape in order to evaluate also the influence of sound reflection on localization and classification.

The 20 scenarios were composed of 10 normal scenarios (everyday life situations, discussion between 2 persons…) and 10 distress scenarios containing the fall of a person (simulated by the actor), some distress expressions or distress sounds. Each scenario has been played in the two rooms and a video recording has been made for easy labeling. The data base has about 34 minutes of recordings.

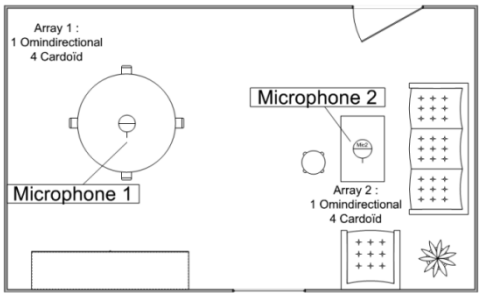


Fig. 4 - CMT microphone layout in rectangular room

C. Results

The proposed system was evaluated on the presented corpus in terms of localization, sound event detection and classification. We present here the example of a Normal Scenario were two persons have a discussion.

1) Localization

For the localization of a sound source with unknown position recorded by the CMT microphone the following strategy is used: the incoming signal is split into blocks of about 20ms length. For each block, a DFT is applied to all 4 signals. The amplitude spectra of the signals stemming from the three pressure gradient transducers are normalized by dividing them by the spectrum of the omnidirectional microphone. Comparing these normalized spectra with all spectra of the database, the direction for the most probable sound incidence is derived. The algorithm used is based on the method described in [10].

In Fig. 5 an example of the localization is shown for a dialogue of 2 speakers. The dialogue shown in Fig. 5 is part of the Normal Scenario 1 listed in Table I. The two speakers were localized such that their voices impinged on the microphone from 300 degrees and 50 degrees azimuth respectively. First a manual tagging of the respective speaker was done. In Fig. 5 the result of the manual tagging is shown with a dashed red line. After the end of each speech section the tagged angle was kept on the last detected azimuth. The automatic detection of the direction of speech sound incidence is shown by the red line in Fig. 5. Apart from minor delays in the attack phase at the beginning of the speech sections the congruence between manually and automatically detected direction of speech sound incidence is very good.

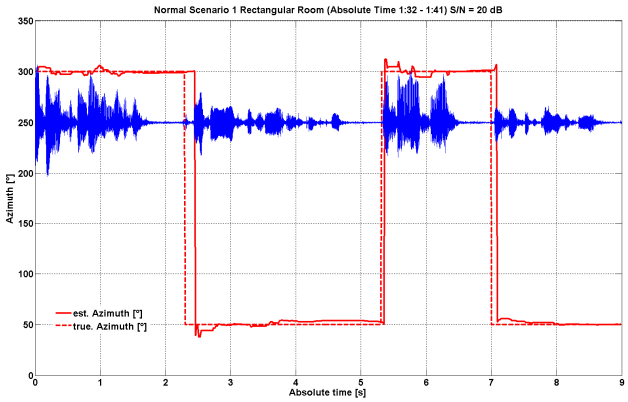


Fig.5 - Localization result for 2 persons speaking

2) Sound detection and classification

The sound event detection is evaluated in terms of number of correct detected events. The recorded signal was manually labeled in SAM format [9]. We consider a correctly detected event if the middle of segmented signal corresponds to a reference segment and if its dimension is at minimum about 50% of the reference one. The Acoustic Event Detection rate (AED) is computed:

$$AED = \frac{N°\ correct\ detected\ events}{N°\ detected\ events} * 100$$

The classification sound/speech and sound classification are evaluated in terms of correctly classified signals through Acoustic Event Classification (AEC):

$$AEC = \frac{N°\ correct\ classified\ detected\ events}{N°\ correct\ detected\ events} * 100$$

The global performances of AED system are evaluated trough Acoustic Events Detected and Classified Rate (AEDC):

$$AEDC = \frac{N°correct\ classified\ events}{N°\ detected\ events} * 100$$

Firstly the proposed sound analysis system is evaluated on the omnidirectional microphone signal, which acquires the signal from all directions. These results are compared with the results obtained from the 24 directions localization files. The analysis is performed on a normal scenario with duration of about 2 minutes (see Table I).

In the Table II we can observe that the classification error rate (1-AEC) decrease from 26.7% in the case of omnidirectional microphone to **11.8%** in the case of the data fusion between different azimuth localization. This can be explained by the fact that SNR is enhanced for some events in some directions (Fig.3).

VI. Conclusion

In this paper we have presented a first approach of an audio based surveillance system for distress situation identification, vocal commands and context awareness detection which was developed in the framework of CompanionAble project. The current proposition uses a CMT microphone which allows localizing the sound source and to enhance the signal. Our first proposition based on the data fusion between different classifications of the same sound event indicates good performances and encourages us to evaluate them on a larger data base.

Acknowledgment

The authors gratefully acknowledge the contribution of European Community's Seventh Framework Programme (FP7/2007-2011), CompanionAble Project (grant agreement n. 216487).

Table I
Normal Scenario 1

Time	Duration	Action
00:00	00:20	person is sitting and reading a book
00:20	00:03	person moves the chair & stands up
00:23	00:20	person walks around
00:43	00:03	person sits down again
00:46	00:15	person is reading a book
01:01	00:20	another person is entering the room and is walking around
01:21	00:15	the person is sitting down to the desk
01:36	01:00	the two persons are talking
02:36	00:15	the first person leaves the room

Table II
Detection and classification on Normal Scenario 1

Signal type	AED	AEC	AEDC
Omnidirectional	66.7 %	73.3 %	48.9 %
Fusion on 24 localization signals	66.7 %	88.2 %	60.2 %

References

[1] G. Valenzise, L. Gerosa, M. Tagliasacchi,. F. Antonacci, A. Sarti, Advanced Video and Signal Based Surveillance, 2007. *AVSS 2007.* Volume 2, Issue, 5-7 Sept. 2007 Page(s):21 - 26

[2] D. Wang, G. Brown, "Computational Auditory Scene Analysis: Principles, Algorithms and Application", *Wiley-IEEE* Press, 2006.

[3] C. Zieger and M. Omologo, "Acoustic event classification using a distributed microphone network with a GMM/SVM combined algorithm", *Interspeech*, Brisbane, September 2008, pp. 115-118.

[4] D. Feil-Seifer, and M.J. Matarić, "Defining socially assistive robotics," in Proc. *IEEE International Conference on Rehabilitation Robotics (ICORR'05),* Chicago, Il, USA, June 2005, pp. 465–468.

[5] J.L. Rouas, J. Louradour, and S. Ambellouis. "Audio Events Detection in Public Transport Vehicle". Proc. *of the 9th International, IEEEConference on Intelligent Transportation System 2006, 17-20 Sept. 2006,pp.733 - 738*

[6] T. Yamada, N. Watanabe, F. Asano, N. Kitawaki, "Voice activity detection using non-speech models and HMM composition," Proc. *Workshop on Hands-free Speech Communication*, Apr. 2001, pp. 131–134, Tokyo, Japan, April 2001.

[7] A. Dufaux, "Detection and recognition of Impulsive Sounds Signals," Ph.D. dissertation, Faculté des sciences de l'Université de Neuchâtel, Switzerland, 2001.

[8] D. Istrate, E. Castelli, M. Vacher, L. Besacier, J.-F. Serignat, " Information extraction from sound for medical telemonitoring" *IEEE Transactions on Information Technology in Biomedicine,* Volume 10, Issue 2, April 2006 Page(s):264 – 274.

[9] D. Well, J. Barry, W. Grice, M. Fourcin, and A. Gibbon, SAM ESPRIT PROJECT2589-multilingual speech input/output assessment, methodology and standardization. *University College London*: Final report. *Technical Report SAM-UCLG004*.

[10] K. Freiberger, A. Sontacchi, M. Opitz, *Acoustic source localization using coincident microphone arrays,* Applied for Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09), Como, Italy, September 11-4, 2009.

[11] Bouman97, "Cluster: An unsupervised algorithm for modeling Gaussian mixtures", available from *http://www.ece.purdue.edu/~bouman* , April, 1997.

© *The Journal on Information Technology in Healthcare* 2008; **6**(3): 204–211

Embedded Implementation of Distress Situation Identification Through Sound Analysis

Dan Istrate, Michel Vacher, Jean-François Serignat**

ESIGETEL, Avon, and *LIG, Grenoble, France.

ABSTRACT

Objective: The development of an embedded system capable of detecting distress sounds, e.g. breaking glass or a cry for help, in a person's home and notifying relevant personnel in the case of a distress situation.

Methods: The system is based on a personal computer (PC) equipped with a sound card and microphone that is capable of performing real time analysis of sound signals. Sounds are processed through 4 modules: Sound Event Detection and Extraction, Sound/Speech Classification, Sound Recognition and Speech Recognition. Training, testing and validation of the model was performed using 2 databases – a life sound database which we created and a French adapted speech corpus (a large and structured set of texts recorded by hundreds of different French speakers).

Results: The system was found to be reliable for detecting and classifying sounds at signal to noise ratios of 10 decibels (dB) or more, with an error rate of 5% or less. However, it was less efficient at sound and speech recognition. The error rate for sound recognition ranged from 9% to 37% at different sound levels. For speech recognition the error rate was 22%. This comprised 6% due to distress words being picked up in a normal sentence (leading to potentially false distress alerts) and 16% due to a distress word not being recognised (resulting in potentially missed distress alerts).

Conclusion: An embedded PC, equipped with a classical sound card and a microphone, is capable of real-time detection and analysis of sounds to detect distress situations. The system requires further refinement to improve its accuracy before it can be evaluated in real-life.

INTRODUCTION

The number of elderly people living alone in their own homes is increasing as a result of the aging population. It is estimated that in 2030, 37% of the European population will be over 60 years and in 2050 the number of persons aged 80 years or more will have increased to 10% from a current level of 3%[1]. In France persons older than 60 make up approximately 20% of the population today, and this is projected to increase to 33% in 2050.

Correspondence and reprint requests: Dan Istrate, Assistant Professor, ESIGETEL 1 rue du Port de Valvins, 77210 Avon, France. E-mail: dan.istrate@esigetel.fr.

Elderly people living alone at home have an increased risk of home accidents such as falls due to cognitive or physical illnesses. One study found that 7% of elderly people have a home accident as a result of everyday activities and in 84% of these accidents a fall occurs[2].

E-Health systems, such as medical remote monitoring, can reduce the consequences of home accidents through the detection of a distress situation and quick transmission of an alarm signal to the emergency services or a nearby relative or friend. Current remote monitoring systems use several fixed sensors (infrared) and mobile sensors (fall detector, movement and pulse) to detect a distress situation[3,4]. We have previously reported a system which extracts information regarding the status of a patient, through sound environment monitoring[5]. The system acquires and analyses data from 5 microphones detecting everyday life sounds and sounds associated with alarm situations such as glass breaking, screams, falls and distress expressions such as "Help", "A doctor quickly!". To preserve patient privacy, the extracted sound or sentences are not recorded, except in the case of an alarm situation.

In this paper we build on our previous model with a new real time implementation of sound monitoring algorithms on an embedded PC using the standard PC sound card and a microphone.

System Description

The proposed remote sound monitoring system consists of sound monitoring algorithms running on an embedded PC, using the PC's standard sound card and a microphone. The acoustical environment is analysed in real time and is made up of four main modules (Figure 1):

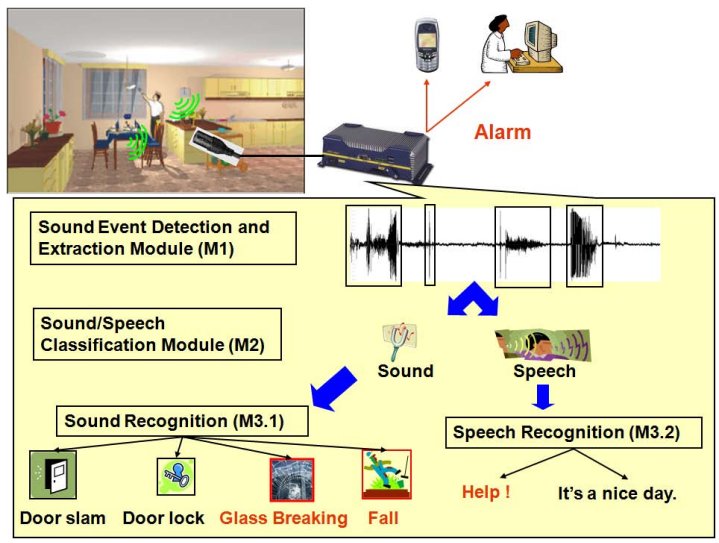


Figure 1. *Sound monitoring architecture*

- Sound Event Detection and Extraction (M1)
- Sound/Speech Classification (M2)
- Sound Recognition (M3.1)
- Speech Recognition (M3.2)

The signal extracted by the M1 module, is run in real time and classified as sound or speech by the M2 module. In the event of the signal being identified as a sound, the sound recognition module M3.1 classifies the signal as one of eight predefined sound classes. In the event that the signal is labelled as speech, the extracted signal is analysed by a speech recognition engine in order to detect distress sentences. In both cases, if an alarm situation has been identified, i.e. the sound or the sentence belongs to an alarm class such as glass breaking or the word Help (Figure 1), a local auditory and/or visual alarm is generated. If the patient does not respond by cancelling the alarm, an alarm message is sent as an e-mail or SMS (short message service) text to a nominated friend or relative and/or a message sent to the medical telemonitoring centre.

Sound Event Detection Module (M1)

The sound flow is analysed through a wavelet based algorithm which is aimed at sound event detection. This algorithm must be robust to neighbourhood environmental noises such as running water, rain, electric shaver, vacuum cleaner, etc. To address this problem, we previously proposed and evaluated an algorithm based on energy of wavelet coefficients[6]. This was shown to be able to precisely detect the beginning and end of a signal using properties of wavelet transformation.

Sound/Speech Classification Module (M2)

The method used by this module is based on Gaussian Mixture Model (GMM)[7]. There are other possibilities for signal classification such as Hidden Markov Model (HMM), Bayesian method, etc. However a major drawback of these other methods is that they are highly complex and take a long time to perform which effectively prevents them from being used in real-time. GMM is able to obtain an estimation in usually less than 20 steps. A preliminary step before signal classification is the extraction of acoustic parameters: LFCC (Linear Frequency Cepstral Coefficients) using 24 filters. The choice of the type of parameters depends on their properties; a bank of filters with constant bandwidth leads to equal resolution of the high frequencies often encountered in life sounds. The BIC (Bayesian Information Criterion) is used in order to find the optimal number of Gaussians[8]. The best performances have been obtained with 24 Gaussians.

Sound Recognition Module (M3.1)

This module is also based on a GMM algorithm. The LFCC acoustical parameters have been used for the same reasons than for sound/speech module and with the same composition, i.e. with 24 filters. The BIC method has been used in order

to determine the optimum number of Gaussians: 12 in the case of sounds. A log-likelihood is computed for the unknown signal according to each predefined sound classes. The sound class with the biggest log likelihood is the output of this module.

Speech Recognition Module (M3.2)

For Speech Recognition, the autonomous system RAPHAEL is used[9]. The language model of this system is a medium vocabulary statistical model (around 11,000 words). The model was created by using textual information extracted from the Internet[10] and from "Le Monde" corpora (a large and structured set of texts taken from the French newspaper *Le Monde*) and optimised for the distress sentences of our corpus (body of text). In order to ensure a good speaker independence, the training of the acoustic models of RAPHAEL have been made with a large structured sets of texts (BREF 80, BREF 120 and BFAF 100 corpora) and recorded with almost 300 French speakers[11].

Sound Database

In order to train, test and validate individual system modules and the complete system, we composed a life sound database and recorded a French adapted speech corpus.

The life sound data was divided into 8 classes corresponding to 2 categories: normal sounds related to usual activities of the person (e.g. door shutting/slamming, phone ringing, sound of steps, washing up sounds, etc.) and abnormal sounds related to distress situations (breaking glass, screams, objects falling, etc). This database contains recordings made at LIG (Laboratoire d'Informatique de Grenoble) (66%), files of "Sound Scene Database in Real Acoustical Environment"[12] (13%), files from the Internet[13] (10%) and files from a commercial CD (11%). An omnidirectional wireless microphone (Sennheiser eW500) was used for the recordings made at LIG. The life sound database has a total duration of 35 minutes and contains 1,985 audio sounds.

The speech database was recorded at LIG by 21 speakers (11 men and 10 women) aged between 20 and 65 years old. It is composed of 126 sentences in French: 66 are characteristic of normal everyday situations, e.g. "Bonjour" (Hello), "Où est le sel" (Where is the salt) and 60 are distress sentence, e.g. "Au secours" (Help), "Un médecin vite" (A doctor quickly). The speech database has a total duration of 38 minutes and contains 2,646 audio files.

With these two databases we generated a noise corpus with 4 levels of signal to noise ratio (SNR) (0 dB, +10 dB, +20 dB, and +40 dB) and evaluated the detection and classification modules. The HIS ("Habitat Intelligent pour la Santé") noise was recorded in an experimental test apartment.

REAL TIME IMPLEMENTATION

The sound telemonitoring system was implemented on an embedded PC using the integrated sound card.

The system is divided in four parallel threads and implemented under LabWindows/CVI (an integrated development environment providing a comprehensive set of programming tools for creating test and control applications). The sound signal acquisition is made through the sound card using the low Win32 functions which allows the use of a double buffer processed via software interruptions. The sample frequency is fixed to 16 KHz and the buffer dimension to 2×2048 samples corresponding to algorithm constraints.

Each time that the sound buffer is full, an interruption calls the detection algorithm. In the case of sound event detection the signal is recorded temporarily on the hard disk as a WAV (waveform audio format) file.

As the file is recorded, the detection thread also sends a message (the file name) through a safe communication queue to the recognition thread. The recognition thread is started in parallel with the detection thread and waits for a message from the detection task. As soon as the message is received, the Sound/Speech Classification algorithm is executed. If the signal is subsequently classified as an everyday sound, the Sound Recognition algorithm is started; alternatively if the signal is classified as speech, the corresponding WAV file is send to the speech recognition engine. In both cases, the Event Analysis sub-module decides on the action to be taken according to the recognised event. If an alarm sound or a distress sentence has been detected, an alarm with the recorded sound is sent using the activated modality (e-mail, SMS or Internet protocol) to the remote monitoring centre. If the processed event does not indicate an alarm situation the recorded file is deleted but the type of event and the corresponding time are written in the history file. The possible choices

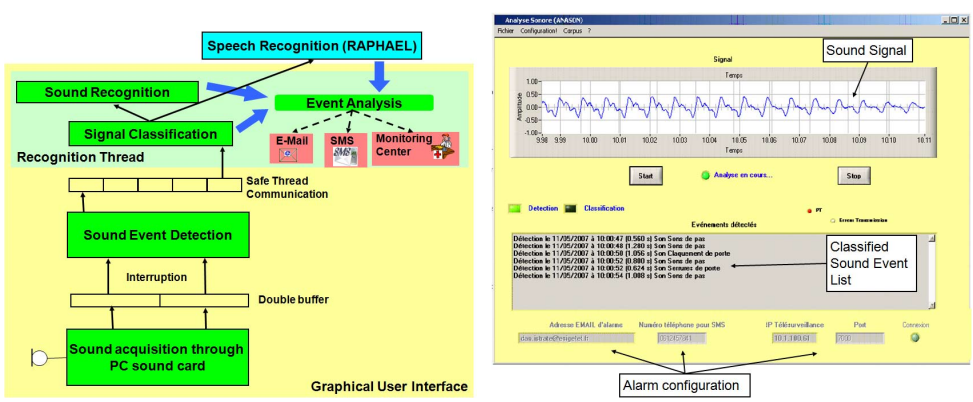


Figure 2. *Real-time architecture for sound monitoring and a view of the front panel of the application*

of the action to carry out in the case of distress event detection allow autonomous utilisation of the remote monitoring system.

The application front-panel, shown in Figure 2, displays in real-time the sound signal, a list of previously detected events and a summary of main alarm action parameters. A special menu allows the user to specify the sound card to use (if more than one), to activate the action(s) to carry out in the case of an alarm and to configure the parameters of these actions (e-mail of a close friend/relative, SMTP (Simple Mail Transfer Protocol) e-mail server, IP (Internet Protocol) address of the remote monitoring centre).

RESULTS

Each module of the proposed sound telemonitoring system was validated separately followed by validation of the complete system. The results of each module, except for the speech recognition module, are shown in Table 1.

Detection

The detection module was evaluated via Receiver Operating Curves (ROC) giving *missed detection rate* as a function of *false detection rate*. The *Equal Error Rate* (EER) was 0% above +10 dB of Signal Noise Ratio (SNR) and 3.7% at 0 dB (Table 1). The timing precision for the beginning of signal detection is less than 30 ms and for the signal end is below 100 ms. The signal sample rate was 16 kHz and the analysis window 2048 samples (128 ms).

Sound/Speech Classification

The analysis window was set to 16 ms (256 samples) with an overlap of 8 ms. The sound/speech classification was evaluated using a cross-validation protocol. Training was accomplished with 80% of the database, and the remaining 20% used in the test stage (no tests were performed using the same speakers or sentences used in training). Training was performed with pure sounds and testing with sounds mixed with HIS noise at 0, +10, +20 and +40 dB levels. Speech/sound discrimination performances were evaluated using the Classification Error Rate (CER). Table 1 shows

Table 1. *Evaluation of the sound telemonitoring modules*

Module	SNR			
	0 dB	10 dB	20 dB	40 dB
Detection	3.7%	0%	0%	0%
Sound/Speech Classification	17.3%	5.1%	3.8%	3.6%
Sound Recognition	36.6%	21.3%	13%	9.3%

dB = Decibels, SNR = Signal to Noise Ratio.

the classification results for 24 LFCC parameters; the CER is 4% above +10 dB and 17.3% at 0 dB.

Sound Recognition
The analysis window was set to 16 ms with an overlap of 8 ms and 24 LFCC parameters ere used. The classification was achieved with a cross-validation protocol – 90% of the database for training and 10% for testing. The module was evaluated using CER and achieved values of 13% at +20 dB and 21.3% at +10 dB (Table 1).

Speech Recognition
It is vital to the functioning of the system that key words related to a distress situation are well recognised. The speech recognition system has been evaluated on sentences pronounced by 5 speakers of our corpus (630 tests). For normal sentences, an unexpected distress key word is introduced by the system in 6% of the cases and leads to a *False Alarm Sentence*. For distress sentences, the distress key word is not recognised but missed in 16% of the cases; this leads to a *Missed Alarm Sentence*. We found that this often occurs in isolated words like "Aïe" (Ouch) or "SOS" or in French syntactically incorrect expressions like "Ça va pas bien" (I am not feeling very well). The Speech Recognition error rate is consequently 22%.

Real Time Evaluation
An initial evaluation of the complete sound remote monitoring system implemented in real time was also performed. The implementation was tested on an Embedded PC (AEON-6810) running Windows XP equipped with a USB (Universal Serial Bus) sound card (Creative 24 bits) and a Sennheiser microphone (ME 104 ANT). The results from this first evaluation were encouraging and will be followed by further systematic testing.

DISCUSSION

In this study we have built on our previous work in developing a system to extract information on the status of a patient through monitoring of environmental sound[5]. The system we have developed consists of an embedded PC enabling real-time implementation of remote sound monitoring. This is performed through continuous analysis of the sound environment permitting the recognition of everyday living sounds and differentiation of normal sound and speech from those associated with distress situations. The system has a number of advantages over our previously described system. The use of an embedded PC provides a compact, silent system which is relatively cheap to implement. Implementation is flexible as it can be installed on any desktop or laptop PC equipped with either an internal or external sound card. In addition the system offers flexibility with respect to the alarm generation due to the software.

Testing of the system demonstrated its ability to reliably detect and classify sounds at signal to noise ratios of 10 decibels or more. The error rate under these conditions was 0% for detection and 5% or less for classification. However, it was less accurate at sound and speech recognition. The error rate for sound recognition ranged from 9% to 37, whereas for speech recognition it was 22%. For speech recognition this was broken down into two components – errors due to distress words being picked up in a normal sentence leading to potentially false distress alerts (6%) and errors due to a distress word not being recognised resulting in potentially missed distress alerts (16%).

The system requires further refinement before it can be implemented into practice. The system may be improved by adding a real time SNR estimator that will allow the adaptation of the GMM models. Future developments also aim at combining this modality with the output of other medical sensors in order to increase the system's reliability.

REFERENCES

1 European Commission. Europe's response to world ageing: promoting economic and social progress in an ageing world. Second World Assembly on Ageing, 2002.

2 Thélot B. Résultats de l'enquête permanente sur les accidents de la vie courante, Réseau EPAC, Institut de Veille Sanitaire, Département Maladies Chroniques et Traumatismes, 2003.

3 Bellego GL, Noury N, Virone G, Mousseau M, Demongeot J. Measurement and model of the activity of a patient in his hospital suite. *IEEE TITB* 2006; **10**: 92–99.

4 Baldinger JL, Boudy J, Dorizzi B, *et al*. Tele-surveillance system for patient at home: the MEDIVILLE system. ICCHP 2004. http://www.esiee-management.fr/recherche/publis_pdf/lacombea-icchp2004.pdf.

5 Vacher M, Serignat JF, Chaillol S, Istrate D, Popescu V. Speech and sound use in a remote monitoring system for health care. *Lecture Notes in Computer Science, Artificial Intelligence, Text Speech and Dialogue*, 2006; **4188**: 711–18.

6 Istrate D, Castelli E, Vacher M, Besacier L, Serignat J. Information extraction from sound for medical telemonitoring. *IEEE TITB* 2006; **10**: 264–74.

7 Reynolds DA. Speaker identification and verification using Gaussian mixture speaker models. *Speech Comm*. 1995; **17**: 91–108..

8 Schwarz G. Estimating the dimension of a model. *Annals of Statistics*, 1978; **6**: 461–64.

9 Akbar M, Caelen J. Parole et traduction automatique: le module de reconnaissance RAPHAEL.COLING-ACL, Montréal, Quebec 1998; **2**: 36–40.

10 Vaufreydaz D, Rouillard J, Akbar M. Internet Documents: a Rich Source for Spoken Language Modelling. IEEE Workshop, Colorado, USA, 1999, pp. 277–81.

11 Gauvain JL, Lamel LF, Eskenazi M. Design considerations and text selection for BREF, a large French read-speech corpus", ICSLP Kobe, Japan, 1990, pp. 1097–100.

12 RealWorld Computing Partnership. CD – Sound Scene Database in Real Acoustical Environments (1998–2001).

13 "Bruitage", Bruitage Gratuits. http://www.sound-fishing.net/bruitages.htm.

Université d'Evry Val d'Essonne

Auteur : Dan ISTRATE
Titre HDR : Contribution à l'analyse de l'environnement sonore et à la fusion multimodale pour l'identification d'activités dans le cadre de la télévigilance médicale
Date de soutenance : le 6 décembre 2011

Résumé : La télévigilance médicale représente un enjeu de la société d'aujourd'hui. En effet l'espérance de vie augmente dans tous les pays industrialisés et les prévisions statistiques annoncent un nombre important de personnes âgées (17% de 60-74 ans en 2030) ou très âgées (12% de plus 75 ans en 2030). Grâce à la progression de la médecine ces personnes peuvent être maintenues plus longtemps à leur domicile mais demeurent plus fragiles et nécessitent donc des solutions techniques permettant d'améliorer leur confort et de faciliter la tâche des aidants.

Ce mémoire donne une synthèse des activités de recherche menées par l'auteur dans le domaine de la télévigilance médicale. Cette recherche est structurée en deux axes : l'analyse de l'environnement sonore et la fusion de données multimodales.

L'environnement sonore est très riche en informations utilisables, directement ou à travers l'analyse des activités de la personne pour détecter ou prévoir une situation de détresse. L'analyse sonore est soumise aux contraintes de l'acquisition sonore distante, à la présence des bruits provenant de l'extérieur et à la grande variabilité des sons à reconnaître. Le manuscrit décrit différentes solutions adoptées, leur mise en œuvre et leur évaluation dans le cadre de plusieurs projets de recherche nationaux et européens.

Le deuxième axe porte sur la fusion de la sortie de l'analyse sonore avec d'autres capteurs en vue d'améliorer la robustesse du système. La fusion de données doit traiter des signaux de natures différentes (signaux binaires ou continus), avec des périodicités différentes et de différentes temporalités (périodiques ou asynchrones). Deux techniques (logique floue et réseaux d'évidence) sont étudiées, adaptées et évaluées dans plusieurs projets de recherche.

Le mémoire se termine avec les perspectives de recherche de l'auteur. Six publications scientifiques sont finalement annexées.

Abstract : The medical remote monitoring is a today's society challenge because life expectancy is increasing in all countries and statistical forecasts announce a significant number of elderly (17% of 60-74 years in 2030) or very elderly (12% from 75 in 2030). With the progress of medicine they may be kept longer in their homes but are more fragile and therefore require technical solutions to make it easier for caregivers and increase the comfort of these people.

This manuscript provides a summary of research activities conducted by the author in the field of medical remote monitoring. These research activities are structured in two themes: sound environment analysis and multimodal data fusion.

The sound environment is very rich in information that can be used to detect or to predict distress, either directly or through the analysis of the activities of the person. The sound analysis is subject to the constraints of the remote audio acquisition, the presence of noise from outside and the large variability in recognizing sounds. The manuscript describes different solutions evaluated and their practical implementation in the framework of several European and national research projects.

The second theme is represented by merging the output of the noise analysis with other sensors to improve the robustness of the system. Data fusion must process signals of different nature (binary or continuous), with different sample rates and different types (periodic or asynchronous). Two techniques (fuzzy logic and evidence networks) are studied, adapted and evaluated in the same research projects.

This manuscript concludes with the research perspectives of the author. Six scientific papers are added in the appendix.