

Localisation d'habitant dans un environnement perceptif non visuel par propagation d'activations multisource

Pedro Chahuara, Michel Vacher, François Portet

► **To cite this version:**

Pedro Chahuara, Michel Vacher, François Portet. Localisation d'habitant dans un environnement perceptif non visuel par propagation d'activations multisource. MajecSTIC 2010, Oct 2010, Bordeaux, France. pp.1-8, 2010. <hal-00953567>

HAL Id: hal-00953567

<https://hal.inria.fr/hal-00953567>

Submitted on 28 Feb 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 environnement perceptif non visuel par propagation d'activation multisource

Pedro Chahuara¹, Michel Vacher¹ et François Portet¹

1 : Laboratoire d'Informatique de Grenoble (UMR 5217 UJF/INPG/CNRS), équipe GETALP, 385 avenue de la Bibliothèque, 38041 Grenoble cedex9 - France.

Contact : Pedro.Chahuara@imag.fr

Résumé

Cet article présente une approche pour localiser une personne dans un environnement perceptif à partir de sources non visuelles. L'information extraite des capteurs (événements) informe sur la localisation d'une personne de manière incertaine. Ces différentes sources sont 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 dimension temporelle. Les résultats préliminaires sur un enregistrement réel montrent que la fusion d'information permet d'atteindre une exactitude pouvant atteindre 90%.

Abstract

In this paper, an approach to locate a person using non visual sensors in a pervasive environment is presented. The information extracted from the sensors (events) gives uncertain evidence about the location of a person. These sources are combined using a dynamic network to represent two levels of location hypotheses and using an adapted spreading activation method that considers the temporal dimension to be able to deal with evidence that expire. The preliminary results on an actual record showed that an accuracy of 90% can be reached using several uncertain sources.

Mots-clés : réseaux dynamiques, propagation d'activation, fusion de données temporelles, Intelligence Artificielle, bâtiment intelligent

Keywords: dynamic networks, spreading activation, temporal data fusion, artificial intelligence, smart home

1. Introduction

L'objectif de ce travail est la localisation de l'habitant dans son domicile en utilisant des sources d'information non visuelles (c.-à-d. sans caméra) et indirectes (la personne ne porte pas de capteur). Il se déroule dans le cadre du projet Sweet-Home¹ qui vise à concevoir un système de contrôle intelligent de la domotique à travers une interface vocale pour un plus grand confort et une plus grande sécurité de la personne. Une bonne localisation de la personne est primordiale pour bien interpréter les commandes vocales (p.ex. si la personne prononce "allume la lumière", la localisation de la lampe doit être déduite). Afin de limiter les coûts et de respecter la vie privée, seuls les capteurs domotiques sont considérés. Cependant, ces capteurs (ici capteurs infrarouges, contacts de porte, et microphones) n'informent qu'implicitement sur la localisation d'une personne. Les sources sont donc incertaines et ont une dimension temporelle. L'approche que nous avons adoptée pour localiser une personne se base sur une fusion multisource des informations obtenues des différents capteurs (événements) à travers un réseau dynamique où les événements des capteurs activent des hypothèses de localisation qui prennent en compte les activations précédentes et l'incertitude des événements.

1. <http://sweet-home.imag.fr/>

Après une brève présentation de l'état de l'art en section 2, l'article introduit les réseaux dynamiques et détaille notre approche en section 3. Cette méthode est évaluée en section 4 sur des données réelles. L'article se termine par une brève discussion des résultats et des perspectives d'amélioration.

2. État de l'art

Les techniques de localisation de personnes dans un environnement perceptif peuvent être divisées en deux catégories : celles qui utilisent des capteurs explicitement dédiés à cette tâche et portés par les personnes tels qu'un bracelet GPS, et, celles qui utilisent des informations implicitement porteuses d'information sur la présence d'une personne dans un espace restreint tels que les capteurs de présence infrarouges ou les caméras vidéos de surveillance. Les capteurs portables sont souvent adoptés dans les situations où la personne a une activité sociale (visite de musée) ou bien pour des raisons professionnelles de santé ou de sécurité (fugue de personnes âgées). Les techniques de localisation qui reposent sur ces capteurs obtiennent généralement de très bonnes performances en localisation (et identification), cependant elles ne sont pas adaptées à un usage informel et confortable. En effet, ces capteurs peuvent être lourds et encombrants et, sauf pour les capteurs passifs, doivent faire l'objet d'une vérification systématique des piles. De plus, si l'objectif domotique est d'améliorer le confort de vie, la contrainte d'un capteur portable constitue une intrusion forte dans la vie intime ; très souvent il est vite oublié ou délaissé. C'est la raison pour laquelle nous nous concentrerons dans cet état de l'art sur les techniques utilisant des capteurs d'environnement (vidéo, son, capteur de mouvements, contact de portes, etc.).

L'analyse vidéo est une modalité très intéressante pour la domotique et est utilisée dans de nombreux projets liés à cette thématique : Place_Lab (MIT) [9], Aware Home (Georgia Tech) [3] et GER'HOME [15]. Cependant le traitement vidéo nécessite des ressources en calcul et en stockage très importantes qui impliquent un coût important d'installation et de maintenance. De plus, notre expérience nous a appris qu'installer des caméras d'enregistrement dans un logement peut être perçu comme une trop grande intrusion dans la vie quotidienne.

Une autre source de localisation peut être dérivée des appareils électroménagers et de surveillance. Par exemple, dans le projet AILISA [11], des capteurs passifs infra rouge prévus pour l'éclairage automatique sont utilisés pour évaluer la position et l'activité de la personne. Par ailleurs, l'utilisation de matériel électroménager et de rangement peut aussi informer sur la localisation d'une personne dans l'habitat (l'ouverture de l'armoire à vêtement indique que la personne est dans la chambre). Cette information est généralement obtenue par des capteurs fixés sur les portes du mobilier ou sur les appareils ménagers considérés. Par exemple, dans [13], un ensemble de capteurs de ce type est utilisé conjointement avec l'analyse audio pour reconnaître les activités de personnes dans un habitat intelligent pour la santé. L'information d'utilisation de ces appareils peut aussi se faire à moindre coût par des nouvelles techniques d'analyse qui permettent de détecter les signatures de mise en route de ces appareils sur le réseau électrique de l'habitat [2]. Cependant, la plupart des appareils sont maintenant programmables (p.ex. four) et/ou utilisable à distance (p.ex. téléphone), l'information de localisation devient donc moins sûre à moins d'utiliser un système domotique complet comme proposé dans le projet Sweet-Home.

Une autre modalité intéressante en domotique est l'analyse du canal audio, qui, en plus de permettre une commande vocale, peut apporter diverses informations sonores, telles que les bris de verre, claquement de portes, etc. [13]. De part sa nature omnidirectionnelle, le microphone est aussi un capteur d'événements et de localisation prometteur. Par exemple, Bian et coll. [3] 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). Cependant, cette solution de grille de micros 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. Dans Vacher et coll. [14], des enregistrements audio, effectués sur une quinzaine de personnes pendant plus de 11 heures dans un appartement, ont montré un nombre très important de bruits extérieurs perturbant le système de reconnaissance. Le canal audio, bien qu'étant une modalité pertinente et peu coûteuse, est donc une source bruitée et parfois

peu informative sur la source du bruit.

Au fil de cet état de l'art, il apparaît donc qu'aucune source ne permet de réaliser une localisation de manière sûre et à peu de frais. Il est donc important 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 [5] ou les modèles de Markov [10]. Cependant, l'information temporelle dans ces modèles est extrêmement pauvre et rarement prise en considération. Des approches à base de règles ont été proposées [6,15] mais l'incertitude sur les données n'est pas considérée. Dans Hong et coll. [8], la théorie de Dempster-Shafer est appliquée pour fusionner l'information provenant de plusieurs capteurs afin de combiner des preuves de manière hiérarchique pour reconnaître deux activités culinaires. Cependant, l'information temporelle n'est, encore une fois, pas prise en considération. Duong et coll. [6] prennent en compte l'aspect temporel à travers d'un modèle semi-Markov caché hiérarchique qui fait la reconnaissance des activités à partir d'une séquence d'événements en modélisant aussi le temps passé dans chaque état (qui représente une sous activité) ; pourtant, dans notre cas, il est difficile de prédire la localisation présente en évaluant une séquence de localisations passées parce que le grand nombre de chemins de déplacements possibles augmenterait l'incertitude. Récemment, Niessen et coll. [12] ont proposés d'appliquer les réseaux dynamiques à la reconnaissance d'événements sonores. Dans cette approche, les événements sonores activent des hypothèses (type d'événements) qui mettent à jour des contextes qui ont eux même été inférés lors des événements précédents (p.ex. concert, hall de gare, etc.). La méthode n'impose pas de structure car elle est purement guidée par les données (c.-à-d. aucun ordonnancement entre les événements) mais la notion temporelle est explicitement prise en compte par une constante de temps qui permet de diminuer le poids d'un événement à mesure de son ancienneté. Étant donné la flexibilité apportée par cette approche, nous avons choisi d'adapter celle-ci à la localisation d'une personne dans un appartement à partir d'informations multi-sources.

3. Méthode

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. Après une courte introduction aux réseaux dynamiques et à la propagation d'activation, nous décrivons la méthode que nous avons adaptée pour travailler avec plusieurs sources temporelles.

3.1. Réseaux dynamiques et propagation d'activation

Un réseau dynamique [4] est un graphe où les nœuds représentent des concepts et où les arcs, généralement pondérés et orientés, représentent des relations entre les concepts. Un réseau dynamique est donc très proche d'un réseau sémantique mais il est plus général. Un réseau est dynamique dans le sens où il évolue en fonction d'événements qui "activent" des nœuds. Un nœud activé va propager son activation aux nœuds voisins. L'activation d'entrée A_{in}^j d'un nœud j se calcule généralement par la formule $A_{in}^j = \sum_{i \neq j} A_{out}^i \cdot w_{i,j}$ où $w_{i,j}$ est le poids (c.-à-d. le degré d'association) entre les nœuds j et i . L'activation de sortie A_{out}^j d'un nœud j peut être calculée à partir d'une fonction d'activation f avec $A_{out}^j = f(A_{in}^j)$. f peut par exemple, être une fonction de seuillage ou une fonction sigmoïde. Dans ce cas, la propagation d'activation est équivalente à la propagation dans les réseaux de neurones *feed-forward* classiques. La principale différence vient du fait qu'il n'y a pas véritablement de notion stricte de nœud d'entrée et de nœud de sortie car n'importe quel nœud peut potentiellement être activé par une source extérieure. Dans le cas d'un graphe non acyclique, l'activation peut être difficile à contrôler à cause de boucles dans le réseau. Dans ce cas l'activation peut être freinée par des contraintes sur le chemin emprunté par l'activation, ou en utilisant la sémantique des arcs ou des nœuds (p.ex. un nœud représentant un concept très général fortement connecté doit déclencher moins d'activation qu'un terme précis dans une application en RI). En RI, la propagation d'activation est souvent utilisée dans un réseau sémantique ou simplement associatif pour trouver des documents répondant à une requête. Dans Aswath et coll. [1], un réseau à deux niveaux où les niveaux de nœuds représentent

successivement des termes et des produits (produits indexés par les termes) est utilisé pour augmenter le nombre de produits associés à une requête. Dans la première étape de l'approche, une *hot phrase* est associée à des produits qui activent des termes qui à leur tour activent d'autres produits. Le nombre de produits associé à la *hot phrase* devient ainsi plus important que dans le cas d'une association simple *hot phrase*-produits (une activation sur les synonymes de la *hot phrase* est aussi effectuée). Les documents trouvés sont ensuite classés par SVM (Séparateurs à Vaste Marge) pour être insérés dans la liste de résultats. La propagation d'activation dans un réseau dynamique offre un cadre flexible et intuitif pour représenter des associations entre concepts, qui devient particulièrement intéressant lorsque plusieurs sources d'association sont considérées. Cependant, à notre connaissance, peu d'approches se sont intéressées aux cas de sources temporelles dont l'activation décroît avec le temps [12]. Nous présentons dans la section 3.2 une approche pour prendre en compte plusieurs sources temporelles pour localiser une personne dans un appartement.

3.2. Réseaux dynamiques temporels pour la fusion multisource

Récemment, Niessen et coll. [12] ont présenté une approche basée sur les réseaux dynamiques permettant de désambigüiser la reconnaissance d'événements sonores. Il s'agit d'un réseau à deux niveaux. Le niveau zéro est constitué des événements sonores, le niveau 1 représente les hypothèses liées à un événement (p.ex. : rebond de ballon ou claquement de mains), et le niveau deux représente le contexte de l'événement (p.ex. match de basket, concert, spectacle). Chaque événement active des hypothèses selon l'événement et les contextes auxquels les hypothèses sont liées. Les hypothèses vont ensuite activer les contextes en les mettant à jour. Ainsi, il y a une relation bidirectionnelle entre les contextes et les hypothèses. Par exemple, si plusieurs sons reconnus sont liés à un concert, les prochains sons reconnus auront des hypothèses liées à un concert beaucoup plus probables. La dimension temporelle est prise en compte par une constante de temps τ qui permet de fixer la décroissance d'une activation. Chaque mise à jour d'une activation entre t et $t + \Delta t$ est pondérée par $e^{-\frac{\Delta t}{\tau}}$. Ainsi, si aucun son n'est émis pendant $5 \cdot \tau$ secondes, l'activation des contextes pourra être considérée comme nulle.

Nous avons adapté cette approche pour localiser une personne à partir de différentes sources d'événements. 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 événement, et le deuxième niveau représente les contextes dont le poids d'activation indique la localisation la plus probable sachant les événements précédents. La figure 1 donne un exemple d'activation pour ce réseau. À l'instant $t[n-1]$, l'événement $e(t[n-1])$ apparaît et génère 3 hypothèses : $h1(t[n-1])$ avec un poids de 0.1 vers le contexte C_A , $h2(t[n-1])$ avec un poids de 0.6 vers le contexte C_B et $h3(t[n-1])$ avec un poids de 0.3 vers le contexte C_C . S'il n'y a pas eu d'événement antérieurement, c'est la localisation C_B qui sera la plus certaine. À l'instant $t[n]$, les poids précédents seront pondérés par $e^{-\frac{t[n]-t[n-1]}{\tau}}$ auxquels vont s'ajouter les poids générés par les hypothèses correspondant à l'événement $e(t[n])$: $h1(t[n])$ avec un poids de 0.2 vers le contexte C_B et $h2(t[n])$ avec un poids de 0.8 vers le contexte C_D . L'introduction du temps permet d'estimer la certitude de trouver une personne dans une pièce en fonction de sa dernière localisation. La diminution de cette certitude est implémentée par la fonction d'oubli $e^{-\frac{\Delta t}{\tau}}$.

La méthode s'appliquera ensuite de la même manière à l'instant $t[n + 1]$ pour lequel le contexte C_D recevra la plus grande activation et sera sélectionné.

La principale différence avec l'approche de Niessen et coll. [12] est la prise en compte de l'activation des hypothèses pour identifier le contexte –ici la localisation la plus probable–, alors que dans leur approche, l'information des contextes est utilisée pour déterminer les meilleures hypothèses. De plus, nous travaillons dans un cas multi-source où les sources, n'ayant pas le même degré de précision, influent différemment sur la génération des hypothèses de localisation. Ainsi, un événement créé par des capteurs de faible précision va générer des hypothèses avec des poids similaires et augmenter l'incertitude (cas de microphones détectant simultanément des sons avec des RSB voisins) ou bien donner un poids trop important à une hypothèse fautive.

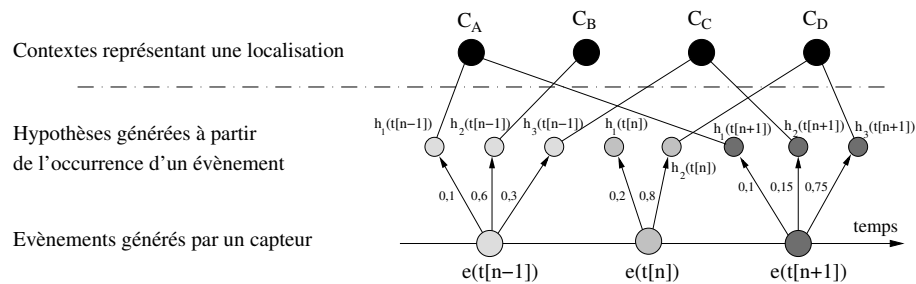


FIGURE 1 – Réseau dynamique.

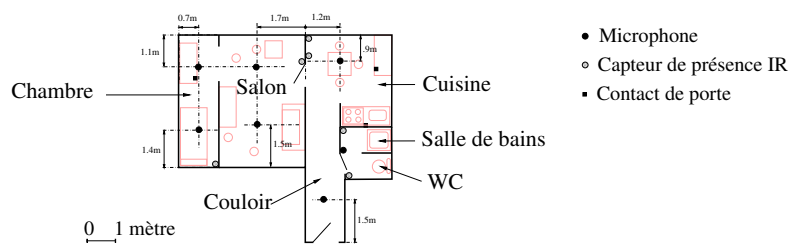


FIGURE 2 – Plan de l'appartement de la Faculté de Médecine de Grenoble et position des capteurs.

4. Expérimentation

La méthode de localisation par propagation d'activations a été testée sur des données réelles enregistrées dans l'habitat intelligent pour la santé (HIS) de la Faculté de Médecine de Grenoble dans le laboratoire TIMC-IMAG [11]. Cette section présente les données considérées dans cette étude, les choix de valeur de paramètres et les résultats obtenus.

4.1. Environnement perceptif et données utilisées

Les données expérimentales ont été acquises lors d'expériences [7] destinées à évaluer la reconnaissance automatique des Activités de la Vie Quotidienne (AVQ) d'une personne à son domicile dans le but, à long terme, de pouvoir détecter automatiquement une perte d'autonomie (c.-à-d. l'incapacité d'une personne à vivre sans assistance). La localisation de la personne joue un rôle primordial pour permettre une prise de décision adéquate aussi bien dans le cas d'une application purement domotique (c.-à-d. commande domotique) que d'une application orientée santé (c.-à-d. appel de détresse), c'est pourquoi nous avons pu utiliser ce corpus bien que l'objectif visé lors de son acquisition soit totalement différent de celui du projet Sweet-Home. La figure 2 décrit l'appartement intelligent. Il est composé de 6 pièces où sont disposés plusieurs capteurs :

- 7 microphones (Mic) placés dans le plafond (hauteur d'environ 3 mètres),
- 3 contacts sur des portes (CP) de meubles (placards dans la cuisine et commode dans la chambre),
- 6 détecteurs de présence infrarouge (PID) placés sur les murs à environ 2 mètres de hauteur.

Chacun des quinze volontaires a effectué au moins une fois 7 AVQ sur une période d'environ 2 heures : (1) dormir, (2) se distraire, (3) s'habiller, (4) se nourrir, (5) aller aux toilettes, (6) faire sa toilette, (7) communiquer avec l'extérieur. Les données ont ensuite été indexées (localisation et activité) grâce à des caméras placées à cet effet, une dans chaque pièce. Pour plus d'informations sur le corpus, le lecteur est invité à consulter l'article de Fleury et coll. [7].

Parmi les données recueillies par les capteurs (Mic, CP et PID), seuls les signaux audio ont subi un traitement. En effet, les données CP et PID délivrent une information booléenne déjà représentative d'événements. Les signaux audio ont été traités par le système AuditHis [13]. Succinctement, les événements audio sont détectés en temps réel par un algorithme à seuil adaptatif basé sur le

calcul du RSB (Rapport Signal sur Bruit). Les événements sont ensuite classés en parole ou son de la vie courante puis subissent d'autres traitements pour déterminer soit la parole prononcée soit le type de son (p.ex. chute d'objet). Les microphones étant omnidirectionnels, un son peut être enregistré en même temps par plusieurs microphones placés dans plusieurs pièces ; AuditHis repère alors ces événements simultanés. Le corpus acquis nous permet de disposer :

- du nom de la pièce et de l'instant d'occurrence des événements des capteurs CP et PID,
- de la description de chaque événement sonore (pièce, rapport signal sur bruit, son ou parole, durée, classe de son ou phrase prononcée),
- des annotations (c.-à-d. vérité terrain) de la localisation du sujet à chaque instant.

4.2. Calcul des poids de relations entre les couches du réseau dynamique

Le réseau dynamique étant constitué de deux niveaux bien définis, deux types de relations existent, la relation *hypothèse-contexte* et la relation *événement-hypothèse*.

La relation *hypothèse-contexte* est dans notre cas une relation biunivoque car une hypothèse de localisation n'est liée qu'à une et 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. Le poids de cette relation sera donc toujours de 1. Il faut aussi noter que cette relation est bidirectionnelle, un contexte influence une hypothèse et une hypothèse affirme ou infirme un contexte de localisation.

La relation *événement-hypothèse* est unidirectionnelle et de type de un vers plusieurs. Le poids et les hypothèses générés varient en fonction de la source de l'événement. Dans le cas des contacts de porte (CP) sur les meubles, un événement n'est lié qu'à une seule hypothèse. En effet, les CP étant placés à un endroit non ambigu et étant parfaitement fiables, un événement généré par un CP active l'hypothèse de la pièce dans laquelle il est placé avec un poids de 1. Les PID sont aussi placés à des endroits bien définis et sont généralement considérés comme fiables. Cependant, nos expériences ont montré un défaut de sensibilité (mais une précision de 100%) et quelques ambiguïtés de localisation dues à leur placement et leur cône de réception. C'est pourquoi, pour chaque événement de PID, l'hypothèse de pièce du capteur reçoit un poids de 0,95 et les 5 autres un poids de 0,01. Ce choix arbitraire permet d'éviter une mise à zéro brutale d'un contexte concurrent. Pour les événements causés par la détection d'un son, les poids attribués aux hypothèses sont calculés dynamiquement. Un événement sonore est généralement composé d'un ensemble d'événements sonores élémentaires détectés simultanément $E = \{e_j\}$ sur plusieurs microphones. Dans cet ensemble, seuls les sons catégorisés comme étant de la parole sont considérés. En effet, notre expérience nous a montré que les sons de la vie courante peuvent être générés par d'autres sources que l'habitant (bruit extérieur). Les hypothèses sont générées à partir des pièces d'occurrence des sons. Le poids des relations événement-hypothèse est calculé pour chaque hypothèse de pièce p_i en utilisant le RSB de chaque détection suivant la formule suivante :

$$P(p_i | E) = \frac{\sum_{e_j \in E \wedge e_j.\text{room} = p_i} \text{rsbl}(e_j)}{\sum_{e_j \in E} \text{rsbl}(e_j)} \quad (1)$$

avec $\text{rsbl}(e_j) = 10^{\frac{e_j.\text{rsb}}{10}}$, et où e_j est un événement sonore élémentaire, p_i est la i^e pièce et $e_j.\text{room}$ (resp $e_j.\text{rsb}$) est la pièce (resp. RSB) où se produit l'événement sonore. Ainsi, l'événement sonore avec le plus grand RSB génère l'hypothèse la plus probable. Le RSB suivant une échelle logarithmique, il doit être linéarisé avant le calcul du poids.

4.3. Résultats

Pour chaque enregistrement des sujets, les événements de CP, PID et Mic ont activé un réseau dynamique. Les traces d'activation de chaque contexte ont été conservées. Les performances de localisation ont été évaluées à l'échelle de la seconde. À chaque seconde, le contexte 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 $TE = \frac{\text{nb}(VP)}{\text{nb}(test)}$ où $\text{nb}(test)$ correspond à la durée de l'évaluation exprimée en secondes. La première ligne de la table 1 montre les résultats sur les 15 sujets pour chaque combinaison de capteurs. Nous pouvons noter que le fait d'ajouter l'information des contacts de porte améliore légèrement la précision par rapport à celle obtenue

Capteurs	PID	Mic	CP	Mic+CP	PID+CP	PID+Mic	PID+Mic+CP
Exactitude globale	95	45	39	56	96	89	90
Exactitude sujet 10	60	63	31	78	61	97	97

TABLE 1 – Exactitude avec plusieurs combinaisons de sources

par les seuls capteurs infrarouges. L'information sonore, quand à elle, dégrade les performances (90% au lieu de 96%). Une des raisons vient du fait que dans les enregistrements la classification en parole des événements sonores du système AuditHis comporte 25% d'erreur (p.ex confusion des harmoniques de la chasse d'eau avec de la parole). Néanmoins, l'information sonore est utile pour améliorer la localisation dans certains cas. En effet, il y a des situations où le changement de pièce n'est pas détecté par les PID alors que les paroles sont bien identifiées, ce qui vient compenser les erreurs de sensibilité des infrarouges. La deuxième ligne de la table 1 montre les résultats pour le sujet 10. Dans ce cas, on voit bien que chaque source prise isolément ne permet pas une bonne localisation mais que la combinaison des sources permet un net gain de performance (de 60% à 97%). Pour ce même participant, le niveau d'activation des contextes au fur et à mesure que les événements arrivent est représentés dans la figure 3.

5. Discussion et perspectives

Les résultats montrent que la fusion d'information par propagation d'activation multisource 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) ou des contacts de porte. L'utilisation de sources de localisation plus incertaines, telles que la reconnaissance de la parole, peut alors améliorer les performances dans certains cas. Cependant, le fort taux d'erreurs de cette source (25%) entraîne une dégradation des performances globales.

Plusieurs pistes d'amélioration de cette méthode vont être suivies. La première consiste à pondérer les événements. En effet pour l'instant, chaque événement a un poids de 1 repartit sur les hypothèses, cependant, une détection (événement) moins certaine qu'une autre devrait avoir un impact moins important. Une extension de la méthode prévoit donc d'inclure l'estimation de la certitude des événements dans le calcul des poids (p.ex. la sortie de la reconnaissance de la parole fournit une probabilité de reconnaissance). Une deuxième piste consiste à agir sur la constante de temps en fonction de la sémantique des événements et des lieux. Par exemple, une détection dans le couloir devrait avoir une constante de temps plus faible que celle des toilettes. Nous prévoyons d'extraire les constantes de temps potentielles à partir des données. Une troisième voie consiste à utiliser la topologie de l'appartement qui est pour l'instant complètement ignorée. Par exemple, un bruit de chasse d'eau, ne peut être que dans les toilettes, un occupant ne peut pas se retrouver de la chambre à la porte d'entrée sans passer par le salon, etc. Ce dernier point peut justement être modélisé par un réseau dynamique.

Sur le plus long terme, la méthode présentée peut être améliorée en utilisant la théorie de Dempster-

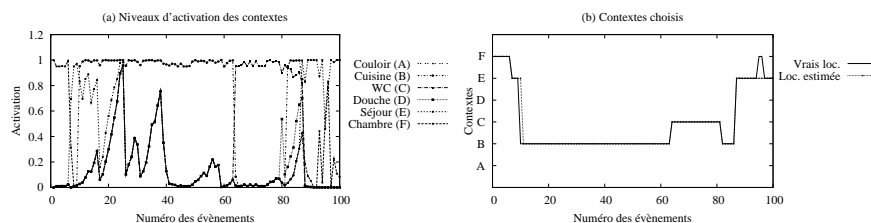


FIGURE 3 – Exemple montrant (a) l'évolution des niveaux d'activation des contextes et (b) les contextes choisis pour les 6 localisations possibles.

Shafer qui permettrait d'utiliser une fusion d'information qui déterminerait 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. Cette méthode a déjà été utilisée pour la reconnaissance des activités [8], pourtant l'aspect temporel n'a pas été pris en compte alors qu'il peut être modélisé par un réseau dynamique temporel. 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.

Bibliographie

1. Dipti ASWATH, James D'CUNHA, Syed Toufeeq AHMED et Hasan DAVULCU : Boosting item keyword search with spreading activation. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI'05)*, pages 704–707, 2005.
2. M. BERENQUER, M. GIORDANI, F. GIRAUD-BY et N. NOURY : Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line. In *HealthCom'08, 10th IEEE Int. Conf. on e-Health Networking, Applications & Service*, 2008.
3. X. BIAN, G. ABOWD et J. REHG : Using sound source localization in a home environment. In *Third International Conference of Pervasive Computing*, pages 19–36, 2005.
4. F. CRESTANI : Application of spreading activation techniques in information retrieval. *Artificial Intelligence Review*, 11:453–482, 1997.
5. S. DALAL, M. ALWAN, R. SEIFRAFI, S. KELL et D. BROWN : A rule-based approach to the analysis of elders' activity data : detection of health and possible emergency conditions. In *AAAI 2005 fall symposium, workshop on caring machines : AI in eldercare*, 2005.
6. T. DUONG, D. PHUNG, H. BUI et S. VENKATESH : Efficient duration and hierarchical modeling for human activity recognition. *Artificial Intelligence*, 173(7–8):830–856, May 2009.
7. A. FLEURY, M. VACHER, F. PORTET, P. CHAHUARA et N. NOURY : A multimodal corpus recorded in a health smart home. In *International Workshop Series on Multimodal Corpora, Tools and Resources, LREC*, pages 99 – 105, Malta, May 18–21 2010.
8. X. HONG, C. NUGENT et M. MULVENNA : Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5:236–252, 2009.
9. S. S. INTILLE : Designing a home of the future. *IEEE Pervasive Computing*, 1(2):76–82, 2002.
10. B. KRÖSE, T. van KASTEREN, C. GIBSON et T. van den DOOL : Care : Context awareness in residences for elderly. In *The 6th International Conference of the International Society of Gerontology*, page 6p., Pisa, Italy, June 4-7, 2008.
11. G. LE BELLEGO, N. NOURY, G. VIRONE, M. MOUSSEAU et J. DEMONGEOT : A model for the measurement of patient activity in a hospital suite. *IEEE Transactions on Information Technologies in Biomedicine*, 10(1):92 – 99, 2006.
12. M. NIESSEN, L. van MAANEN et T. ANDRINGA : Disambiguating sounds through context. In *IEEE International Conference on Semantic Computing*, pages 88 – 95, 4-7 Aug. 2008.
13. M. VACHER, A. FLEURY, F. PORTET, J.-F. SERIGNAT et N. NOURY : *New Developments in Biomedical Engineering*, chapitre Complete Sound and Speech Recognition System for Health Smart Homes : Application to the Recognition of Activities of Daily Living, pages 645 – 673. Intech Book, Feb. 2010. ISBN : 978-953-7619-57-2.
14. M. VACHER, F. PORTET, A. FLEURY et N. NOURY : Challenges in the processing of audio channels for ambient assisted living. In *12th International Conference on E-Health Networking, Applications and Services*, Lyon, France, 1-3 Jul. 2010.
15. N. ZOUBA, F. BREMOND, M. THONNAT, A. ANFONSO, E. PASCUAL, P. MALLEA, V. MAILLAND et O. GUERIN : A computer system to monitor older adults at home : preliminary results. *International Journal of Gerontechnology*, 8(3), Jul. 2009.

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

Les auteurs remercient particulièrement Anthony Fleury et le laboratoire TIMC-IMAG pour la collecte des données dans un environnement réel ainsi que pour l'aide apportée lors de leur indexation.
