



**HAL**  
open science

## Traitement acoustique pour la télévigilance médicale à domicile

Toufik Guettari, Hamid Medjahed, Jérôme Boudy, Dan Istrate, Jean-Louis Baldinger, Imad Belfeki

► **To cite this version:**

Toufik Guettari, Hamid Medjahed, Jérôme Boudy, Dan Istrate, Jean-Louis Baldinger, et al.. Traitement acoustique pour la télévigilance médicale à domicile. SFTAG 2009: 1er Congrès de la Société Française des Technologies pour l'Autonomie et de Gérontechnologie, Nov 2009, Troyes, France. hal-01366482

**HAL Id: hal-01366482**

**<https://hal.science/hal-01366482>**

Submitted on 14 Sep 2016

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

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

# Traitement Acoustique pour la Télévigilance Médicale à Domicile

T.GUETTARI<sup>1</sup>, H.MEDJAHED<sup>1,2</sup>, J.BOUDY<sup>1</sup>, D.ISTRATE<sup>2</sup>, J.L.BALDINGER<sup>1</sup>, I.BELFEKI<sup>1</sup>.  
(1)EPH-TMSP (TELECOM & Management SudParis) 9, Rue Charles Fourier, 91011 Evry, France  
(2)LRIT/ESIGETEL 1, Rue du Port de Valvins, 77210 Avon, France

{toufik.guettari, hamid.medjahed, jerome.boudy, jean-louis.baldinger, imad.belfeki}@it-sudparis.eu  
{hamid.medjahed, dan.istrate}@esigetel.fr

*Mots clés:* Télévigilance médicale, Fusion multimodale, traitement acoustique.

## 1. Introduction

Le traitement acoustique est de plus en plus utilisé dans plusieurs domaines de recherche. Ces dernières années l'utilisation de traitement de son dans les systèmes de télévigilance médicale attire de plus en plus les équipes de recherche spécialisées dans ce domaine. Anason [1] utilise un ensemble de microphones pour l'analyse de l'environnement acoustique de la personne, afin de détecter des situations de détresse et/ou des appels à l'aide. Pour améliorer la robustesse de ce système nous proposons un ensemble d'approches de fusion basés sur la technique de fusion des scores.

Cet axe de recherche est motivé par le fait qu'aujourd'hui nous avons à faire face à un réel problème d'acceptation et d'adoption des dispositifs de vigilance et de soins à distance par les utilisateurs finaux et les soignants. En effet, ce genre de systèmes est souvent jugé comme intrusif et non respectueux de la vie intime des personnes équipées. Les capteurs sonores, quant à eux, sont plus facilement acceptés par les patients et/ou les personnes âgées, leur famille et les soignants. Ils sont considérés comme moins intrusifs que les caméras, les vêtements intelligents, etc.

## 2. Le capteur sonore intelligent

Ce capteur comporte un ensemble de microphones connectés à un soft intelligent qui est composé de trois modules :

Le premier module est en permanence à l'écoute de l'environnement sonore dans le but de détecter et d'extraire des sons ou des phrases utiles. Le deuxième module est un module de classification à bas niveau (parole/son). Il traite le son reçu par le premier module afin de séparer les signaux de parole des signaux de son. A titre d'exemple un discours prononcé par le patient sera classé comme 'parole' mais un claquement de porte ou des sons de pas seront classés comme 'son'. Le troisième module traite chaque signal classifié par le deuxième module. Il consiste en deux sous-modules. Dans le cas où un label de son a été attribué au signal, le sous module de reconnaissance des sons classe le signal entre huit classes sonores prédéfinies (bris de verre, claquement

de porte, bruit de vaisselle, etc.). Dans le cas de l'attribution d'une étiquette de parole, le signal extrait est analysé par un moteur de reconnaissance vocale en vue de détecter les expressions de détresse « A l'aide ! », « Un docteur ! », « A moi ! », « Aidez-moi ! », etc.

Dans le but d'améliorer les performances de détection (sensibilité/spécificité) de ce système de télévigilance, nous nous sommes orientés naturellement vers des approches fondées sur la fusion statistique de scores de reconnaissance introduites par [2], [3] et [4] dans le contexte de la reconnaissance bi-modale audio-visuelle.

## 3. Les techniques de fusion

Ces différentes formules de fusion sont basées sur la fusion des scores de reconnaissance obtenus à la sortie de chaque classifieur considéré, on doit prendre comme hypothèse " $H_i$ " l'indépendance conditionnelle entre les deux principales sorties des classifieurs " $X_1$ ", " $X_2$ " obtenues dans le but de détecter avec le plus de robustesse possible un événement de chute du patient. Cette hypothèse fondamentale pour la suite s'exprime alors de la manière suivante:

Elle peut être en effet considérée comme plausible dans la mesure où les deux sorties ou modalités n'ont pas de corrélations directes par rapport à l'hypothèse " $H_i$ ". En appliquant la règle de Bayes, on obtient la formule de fusion de base, dite *UnWeighted Bayesian Product UBP*.

$$p_{UBP}(H_i/x_1, x_2) = \frac{p(H_i/x_1)p(H_i/x_2)}{P(H_i)} \cdot \eta \quad (2)$$

Où  $\eta$  est le paramètre de normalisation [1].

Les deux modalités contribuent de manière équivalente à la prise de décision quelque soit leur niveau de perturbation par le bruit de mesure. On introduit alors un facteur de pondération  $\lambda$  de pondération mutuelle des probabilités obtenues en sortie de chacun des deux classifieurs ici considérés. On obtient alors des versions améliorées de la formule de fusion *UBP* selon l'ajout d'hypothèse sur

les probabilités *a priori*, qui se traduit par la formule suivante :

$$p_{SWP_\lambda}(H_i/x_1, x_2) = \frac{p^\lambda(H_i/x_1)p^{1-\lambda}(H_i/x_2)}{\sum_{j=1}^N p^\lambda(H_i/x_1)p^{1-\lambda}(H_i/x_2)} \quad (3)$$

Où  $\lambda$  est calculées par la formule d'entropie définie dans [3].

On a une autre approche qui consiste à ne plus faire des produit de probobalité mais des sommes de probabilités de sortie de chaque classifieurs comme suit:

$$P_{FC}(H_i/x_1, x_2) = a_1P(H_i/x_1, x_2) + a_2P(H_i/x_1) + a_3P(H_i/x_2) + a_4P(H_i) \quad (4)$$

Et le premier terme de la somme va être remplacé par sa formule défini dans l'équation (2) et on obtient  $P_{FCA}$ .

$$P_{FCA}(H_i/x_1, x_2) = a_1 \frac{P(H_i/x_1)P(H_i/x_2)}{P(H_i)} \cdot \eta + a_2P(H_i/x_1) + a_3P(H_i/x_2) + a_4P(H_i) \quad (5)$$

Dans [3] on peut trouver les expressions des différentes pondérations  $a_i$  en fonction de l'entropie calculée a la sortie de chaque classifieur relative à chaque modalité.

#### 4. Application

Le système de classification utilisé par Anason est basé sur des GMMs générées et appris dans des conditions spécifiques de rapport signal sur bruit (RSB) 0 db, 10 db, 20 db, 40 db ou 60 db. Comme Anason pour la prise de décision se base sur une seule catégorie de GMMs générées sous les mêmes conditions de RSB, nous avons procédé à l'application des techniques de fusion décrites précédemment afin de prendre en compte toutes les cinq catégories de GMMs spécifiées par le RSB.

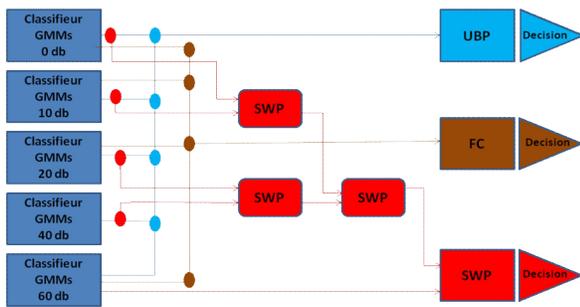


Figure 1 : Architecture des technique de fusion

Pour cela on prend comme hypothèse  $H_i$  l'indépendance conditionnelle des cinq sorties issues de chaque classification basé sur l'une des cinq catégories de GMMs par rapport à l'événement sonore détecté. La technique *UBP* est généralisé sur ces sorties. La technique *SWP* est aussi appliquée, en combinant ces cinq sorties deux par deux. Ensuite la

technique *FC* qui est adaptable pour combiner plusieurs sorties à été testée.

La figure 1 résume ces différentes techniques de fusion appliquée sur Anason.

#### 5. Résultats

En appliquant ces techniques de fusion à la sortie de système Anason on a eu les résultats décrits dans le tableau 1.

Technique	Sans fusion	Après la fusion
UBP	0.13	0.02
SWP $_\lambda$	0.13	0.03
SWP $_{\alpha\beta}$	0.13	0.03
FC	0.13	0.000001

Tableau 1 les résultats d'Anason (Taux d'erreur)

La technique *UBP* a ignoré 2 signaux à 40 dB, nous pouvons dire que cette technique est moins performante dans des conditions moins bruités. En revanche la technique *SWP $_\lambda$*  et *SWP $_{\alpha\beta}$*  ont ignoré 3 signaux à 00 dB ce qui montre que ces deux techniques sont moins performantes dans des conditions bruités. La méthode de *FC* a bien classé tous les signaux, indépendamment des conditions d'apprentissage des GMMs.

#### 5. Conclusion

Ces techniques de fusion nous ont permis d'améliorer le système de télésurveillance de plus 10 %, ce qui nous motive à généraliser ces techniques sur les systèmes de télévigilance à plusieurs modalités tel que le système EMUTEM [5].

#### Références

- [1]. Istrate, E. Castelli, *Information extraction from sound for medical telemonitoring*, IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 2, pp. 264-274, April 2006.
- [2] Kroschel.K, Heckmann M. *Noise Adaptive Stream Weighting in Audio-Visual Speech Recognition*. *La Revue*, 2002:11, 1260–1273
- [3]. Bourlard .H, Mistra. H, *New Entropy Based Combination Rules In HMWANN Multi-stream*, ASR ICASSP 2003.
- [4]. Josef Kittler, Member, *On Combining Classifiers* IEEE Computer Society vol. 20, N. 3, March 1998.
- [5]. 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;