

# Localisation d'habitant dans un espace perceptif par réseau dynamique

Pedro Chahuara, Michel Vacher, François Portet

► **To cite this version:**

Pedro Chahuara, Michel Vacher, François Portet. Localisation d'habitant dans un espace perceptif par réseau dynamique. Représentation et Raisonnement sur le Temps et l'Espace (RTE 2011), May 2011, Chambéry, France. pp.1-10, 2011. <hal-00953559>

**HAL Id: hal-00953559**

**<https://hal.inria.fr/hal-00953559>**

Submitted on 3 Mar 2014

**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.

# Localisation d’habitant dans un espace perceptif par réseau dynamique <sup>\*</sup>

Pedro Chahuara, Michel Vacher, François Portet

Laboratoire d’Informatique de Grenoble (UMR 5217 UJF/INPG/CNRS)  
équipe GETALP, 385 avenue de la Bibliothèque  
38041 Grenoble - France

{Pedro.Chahuara,Michel.Vacher,Francois.Portet}@imag.fr

**Résumé** : Cet article présente une approche de fusion de données temporelles pour localiser une personne dans un environnement perceptif à partir de sources non visuelles. Ces sources informent sur la localisation de manière incertaine et sont donc combinées en utilisant un réseau dynamique à deux niveaux d’hypothèses de localisation et en adaptant une méthode de propagation d’activation pour prendre en compte la validité éphémère et l’ambiguïté des sources. Les résultats sur des enregistrements réels montrent l’intérêt de l’approche.

**Mots-clés** : réseaux dynamiques, propagation d’activation, fusion de données temporelles, Intelligence Artificielle, informatique ubiquitaire.

## 1 Introduction

L’objectif de ce travail est de définir une méthode de fusion de données pour la localisation de l’habitant à l’intérieur de son domicile en utilisant des sources d’informations non visuelles (c.-à-d. sans caméra), événementielles et indirectes (c.-à-d. pas de capteur porté par la personne). Il se déroule dans le cadre du projet SWEET-HOME qui vise à concevoir un système de contrôle intelligent de la domotique à travers une interface de commande utilisant la voix pour un plus grand confort et une plus grande sécurité de la personne. La localisation de la personne est primordiale pour interpréter les commandes vocales (p.ex., si la personne prononce « allume la lumière », il est nécessaire de localiser la personne pour déduire la lampe à activer). Les approches présentées dans la littérature reposent souvent sur l’utilisation de la vidéo (Friedland *et al.*, 2010). Cependant, pour limiter les coûts et pour respecter la vie privée, seuls les capteurs domotiques sont considérés dans le projet. Ces sources (ici capteurs de présence, capteurs de contact, et microphones) sont donc ambiguës (p.ex., observation d’un même évènement par plusieurs capteurs éloignés), ont une dimension temporelle (p.ex., dans le couloir à 15 : 23 : 45) et dépendent de l’espace physique (p.ex., nombre de pièces, réverbération).

Parmi les capteurs considérés, le microphone est un capteur d’évènements et de localisation prometteur qui, par sa nature omnidirectionnelle ou unidirectionnelle, peut être

---

\*. Ce travail a été réalisé dans le cadre du projet Sweet-Home qui bénéficie d’une aide de l’Agence Nationale de la Recherche portant la référence ANR-09-VERS-011

très sensible ou très spécifique. Cependant, très peu d'approches du domaine utilisent uniquement l'information sonore. On peut citer Bian *et al.* (2005) qui ont développé une grille de microphones dans une pièce pour déterminer l'origine d'un son dans un espace 3D avec une erreur moyenne de moins de 27 cm. Comme souligné par les auteurs, l'information audio nécessite beaucoup moins de bande passante que l'information vidéo et permet de détecter facilement certaines activités (conversations, sonneries de téléphone). Néanmoins, cette solution est trop complexe et trop coûteuse à mettre en place dans un appartement classique. De plus, si la vidéo est sensible aux changements de luminosité, le canal audio est sensible au bruit environnemental (Vacher *et al.*, 2011). Le canal audio, bien qu'étant une modalité pertinente et peu coûteuse, est donc une source bruitée et ambiguë. Il devient donc essentiel de mettre en place un système de localisation multisource qui tirerait partie des redondances et des complémentarités des sources choisies. Plusieurs travaux ont été dans ce sens, en utilisant principalement des méthodes de classification probabilistes telles que les réseaux bayésiens (Dalal *et al.*, 2005) ou les modèles de Markov (Kröse *et al.*, 2008). Cependant, l'information temporelle dans ces modèles est assez pauvre et rarement prise en considération. Duong *et al.* (2009) prennent en compte l'aspect temporel à travers d'un modèle semi-Markov caché hiérarchique pour la reconnaissance d'activités à partir de séquences d'événements en modélisant aussi le temps passé dans chaque état (représentant une sous-activité). Cependant, dans notre cas, il est difficile de prédire la localisation présente en évaluant une séquence de localisations passées car le nombre de chemins de déplacements possibles augmenterait l'incertitude. De plus, ces approches nécessitent un grand nombre de données étiquetées pour l'apprentissage. Certaines applications mettant en jeu des traitements sur la connaissance, telles que la recherche d'information (Crestani, 1997; Aswath *et al.*, 2005), la reconnaissance de forme (Niessen *et al.*, 2008) ou la navigation d'agents animés (Kortenkamp & Chown, 1993), emploient une approche de propagation d'activation à travers un réseau sémantique. Ces réseaux sont constitués de nœuds représentant des concepts et de liens représentant l'intensité de la relation entre concepts. L'activation d'un nœud se dilue de proche en proche en activant les nœuds voisins jusqu'à ce qu'elle s'épuise. Des travaux récents (Niessen *et al.*, 2008) ont montré l'intérêt de la propagation d'activation comme mémoire à court terme pour désambiguïser des événements sonores.

L'approche, que nous avons adoptée pour localiser une personne, se base sur un réseau dynamique où les événements des capteurs activent des hypothèses de localisation qui se périment au cours du temps. Ces hypothèses prennent en compte les activations précédentes, l'incertitude des événements et la durée de validité d'un événement. Dans notre approche, le réseau est dynamique dans le sens où il évolue en fonction des données mais aussi où il prend en compte des connaissances a priori sur l'environnement dans lequel évolue l'habitant. Cette approche est détaillée dans la section 2. La méthode proposée est ensuite évaluée en section 3 sur des données réelles. L'article se termine par une brève discussion.

## 2 Localisation d'habitant par propagation d'activations multisource

La méthode mise en œuvre pour la localisation d'une personne à partir de plusieurs sources d'information non-visuelles repose sur la modélisation des liens entre les événements et les hypothèses de localisation par un réseau dynamique à deux niveaux.

Chaque nouvelle information est propagée dans le réseau pour mettre à jour les hypothèses de localisation. Dans la plupart des approches de propagation d'activation, le réseau est bien établi, mais, dans notre cas, les informations ont une validité temporaire et les liens entre les nœuds évoluent en fonction du temps. La section 2.1 présente comment le réseau évolue au fil du temps en intégrant les données de plusieurs sources temporelles et comment cette information se « périmé » grâce à une fonction d'oubli. La section 2.2 détaille comment des connaissances a priori sur l'environnement peuvent être prise en compte avec l'information immédiate des capteurs pour calculer les relations entre les différents nœuds du réseau.

## 2.1 Réseaux dynamiques et propagation d'activation

La méthode que nous utilisons est basée sur celle de Niessen *et al.* (2008) qui ont présenté une approche basée sur les réseaux dynamiques permettant de désambigüiser la reconnaissance d'évènements sonores. Il s'agit d'un réseau à deux niveaux. Le niveau zéro est constitué des évènements sonores, le niveau 1 représente les hypothèses liées à un évènement (p.ex. : rebond de ballon ou claquement de mains), et le niveau deux représente le contexte de l'évènement (p.ex, match de basket, concert). Chaque évènement active des hypothèses selon l'évènement et les contextes auxquels les hypothèses sont liées. L'activation des hypothèses se propage ensuite aux contextes. Nous avons adapté cette approche pour localiser une personne à partir de différentes sources d'évènements. Le réseau dynamique, que nous avons conçu, est organisé en deux niveaux : le premier niveau correspond aux *hypothèses de localisation* générées à partir d'un évènement, et le deuxième niveau représente les *contextes d'occupation* de chaque pièce ; le poids d'activation des différents contextes indique la localisation la plus probable sachant les évènements précédents. La méthode utilise les définitions suivantes :

### Définition 1 (observation)

Une observation  $o_{t_n}$  est une donnée structurée générée par un capteur ayant réagit à un évènement  $e_{t_n}$  au temps  $t_n$  avec  $t, n \in N$ . À chaque observation  $o$  est associée le type de capteur  $o.type$ , l'origine du capteur  $o.captteur$  ainsi que le rapport signal-à-bruit (RSB)  $o.rsb$  dans le cas de signaux numériques.

### Définition 2 (observations simultanées)

Deux observations  $o_{t_n}^i$  et  $o_{t_k}^j$  sont dites simultanées si  $t_n = t_k$  et  $i \neq j$ .

### Définition 3 (hypothèse de localisation)

$h_{t_n}^i \in L$ , où  $L = \{Loc_1, \dots, Loc_R\}$  est le nœud hypothèse que l'habitant se trouve à la  $i^e$  localisation à l'instant  $t_n$ . Les hypothèses sont générées uniquement à partir des observations au temps  $t_n$ .

### Définition 4 (contexte d'occupation)

$c^i \in P$  où  $P = \{Piece_1, \dots, Piece_R\}$  est le nœud de contexte d'occupation de la  $i^e$  pièce. La valeur d'activation du contexte varie en fonction du temps et des hypothèses.

### Définition 5 (poids de relation)

$w \in [0, 1]$  est la force de la relation entre deux nœuds du réseau. Ainsi,  $w_{o, h^i}$  est le poids du lien entre une observation  $o$  et la  $i^e$  hypothèse et  $w_{h^i, c^j}$  est le poids du lien entre la  $i^e$  hypothèse et le  $j^e$  contexte.

**Définition 6 (fonction d'oubli)**

La fonction d'oubli  $f(t_n, t_{n-1}) = e^{-\frac{\Delta_t}{\tau}}$ , avec  $\Delta_t = t_n - t_{n-1}$ , représente la décroissance d'un contexte au cours du temps. Elle permet de conserver une mémoire à court terme des contextes. Plus les observations seront espacées au cours du temps (p.ex.,  $\Delta_t > 3 \times \tau$ ), plus les contextes seront oubliés.

**2.1.1 Évolution du réseau dynamique**

L'algorithme fonctionne ainsi :

1. pour toute nouvelle observation  $o_{t_n}^k$  un nouveau nœud est créé ;
2. des nœuds hypothèses  $h_{t_n}^i$  sont ensuite créés et connectés aux  $o_{t_n}^k$  avec les poids  $w_{o^k, h^i}$  ;
3. les hypothèses  $h_{t_n}^i$  sont connectées aux contextes de localisation  $c^j$  avec les poids  $w_{h^i, c^j}$  ;
4. les activations sont propagées des  $o_{t_n}^k$  aux  $h_{t_n}^i$  et l'activation des  $h_{t_n}^i$  est calculée ;
5. les activations sont propagées de  $h_{t_n}^i$  à  $c^j$  et l'activation de  $c^j$  est recalculée ;
6. le  $c^j$  avec la plus forte activation devient la localisation courante ;
7. tous les nœud  $h_{t_n}^i$  et  $o_{t_n}^k$  sont supprimés du réseau.

Un exemple de réseau dynamique est donnée par la figure 1. À l'instant  $t_{n-2}$ , l'évènement  $e_{t_{n-2}}$  est capté par un capteur qui génère une observation  $o_{t_{n-2}}$  de laquelle 3 hypothèses sont déduites :  $h_{t_{n-2}}^1$  avec une relation de 0.1 vers le contexte  $c^1$ ,  $h_{t_{n-2}}^2$ , avec 0.6 vers  $c^2$  et  $h_{t_{n-2}}^3$  avec 0.3 vers  $c^3$ . S'il n'y a pas eu d'évènement antérieur,  $c^2$  est la localisation la plus probable. À l'instant  $t_{n-1}$ , l'évènement  $e_{t_{n-1}}$  est observé par deux capteurs. Les nœuds générés à  $t_{n-2}$  disparaissent, seul les contextes sont conservés. Si ces contextes sont encore actifs leur activation sera pondérée par  $f(t_{n-1}, t_{n-2})$  auxquels vont s'ajouter les activations générées par les hypothèses  $h_{t_{n-1}}^1$  avec une relation de 0.2 vers le contexte  $c^2$  et  $h_{t_{n-1}}^2$  avec 0.8 vers  $c^4$ . La méthode s'appliquera ensuite de la même manière à l'instant  $t_n$ , etc.

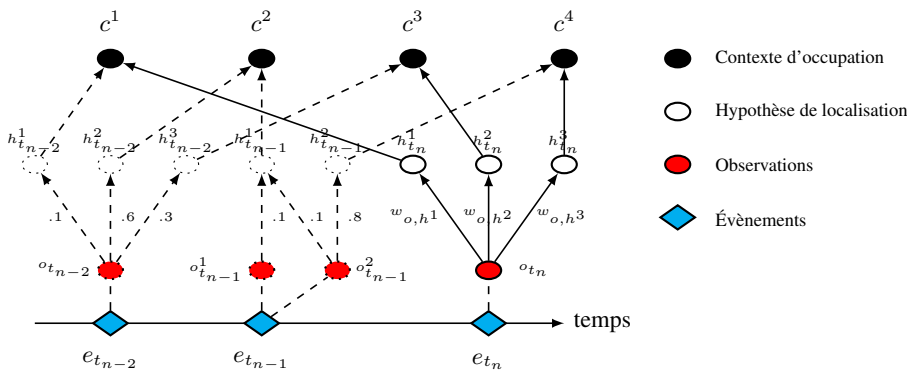


FIGURE 1: Exemple de réseau dynamique

### 2.1.2 Propagation de l'activation

L'activation recueillie par un nœud est classiquement définie (Crestani, 1997; Niessen *et al.*, 2008) par la formule  $n_i(t) = \sum_{i \neq j} w_{i,j} \times A^j(t)$  où  $w_{i,j}$  est le poids,  $j$  correspond à un voisin de  $i$  et  $A^j(t)$  est l'activation de ce voisin au temps  $t$ . Bien entendu, un nœud qui a été activé par un voisin ne peut communiquer de nouveau cette activation à ce voisin. Dans notre cas, les activations sont toujours initiées par une observation  $o$  qui aura à chaque fois une activation maximale  $A^o(t) = 1$  quelque soit son type et se propagera de bas en haut. Une fois l'activation  $n(t)$  provenant des voisins accumulée, l'activation de sortie du nœud doit être calculée. Celle-ci diffère selon le niveau des nœuds. Pour les hypothèses de localisation, l'activation  $A^{h^i}(t) \in [0, 1]$  sera calculée selon la formule 1 :

$$A^{h^i}(t_n) = n_i(t_n) = \sum_{o \in O_{t_n}} w_{o,h^i} A^o(t_n) \quad (1)$$

où  $O_{t_n}$  est l'ensemble des observations simultanées à  $t_n$  liées à  $h^i$  et  $\sum_{o \in O_{t_n}} w_{o,h^i} \leq 1$ . Pour les contextes d'occupation de pièce, l'activation de sortie résulte de l'accumulation des activations provenant des hypothèses et de l'activation précédente pondérée par un facteur d'oubli. L'équation 2 décrit la loi d'activation du contexte de localisation  $A^{c^i}$  suite à une activation externe au temps  $t_n$ .

$$A^{c^i}(t_n) = n_i(t_n) \times [M - e^{-\frac{\Delta t}{\tau}} A^{c^i}(t_n - \Delta t)] + e^{-\frac{\Delta t}{\tau}} A^{c^i}(t_n - \Delta t) \quad (2)$$

où  $A^{c^i}(t_n - \Delta t)$  est l'activation précédente,  $M = 1$  est l'activation maximale et  $e^{-\frac{\Delta t}{\tau}}$  est la fonction d'oubli. Ainsi, si aucun évènement n'apparaît pendant  $5 \cdot \tau$  secondes, l'activation des contextes pourra être considérée comme nulle. L'introduction de  $M$  permet de maintenir toutes les activations entre 0 et 1.

## 2.2 Calcul des relations entre les différentes couches du réseau

Le réseau dynamique étant constitué de deux niveaux bien définis (cf. figure 1), deux types de relations existent, la relation *observation-hypothèse* et la relation *hypothèse-contexte*. Les liens entre les différentes couches dépendent fortement de l'application et de l'environnement considéré, c'est pourquoi il est nécessaire d'introduire tout d'abord l'environnement applicatif pour faciliter la compréhension de l'article.

### 2.2.1 Évènements considéré dans l'environnement perceptif

Un exemple d'environnement perceptif considéré dans cette étude est l'Habitat Intelligent pour la Santé (HIS) de la Faculté de Médecine de Grenoble (Le Bellego *et al.*, 2006). La figure 2 décrit l'appartement intelligent. Il est composé de 6 pièces contenant un ensemble de capteurs pour percevoir l'environnement : 7 microphones (*Mic*) placés dans le plafond (hauteur d'environ 3 mètres) ; 3 capteurs de contact sur des portes (*CP*) de meubles ; et 6 détecteurs de présence infrarouge (*PID*) placés sur les murs. Parmi les données recueillies par les capteurs (*Mic*, *CP* et *PID*), seuls les signaux sonores doivent subir un pré-traitement. En effet, les données *CP* et *PID* délivrent une information booléenne déjà représentative d'évènements. Les signaux audio ont été traités par le système AUDITHIS (Vacher *et al.*, 2010). Succinctement, les évènements audio

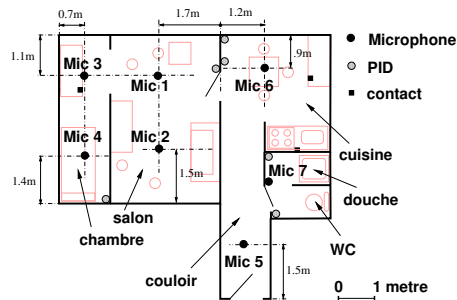


FIGURE 2: Plan de l'appartement et position des capteurs.

sont détectés, en temps réel, et sont ensuite classés comme parole ou comme son de la vie courante (p.ex. chute d'objet). Les microphones étant omnidirectionnels, un événement sonore est généralement capté par plusieurs microphones en même temps. Pour chaque observation  $o$  de notre étude,  $o.type \in \{CP, Mic, PID\}$ .

### 2.2.2 Relation hypothèse-contexte

La relation *hypothèse-contexte* est dans notre cas une relation univoque car une hypothèse de localisation n'est liée qu'à une seule pièce. Il s'agit d'un choix expérimental car certaines hypothèses sur des pièces faiblement séparées (p.ex. salon/chambre) pourraient activer plusieurs contextes de pièce. Par ailleurs, l'application visée étant la localisation d'un habitant dans une pièce, tous les nœuds contextes de pièces sont créés à l'initialisation du réseau et en font partie en permanence.

### 2.2.3 Relation observation-hypothèse

La relation *observation-hypothèse* est unidirectionnelle et de type de un vers plusieurs. Le poids et les hypothèses générés varient en fonction de la source des observations. Pour représenter cette relation dans le réseau, deux types d'information sont considérées : celles qui sont dynamiques et dépendent des événements, et celle qui sont statiques provenant des connaissances *a priori* sur les capteurs.

Certaines connaissances sur l'environnement peuvent être utilisées pour définir les liens qui existent entre concepts. Pour une observation  $o$ , avec  $o.type \in \{CP, PID\}$ , un seul nœud hypothèse de localisation est créé. Les informations spatiales des *PID* et des *CP* sont univoques et certaines. Par exemple, l'ouverture du frigo ne peut se produire que si la personne se trouve dans la pièce. Par conséquent, un événement est lié avec un poids  $w_{o,h} = 1$  à une hypothèse unique de localisation, et, l'hypothèse est liée avec un poids de  $w_{h,c} = 1$  au contexte. Il n'en va pas de même pour les informations des microphones. En effet, les microphones peuvent théoriquement capter toutes les ondes acoustiques générées dans les pièces d'une habitation. L'énergie du signal sonore détecté fournit bien une information dynamique sur la proximité de la source mais, il ne donne, par contre, aucune indication sur sa direction. Par exemple, un bruit capté à 1 mètre d'un microphone peut tout à fait avoir été généré dans une pièce différente de celle dans laquelle le microphone est fixé. Pour prendre en compte ce phénomène, nous avons modélisé l'ambiguïté d'un microphone en fonction de sa distance avec les

autres pièces par les poids  $w_{o,h^i}$  des relations dans le réseau dynamique. La valeur de ces poids a été calculée en estimant la probabilité  $P(loc = i | Mic = j)$  que l'habitant soit à la  $i^e$  localisation sachant qu'un évènement a été détecté sur le  $j^e$  microphone. Le poids de la relation entre l'hypothèse  $h^i$  et l'observation  $o$  est donné par la formule 3 :

$$w_{o^k,h^i}(t_n) = \frac{P(loc = i | o_{t_n}^k . capteur) \times o_{t_n}^k . rsb}{\sum_{l \in L} \sum_{o \in O_{t_n}} P(loc = l | o . capteur) \times o . rsb} \quad (3)$$

où  $O_{t_n}$  est l'ensemble des observations simultanées provenant des microphones à l'instant  $t_n$ ,  $P(loc | Mic)$  est une connaissance *a priori* décrite en section 2.2.4,  $L$  est l'ensemble des localisations possibles et  $o . rsb$  le rapport signal sur bruit linéarisé de  $o$ . Ainsi  $w_{o^k,h^i}$  est un poids normalisé ( $\sum_{i,k} w_{o^k,h^i} = 1$ ) qui représente le lien entre les observations sonores simultanées et les hypothèses de localisation.

### 2.2.4 Acquisition des informations statiques

Pour calculer la probabilité  $P(loc = i | Mic = j)$ , deux approches ont été testées : l'approche naïve et l'approche par analyse statistique sur corpus. Pour l'approche naïve, la référence sonore a été choisie à 1 mètre au dessous du microphone. À partir de cette référence, un cercle de 2 mètres de rayon a été tracé autour de chaque microphone, il correspond à une atténuation de  $-6dB$  en considérant l'atténuation quadratique classique. Au delà de ce cercle, la perte d'énergie est supérieure à 75%. Le poids est calculé selon la surface de l'intersection entre le cercle et les pièces, avec une pénalité de 2 lorsqu'une cloison est traversée. L'autre approche est statistique, elle calcule directement les probabilités à partir du corpus annoté. Le tableau 1 indique les poids obtenus pour l'approche naïve et pour l'approche statistique.

Le mélange d'informations a priori et d'informations dynamiques permet une meilleure désambiguïsation. Par exemple, dans le cas de deux observations sonores simultanées captées par le micro 6 de la cuisine et le micro 7 de la salle de bain avec un RSB identique égal à  $12dB$ , les formules 1 et 3 et l'a priori naïf donnent pour l'activation  $A^{h^{cuisine}} = rsb \times (.87 + .5) / [rsb \times (.87 + .5 + .18 + .13 + .03 + .18 + .1)] = 0,69$  qui sera supérieure à celle de la salle de bain  $A^{h^{sdb}} = 0,09$  bien que le rapport signal à bruit soit le même.

## 3 Résultats

L'approche a été testée sur des données réelles acquises lors d'expériences (Fleury *et al.*, 2010) destinées à évaluer la reconnaissance automatique des activités de la vie

Mic	$P(Loc Mic)$ estimé par la méthode naïve							$P(Loc Mic)$ estimé à partir du corpus						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7
chambre	.14	.07	.70	.85				.28	.29	.42	.43	.25	.18	.20
salon	.86	.93	.27	.14	.01	.13	.03	.59	.56	.47	.41	.07	.07	.09
cuisine			.03	.02	.10	.87	.50	.05	.08	.06	.09	.45	.63	.37
douche					.06		.18	.06	.05	.04	.04	.09	.04	.10
wc					.06		.18	.01	.01	.01	.01	.12	.05	.21
couloir					.77		.10		.02	.01	.02	.03	.03	.02

TABLE 1: Estimation de  $P(Loc|Mic)$



Capteurs	Mic	PID	CP	Mic+CP	PID+Mic	PID+CP	PID+Mic+CP
sans a priori	25.7	<b>88.9</b>	26.5	32.8	87.7	89.4	88.2
a priori naïf	30.2	88.9	26.5	34.1	89.0	89.4	<b>89.5</b>
a priori naïf sujet 4	25.3	92.8	19.5	25.4	96.8	92.8	<b>96.8</b>
a priori statistique	30.9	88.9	26.5	34.8	89.7	89.4	<b>90.1</b>

TABLE 2: Exactitude (%) avec plusieurs combinaisons de sources

quotidienne d'une personne à son domicile dans le but de pouvoir détecter automatiquement une perte d'autonomie. Dans le HIS (cf. section 2.2.1), 15 volontaires ont effectué chacun au moins une fois 7 activités prédéfinies (p.ex., dormir, faire sa toilette, communiquer avec l'extérieur) sur une période d'environ 1 heure, sans consigne sur la manière ou l'ordre dans lequel elles devaient être effectuées. Les données ont ensuite été annotées grâce à des caméras placées uniquement à cet effet. Pour chaque enregistrement, les observations enregistrées par les capteurs *CP*, *PID* et *Mic* ont activé un réseau dynamique. Les traces d'activation de chaque contexte ont été conservées et l'emplacement de départ de l'expérimentation est supposé connu (ici la cuisine). Les performances de localisation ont été évaluées à l'échelle de la seconde. À chaque seconde, le contexte de localisation de plus haut poids est comparé à la vérité terrain. S'il y a correspondance, alors il s'agit d'un vrai positif (VP), sinon il s'agit d'une confusion. Ceci nous permet de générer une table de confusion à partir de laquelle le taux d'exactitude est calculé par  $TE = \frac{nb(VP)}{nb(test)}$ , où  $nb(test)$  correspond à la durée de l'évaluation exprimée en secondes. Le tableau 2 montre les résultats sur les 47226 secondes (13h 7min 6s) d'enregistrement des 15 participants pour chaque combinaison de capteurs. Dans le cas sans a priori (c.-à-d.,  $P(loc = i | Mic = j) = 1$  si le capteur  $j$  est dans  $i^e$  pièce 0 sinon), l'ajout des contacts de porte *CP* améliore légèrement la précision obtenue avec seulement les capteurs infrarouges *PID* mais l'information sonore dégrade légèrement les performances (88,2% au lieu de 88,9%). Lorsqu'un a priori sur les relations événement sonore - hypothèse de localisation est ajouté et calculé par la méthode naïve, dans ce cas, comme dans la méthode statistique, nous observons une amélioration des performances. La connaissance a priori du lien entre un son et un capteur permet de corriger un grand nombre d'erreurs de localisation. Le gain reste faible néanmoins, lorsque le changement de pièce n'est pas détecté par les *PID* mais que les paroles sont bien identifiées, l'information sonore est utile pour améliorer la localisation (comme dans le cas du sujet 4) en compensant le défaut de sensibilité des infrarouges. Ceci illustre l'intérêt de la combinaison de plusieurs sources.

Pour le participant 4 et dans le cas a priori naïf, la valeur de l'activation des contextes de localisation pour chacune des occurrences d'événements est tracée sur la figure 3. Dans cet extrait, le participant est dans le séjour (événements 1 à 9), puis il se rend à la cuisine (événements 10 à 15) avant de revenir au séjour. Les événements donnant des informations sur cette transition sont sonores (moins précis que les *PID*). La méthode estime la localisation dans la douche (événements 16 à 18) avant de trouver la vraie localisation. Ensuite, le participant reste dans le salon (événements 16 à 60) pour effectuer un appel téléphonique. Au cours de la période dans le séjour, les deux contextes séjour et chambre s'affrontent car la conversation téléphonique se tient à proximité de la limite du salon et de la chambre à coucher. Le participant étant resté presque immobile pendant la conversation, il n'y a pas eu d'événement détecté par les *PID* et les microphones ont donc constitué la seule source d'information.

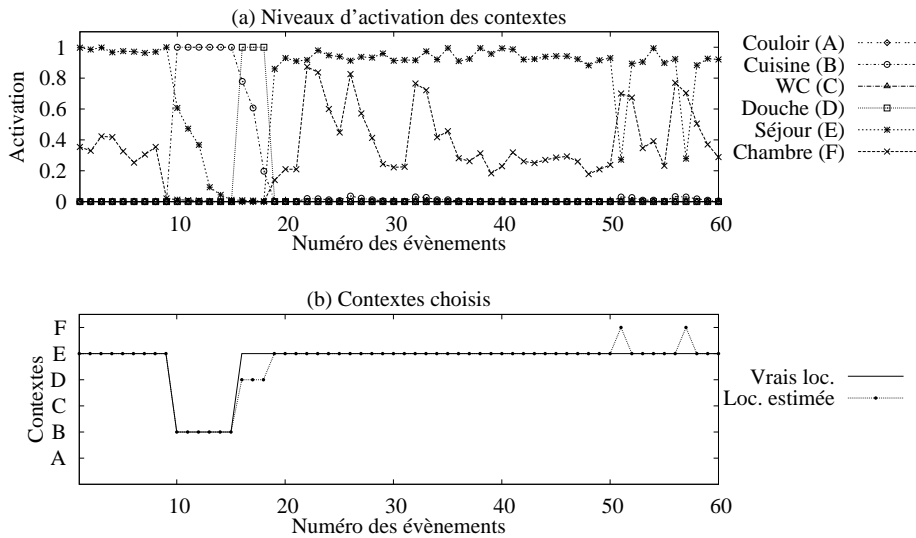


FIGURE 3: Extrait de l'évolution des niveaux d'activation des contextes (a) et contextes choisis pour les 6 localisations possibles (b).

## 4 Discussion

L'article présente une méthode de fusion d'informations multisources multimodales et temporelles par propagation d'activation dans un réseau dynamique qui permet de modéliser la connaissance statique (liée à l'organisation de l'espace) et dynamique (temporelle). Cette approche permet de fusionner des sources non visuelles avec une grande flexibilité et n'impose pas une perception continue de l'espace ni de patron d'évènements. De plus, cette méthode permet de prendre en compte facilement de l'a priori dans le réseau afin de tenir compte de connaissances expertes.

Les résultats montrent que la fusion d'information par propagation d'activation multi-source présente un intérêt même dans le cas où les sources ont une très bonne précision, c'est le cas des capteurs infrarouges (mais avec une sensibilité imparfaite). L'utilisation de sources de localisation plus incertaines, telles que la détection de parole, peut alors améliorer les performances. Malgré l'ambiguïté inhérente aux microphones omnidirectionnels, la prise en compte de la connaissance a priori du lien entre un son et un microphone permet de corriger un grand nombre d'erreurs de localisation. Cependant, l'utilisation unique des informations de parole conduit cette source à être très peu sensible. Dans notre étude, quasiment 60% des confusions sont dues à des périodes durant lesquelles le sujet ne parle pas, n'utilise pas d'objet et dont les mouvements ne sont pas captés par les PID. Nous prévoyons donc d'inclure la classification de différents types de sons afin de capter les mouvements générateurs de bruits qui ne peuvent être détectés par les PID.

Plusieurs autres pistes d'amélioration de cette méthode vont être suivies. Une première piste consiste à agir sur la constante de temps en fonction de la sémantique des évènements et des lieux. Par exemple, une détection dans le couloir devrait avoir une

constante de temps plus faible que dans le cas des toilettes. À cet effet, nous prévoyons d'extraire les constantes de temps potentielles à partir des données. Sur le plus long terme, la méthode présentée peut être améliorée en utilisant la théorie de Dempster-Shafer. Une fusion d'information serait alors possible pour déterminer la localisation par consensus entre le niveau de croyance des sources d'information et les poids d'activation des hypothèses. Cette théorie de combinaison de preuves offre un cadre de travail plus formel que la combinaison de poids. Enfin, nous prévoyons d'appliquer cette méthode à la classification des sons de la vie courante en utilisant le contexte de la localisation pour désambiguïser les hypothèses de classification des sons.

## Références

- ASWATH D., D'CUNHA J., AHMED S. T. & DAVULCU H. (2005). Boosting item keyword search with spreading activation. In *WI'05*, p. 704–707.
- BIAN X., ABOWD G. D. & REHG J. M. (2005). Using sound source localization in a home environment. In *Pervasive 2005*, p. 19–36.
- CRESTANI F. (1997). Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, **11**(6), 453–482.
- DALAL S., ALWAN M., SEIFRAFI R., KELL S. & BROWN D. (2005). A rule-based approach to the analysis of elders activity data : Detection of health and possible emergency conditions. In *AAAI Fall 2005 Symposium*.
- DUONG T., PHUNG D., BUI H. & VENKATESH S. (2009). Efficient duration and hierarchical modeling for human activity recognition. *Artificial Intelligence*, **173**(7-8), 830–856.
- FLEURY A., VACHER M., PORTET F., CHAHUARA P. & NOURY N. (2010). A multi-modal corpus recorded in a health smart home. In *LREC MMC Workshop*, p. 99–105, Matla.
- FRIEDLAND G., YEO C. & HUNG H. (2010). Dialocalization : Acoustic speaker diarization and visual localization as joint optimization problem. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **6**(4), 27 :1–27 :18.
- KORTENKAMP D. & CHOWN E. (1993). A directional spreading activation network for mobile robot navigation. In *Second International Conference on Simulation of Adaptive Behavior*, p. 218–224.
- KRÖSE B., VAN KASTEREN T., GIBSON C. & VAN DEN DOOL T. (2008). Care : Context awareness in residences for elderly. In *The 6th International Conference of the International Society of Gerontology*, Pisa, Italy.
- LE BELLEGO G., NOURY N., VIRONE G., MOUSSEAU M. & DEMONGEOT J. (2006). A model for the measurement of patient activity in a hospital suite. *IEEE Transactions on Information Technology in Biomedicine*, **10**(1), 92 – 99.
- NIESSEN M. E., VAN MAANEN L. & ANDRINGA T. C. (2008). Disambiguating sounds through context. In *ICSC2008*, p. 88–95 : IEEE Computer Society.
- VACHER M., FLEURY A., PORTET F., SERIGNAT J.-F. & NOURY N. (2010). *Complete Sound and Speech Recognition System for Health Smart Homes : Application to the Recognition of Activities of Daily Living*, p. 645 – 673. Intech Book.
- VACHER M., PORTET F., FLEURY A. & NOURY N. (2011). Development of audio sensing technology for ambient assisted living : Applications and challenges. *International Journal of E-Health and Medical Communications*, **2**(1), 35–54.