Reconnaissance d'ordres domotiques en conditions bruitées pour l'assistance à domicile (Recognition of Voice Commands by Multisource ASR and Noise Cancellation in a Smart Home Environment) [in French]

Benjamin Lecouteux, Michel Vacher, François Portet

▶ **To cite this version:**

Benjamin Lecouteux, Michel Vacher, François Portet. Reconnaissance d'ordres domotiques en conditions bruitées pour l'assistance à domicile (Recognition of Voice Commands by Multisource ASR and Noise Cancellation in a Smart Home Environment) [in French]. JEP-TALN-RECITAL 2012, Atelier ILADI 2012: Interactions Langagières pour personnes Agées Dans les habitats Intelligents, Jun 2012, Grenoble, France. pp.31–39. hal-00953515

HAL Id: hal-00953515
https://inria.hal.science/hal-00953515

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.

Reconnaissance d'ordres domotiques en conditions bruitées pour l'assistance à domicile

Benjamin Lecouteux, Michel Vacher, François Portet
Laboratoire Informatique de Grenoble, équipe GETALP
prénom.nom@imag.fr

RÉSUMÉ

Dans cet article, nous présentons un système de reconnaissance automatique de la parole dédié à la reconnaissance d'ordres domotiques dans le cadre d'un habitat intelligent en conditions réelles et bruitées. Ce système utilise un étage d'annulation de bruit qui est à l'état de l'art. L'évaluation du système proposé est effectuée sur des données audio acquises dans un habitat intelligent où des microphones ont été placés proche des sources de bruit (radio, musique...) ainsi que dans les plafonds des différentes pièces. Ce corpus audio, a été enregistré avec 23 locuteurs prononçant des phrases banales, de détresse ou de type domotique. Les techniques de décodage utilisant des connaissances *a priori* donnent des résultats en conditions bruitées comparables à ceux obtenus en conditions normales, ce qui permet de les envisager en conditions réelles. Cependant l'étage d'annulation de bruit semble beaucoup plus efficace pour annuler les bruits issus de la radio (parole) que ceux de type musicaux.

ABSTRACT

In this paper, we present a multisource ASR system to detect home automation orders in various everyday listening conditions in a realistic home. The system is based on a state of the art noise cancellation stage that feeds recently introduced ASR techniques. The evaluation was conducted on a realistic noisy dataset acquired in a smart home where a microphone was placed near the noise source and several other microphones were set in the ceiling of the different rooms. This distant speech French corpus was recorded with 23 speakers uttering colloquial or distress sentences as well as home automation orders. Techniques acting at the decoding stage and using a priori knowledge gave the best results in noisy condition compared to the baseline reaching good enough performance for a real usage. If broadcast news is easily handled by the noise canceller, improvements still need to be made when music is used as background noise.

MOTS-CLÉS : Domotique, habitat intelligent, parole distante, SRAP multisource, détection de mots clefs.

KEYWORDS: Home automation, smart home, distant speech, multisource ASRs, keyword detection.

JEP-TALN-RECITAL 2012, Atelier ILADI 2012: Interactions Langagières pour personnes Agées Dans les habitats Intelligents, pages 31–39,
Grenoble, 4 au 8 juin 2012. ©2012 ATALA & AFCP

1 Introduction

Les changements démographiques et le vieillissement dans les pays développés invitent à la réflexion sur la prise en charge de notre population. Par ailleurs, l'évolution technologique permet d'envisager de nombreuses possibilités en vue d'améliorer la qualité de vie et de soutenir les personnes âgées, afin de vivre dans leur propre maison avec un maximum d'autonomie. Une solution pour apporter cette assistance au quotidien est le développement des maisons intelligentes qui sont équipées de capteurs, d'actionneurs, d'automates et de logiciels centralisés, contrôlant une partie des appareils ménagers. Diverses méthodes d'interaction sont en cours d'élaboration dans ce cadre, mais l'une des plus prometteuses est la reconnaissance automatique de la parole (RAP). En effet, les interfaces vocales sont beaucoup plus adaptées pour les personnes ayant des difficultés à se déplacer ou à voir. La commande vocale est aussi particulièrement adaptée aux situations de détresse où une personne ne peut plus se déplacer après une chute : elle a la possibilité de demander de l'aide (Hamill *et al.*, 2009) . En outre, étant donné la complexité croissante des appareils électroménagers, une interface vocale semble plus naturelle qu'une interface tactile (Vovos *et al.*, 2005).

Alors que l'interaction vocale est une caractéristique souhaitable pour une maison intelligente, de nombreux verrous doivent être soulevés pour transférer cette technologie du laboratoire au domicile. L'un des enjeux majeurs est la mauvaise performance de la RAP en environnement bruyant (Vacher *et al.*, 2011). En effet, dans des conditions réalistes, la performance des systèmes de RAP (SRAP) diminue significativement dès que le microphone s'éloigne du locuteur. Cette détérioration est due à une grande variété d'effets, tels que la réverbération et la présence de bruits de fond (télévision, radio etc. (Wölfel et McDonough, 2009)). Tandis que les aspects linguistiques, de dialogue ou d'interface ont été étudiés en fonction de l'âge, (Vovos *et al.*, 2005; Hamill *et al.*, 2009; Vipperla *et al.*, 2009), la RAP dans les habitats intelligents n'a reçu d'attention que très récemment au sein de la communauté du traitement de la parole (Barker *et al.*, 2011).

Dans cet article nous présentons un système reconnaissant les ordres domotiques dans un habitat intelligent en conditions bruitées. Ce travail fait parti du projet Sweethome [1] introduit dans la section 2. L'approche proposée est basée sur un étage d'annulation de bruit et un SRAP multi-source qui utilise des connaissances *a priori* pour améliorer la reconnaissance des ordres domotiques. Cette plateforme est présentée dans la section 3. Les expériences et résultats sont présentés dans la section 4, puis nous concluons.

2 Contexte de l'étude et des données utilisées pour l'évaluation

Cette étude a été effectuée dans le contexte du projet Sweethome (`http://sweet-home.imag.fr`) et s'articule autour de la conception d'un habitat intelligent doté de RAP ; permettant une interaction (ordres domotiques) et une détection des situations de détresse. Grâce à cet habitat, les personnes seront capables de piloter de n'importe quel endroit de la maison leur environnement.

1. Cette étude a été financée par l'Agence Nationale de la Recherche dans le cadre du projet Sweet-Home (ANR-2009-VERS-011). Nous remercions particulièrement les différentes personnes qui ont accepté de participer aux enregistrements.

```
Commande          = clef commande_départ objet |
                    clef commande_arrêt [objet] |
                    clef commande_aide
clef               = "Nestor" | "maison"
commande_arrêt      = "stop" | "arrête"
commande_départ   = "ouvre" | "ferme" | "baisse" | "éteins" | "monte" |
                    "allume" | "descend" | "appelle"
commande_aide = "au secours" | "à l'aide"
objet            = [determiner] ( appareil | personne | organisation)
déterminant       = "mon" | "ma" | "l'" | "le" | "la" | "les" | "un" | "des" |
                    "du"
appareil           = "lumière" | "store" | "rideau" | "télé" | "télévision" |
                     "radio"
personne           = "fille" | "fils" | "femme" | "mari" | "infirmière" |
                    "médecin" | "docteur"
organisation      = "samu" | "secours" | "pompiers" | "supérette" | "supermarché"
```

FIGURE 1 – Grammaire générant les ordresdomotiques

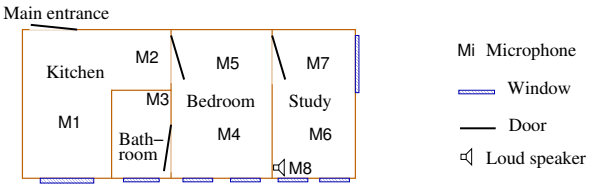

FIGURE 2 – Position des microphones dans l'appartement domus

Dans cette étude les ordres domotiques ont été définis en utilisant une grammaire (Figure 1). Nos précédentes études ont montré que les utilisateurs préfèrent des phrases courtes à des phrases plus longues (et naturelles) pour piloter leur environnement (Portet *et al.*, ress). Chaque ordre est classifié dans une des catégories suivantes : commande initiale, commande d'arrêt, appel de détresse. A l'exception des appels de détresse, toutes les commandes commencent par un mot clef permettant de lever l'ambiguïté sur les phrases prononcées. Dans nos expériences, le mot clef est **Nestor** (qui représente l'entité intelligente de l'habitat) :

L'environnement dans lequel la RAP est effectuée est montré dans la figure 2. C'est un appartement de plein pied de 30 m² mis en place par l'équipe MULTICOM du LIG. Cet appartement comprend une salle de bain, une cuisine, une chambre et un bureau. Toutes les pièces sont équipées de capteurs et d'interrupteurs reliés à un réseau central. Par ailleurs 7 microphones ont été installés dans les plafonds et enregistrent en temps réel le flux grâce à un pc muni d'une carte audio 8 canaux (Vacher *et al.*, 2011). Pour les expériences proposées dans cet article, un huitième microphone a été placé face au haut parleur qui sera source de bruit (radio, musique). Cette étude se rapproche ainsi du cadre réel où l'utilisateur est distant des micros, et en présence d'un bruit environnant issu de sa chaîne Hi-Fi.

A notre connaissance aucune base d'ordres domotiques en Français en conditions bruitées n'a déjà été créée. Nous avons conduit une phase d'enregistrements de phrases mélangeant à la fois des phrases neutres, des ordres domotiques et des appels de détresse. Afin d'être dans des conditions aussi réalistes que possible, deux types de bruits de fond ont été générés lorsque l'utilisateur parlait : un journal radiophonique et de la musique classique. Cette source de bruit a été générée par les deux haut-parleurs de la chaîne Hi-Fi de l'appartement ; l'un des hauts parleurs est enregistré dans le cadre de l'expérience. Il est à noter que ces conditions n'ont rien à voir avec des mélanges de sources artificiellement mélangées.

Bruit de fond	Bureau	Chambre	Salle de bains	Cuisine
Rien	30	30	30	30
Musique (Radio)	30	30	-	-
Parole (journal à la radio)	30	30	-	-

TABLE 1 – Nombre de phrases en fonction de la pièce (Phase 2)

Le protocole se décompose en deux phases. Au cours de la **Phase 1**, les participants sont placés dans le bureau, ferment la porte et lisent un texte de 285 mots. Ce texte sera utilisé par la suite dans l'objectif d'adapter les modèles acoustiques du SRAP. Dans la **Phase 2**, les participants doivent prononcer 30 phrases par pièce dans différentes conditions : sans bruit, avec radio ou avec musique. La Table 1 résume les conditions et les lieux d'enregistrement. Chaque séquence de 30 phrases a été composée par une sélection aléatoire de 21 ordres domotiques (9 sans mots clefs initiaux), 2 appels de détresse ("À l'aide", "Appelez un docteur"), et 7 phrases neutres ("Bonjour" , "J'ai bien dormi"). Les participants ne prononcent pas les même phrases. La radio et la musique sont par contre toujours identiques, mais démarrés au bout d'un laps de temps aléatoire pour chaque participant.

23 personnes (dont 9 femmes) ont participé aux expériences. L'âge moyen des participants est de 35 ans (19 à 64 ans). Aucune instruction n'a été donnée aux participants sur la manière dont ils devaient prononcer leurs phrases. La distance entre le locuteur et le microphone le plus proche est généralement de 2 mètres. La durée totale des expériences est quant à elle de 5 heures.

A la fin de l'expérience, la phase 1 représente un total de 36 mn pour 351 phrases, et la phase 2 représente un total de 2h30 pour 5520 phrases dont 38 minutes en condition radio et 37 mn en condition musique (2760 phrases en conditions bruitées). Chaque phrase a été manuellement annotée sur le canal de meilleur rapport signal bruit.

En conditions normales, 1076 ordres domotiques sont prononcés et 348 appels de détresse sont effectués contre respectivement 489 et 192 pour la radio en fond ; 412 et 205 avec la musique en fond.

3 Approches proposées pour un SRAP robuste

Afin de détecter les ordres domotiques dans le contexte de Sweethome, nous proposons une approche à trois étages. Le premier détecte les activités de parole dans le flux audio, le second extrait les meilleures hypothèses en utilisant un SRAP et le dernier identifie l'objet de ce qui est prononcé. Le premier étage est décrit dans (Vacher *et al.*, 2011).

Pour résoudre les problèmes du contexte Sweethome (bruit, parole distante) et pour bénéficier de ce dernier (microphones multiples), nous proposons de tester l'impact de techniques nouvelles ou état de l'art permettant de fusionner les flux à trois niveaux différents du processus de décodage : acoustique, décodage et sorties des SRAP (**?**). Malgré de bons résultats les méthodes proposées n'incluaient pas de traitement du bruit de fond. Cette section se concentre donc sur les techniques implémentées pour l'annulation de bruit dans le cas de sources connues ; en réutilisant les méthodes proposées dans nos précédents travaux.

3.1 Suppression de bruits dont la source est connue

L'écoute de la radio ou de la TV sont des activités fréquentes dans la vie quotidienne. Mais ces appareils peuvent impacter un SRAP à deux niveaux : les sons émis par la personne dans l'appartement sont altérés par le bruit de fond ; quant aux informations issues de la radio ou télévision elles peuvent être décodées à tort par le SRAP. Dans cette configuration, nous proposons d'annuler le bruit de fond en utilisant des méthodes d'annulation d'écho (AEC).

Le processus AEC est décrit dans la Figure 3. Le son émis par une source de bruit $x(n)$ (dans notre cas le haut parleur de la chaîne hi-fi) est altéré par l'acoustique de la pièce. Le bruit résultant de cette altération $y_b(n)$ peut être exprimé sous la forme d'un produit de convolution dans le domaine temporel $y_b(n) = h(n) * x(n)$, h étant la réponse impulsionnelle de la pièce et n le temps discrétisé. Ce bruit est alors mélangé avec le signal qui nous intéresse $e(n)$ émis dans la pièce (la parole). Le signal enregistré par le microphone devient alors : $y(n) = e(n) + h(n) * x(n)$. Afin d'annuler le bruit, un filtre adaptatif (Vacher *et al.*, 2009) estime la réponse impulsionnelle de la pièce $\hat{h}(n)$ et utilisé pour restituer une estimation du signal original $e(n)$ suivant l'équation : $v(n) = e(n) + y_b(n) - \hat{y}(n) = e(n) + h(n) * x(n) - \hat{h}(n) * x(n)$

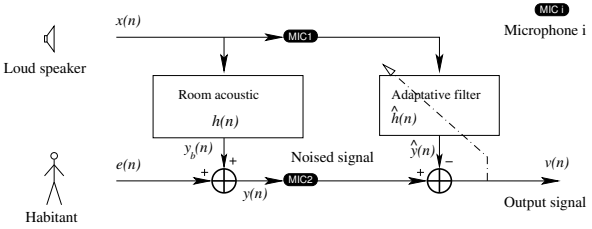

FIGURE 3 – Principe de l'annulation d'écho appliquée à l'annulation de bruit

Le filtre de bruit est adapté en utilisant la valeur résiduelle $v(n)$ conduisant à un système d'adaptation où v est la rétroaction. Au début du processus, un certain temps est nécessaire pour estimer $\hat{h}$ et minimiser l'erreur : c'est le temps de convergence. Si aucun signal $x(n)$ n'est présent, l'adaptation du filtre tend à diverger en raison des paramètres y_b et $\hat{y}$ qui deviennent nuls. Ainsi l'AEC doit être appliqué uniquement lorsque le bruit de fond est présent. Il est aussi nécessaire que e reste proche de zéro sinon le signal utile (ici l'ordre domotique) est considéré comme un bruit additif et l'adaptation devient instable. Pour éviter ce problème, l'annulation d'écho robuste nécessite un réglage. Nous avons utilisé la bibliothèque SPEEX dont l'étape d'AEC est basée sur l'algorithme de (Valin et Collings, 2007). La suppression du bruit a été effectuée séparément sur les 7 canaux.

3.2 Détection des ordres domotiques

Une fois le bruit filtré les canaux sont décodés par le SRAP Speeral (Linarès *et al.*, 2007) du LIA (Laboratoire Informatique d'Avignon). Étant donné le champ d'application de Sweethome et ses contraintes temps-réel, Speeral est configuré en 1xRT (temps réel). Au niveau du décodage une nouvelle version du DDA (Driven Decoding Algorithm) a été utilisé dans Speeral. DDA aligne et

corrige des transcriptions a priori en utilisant un SRAP (Lecouteux *et al.*, 2008). Cet algorithme améliore les performances du décodage en s'appuyant sur la disponibilité des informations *a priori* (prompts, scénarios etc.)

Dans le contexte d'un habitat intelligent, nous n'avons pas la connaissance de ce que les locuteurs vont prononcer. Cependant le système connaît la composition des ordres domotiques et possède plusieurs micros. Il est alors possible d'utiliser DDA pour combiner efficacement plusieurs microphones en guidant un microphone par le décodage issu d'un autre. Ainsi à chaque nouvelle hypothèse explorée par le SRAP, cette dernière est alignée avec le décodage d'un autre micro. Un score d'alignement α est alors calculé pour biaiser le modèle de langage (Lecouteux *et al.*, 2008) :$\tilde{P}(w_i|w_{i-1}, w_{i-2}) = P^{1-\alpha}(w_i|w_{i-1}, w_{i-2})$ où $\tilde{P}(w_i|w_{i-1}, w_{i-2})$ est la nouvelle probabilité du trigramme et $P(w_i|w_{i-1}, w_{i-2})$ est sa probabilité initiale.

La stratégie proposée est dynamique et utilise pour chaque phrase à décoder deux canaux. Cette approche a été améliorée pour prendre en compte des connaissances *a priori* sur les phrases attendues. Le SRAP est alors guidé par les ordres vocaux reconnus durant la première passe : les segments de parole du premier canal sont projetés dans les 3 meilleurs ordres domotiques (la mesure étant une distance d'édition) possibles qui vont alors guider la seconde passe, comme présenté dans (**?**). Cette approche présente plusieurs avantages : la vitesse du second SRAP est augmentée par la présence de la transcription approchée (seulement 0.1x le temps réel), DDA permet de combiner efficacement l'information des deux canaux, enfin DDA permet d'introduire une grammaire flexible au sein du décodage.

4 Expériences et résultats

Dans toutes les expériences, le corpus de la phase 1 a été utilisé pour le développement et l'apprentissage. La Phase 2 a été utilisée pour l'évaluation. Cette section présente les spécifications du SRAP et les résultats expérimentaux pour les différentes approches.

4.1 Les modèles du SRAP

Les modèles acoustiques ont été entraînés sur 80 heures de parole annotées. Par ailleurs, les modèles sont adaptés pour chacun des locuteurs via une MLLR sur les données de la phase 1. Pour le décodage, un modèle de langage 3-gramme est utilisé avec un lexique de 10K mots. Ce modèle est issu de la combinaison d'un modèle générique (10%) avec un modèle spécialisé appris sur la grammaire d'ordres domotiques (90%). Le modèle générique (1000M mots) a été appris sur des données issues du journal *Le Monde* et du corpus *Gigaword*. La combinaison biaise le comportement du SRAP tout en permettant de décoder des phrases "hors domaine". Un modèle probabiliste a été préféré à une grammaire déterministe afin d'avoir plus de flexibilité dans des situations parfois éloignées d'un cadre expérimental parfait.

4.2 Résultats

Les résultats des différentes approches sont présentés dans la Table 2. Dans cette étude, nous nous intéressons à la détection des ordres domotiques ou de détresse. Nous avons ainsi trois

Méthode	rappel ordres domotiques	rappel détresse	précision détresse	rappel phrases neutres
Sans bruit	62.1(±16.9)	84.2(±29.2)	88.8(±18.5)	97.5(±5.2)
Sans bruit+DDA	**92.7(±10.1)**	**87.2(±27.3)**	**89.0(±18.1)**	**97.9(±5.2)**
radio	29.3 (±23.5)	74.3(±22)	73.7(±19.8)	94.5(±4.8)
radio + DDA	**57.2(±30.8)**	**75.2(±22.1)**	**74.7(±19.9)**	**94.6(±5)**
radio+débruitage	42.6 (±21.1)	79.4(±19.4)	87.5(±17.6)	97.2(±3.8)
radio+DDA+débruitage	**83.5(±16.1)**	**81.2(±19.1)**	**88.0(±18.2)**	**97.8(±3.9)**
Music	59.0(±21)	81.6(±27.6)	87.3(±16.2)	96.8(±4)
Musique+DDA	**90.6(±15)**	**82.5(±26.1)**	**87.6(±16.1)**	**97.1(±3.9)**
Musique+débruitage	46.9(±23.8)	64.5(±36.4)	79.7(±27.1)	94.8(±5.3)
Musique+DDA+débruitage	**79.2(±16.5)**	**66.5(±34.3)**	**80.7(±27.2)**	**95.1(±4.8)**

TABLE 2 – Détection d'ordres domotiques et de phrases de détresse dans trois configurations : musique, radio et sans bruit

classes : ordre domotique, appels de détresse et phrases neutres. La reconnaissance est évaluée en utilisant le triplet rappel/précision/F-mesure :

$$rappel = \frac{nombre\ de\ phrases\ correctement\ attribuees\ a\ une\ classe}{nombre\ de\ phrases\ de\ la\ classe}$$
$$precision = \frac{nombre\ de\ phrases\ correctement\ attribuees\ a\ une\ classe}{nombre\ de\ phrases\ attribuees\ a\ la\ classe}$$
$$F - mesure = \frac{2.rappel.precision}{precision + rappel}$$

Au cours de la détection un ordre domotique ou de détresse est reconnu comme tel uniquement s'il correspond parfaitement à la grammaire. Dans tous les autres cas, il est classifié comme neutre. Pour chaque approche, les résultats présentés sont moyennés pour les 23 locuteurs. Pour comparaison, des expériences en conditions non bruitées sont présentées.

Les expériences sans bruits environnants montrent un rappel des ordres domotiques de 62% et 84% pour les phrases de détresse. La meilleure détection des phrases de détresse s'explique par le petit nombre de possibilités d'expressions de détresses comparé au nombre possible d'ordres domotiques (400). Quand le DDA est utilisé, la détection des ordres domotiques monte à 92.7% et la détection des ordres de détresse augmente très légèrement. L'impact est plus grand pour les ordres domotiques car DDA introduit directement une grammaire dans le SRAP et dans le cas des phrases de détresse, il agit comme une simple combinaison entre deux canaux.

Dans le cas d'un environnement bruité de type radio, le rappel des ordres domotiques est de 29.3% tandis que le rappel des ordres de détresse baisse à 74.3%. L'introduction du DDA améliore la détection des ordres domotiques de 57.2% relatifs mais n'a pas d'effet sur les appels de détresse. En utilisant le système d'annulation de bruit, la détection des ordres domotiques et de détresse sont très largement améliorés. Finalement, la meilleure configuration est obtenue en combinant les deux approches : la détection d'ordres domotiques atteint ainsi 83.5% et la détection des phrases de détresse atteint 81.2%.

Dans le contexte musical les résultats sont surprenants sur deux aspects. La musique ne semble pas impacter outre mesure les résultats du SRAP. Lorsque l'annulation de bruit est utilisée les performances se dégradent sensiblement. Le seul locuteur pour lequel nous avons observé une amélioration était dans un environnement où la musique avait été réglée extrêmement fortement (par erreur). Quant au DDA, il améliore les résultats dans toutes les conjonctions excepté dans le

cas d'une combinaison avec l'annulation de bruit.

Dans toutes les configurations, la précision de la détection des ordres domotiques est améliorée en utilisant DDA : le taux de reconnaissance est supérieur à 80%. L'approche basée sur l'annulation de bruit montre des améliorations significatives dans le cadre d'émissions radio mais n'est pas adaptée à d'autres types de bruits. Ce dernier point n'est pas surprenant pour une méthode avant tout dédiée à traiter des données audio de type voix.

5 Conclusion

Cet article décrit une approche pour la détection d'ordres domotiques dans un habitat intelligent où les informations audio sont capturées par des microphones distants du locuteur. Notre approche permet d'être efficace dans des conditions bruitées. L'équipement de cet habitat permet de connaître les médias allumés (télévision, radio etc.) : nos expériences prennent en compte cet aspect en utilisant à la source les équipements générant du bruit (via des micros). L'approche proposée s'effectue donc à deux niveaux : acoustique (annulation de bruit parasite) et décodage par introduction à la volée d'une grammaire d'ordres domotiques (DDA).

D'excellents résultats ont été obtenus. En utilisant le DDA, plus de 80% des ordres domotiques ou phrases de détresses sont détectés à la fois en conditions non bruitées et bruitées. Le DDA a montré les améliorations les plus importantes, tandis que la méthode d'annulation de bruit ne fonctionne que dans le cas d'émissions radiophoniques et montre des dégradations en cas d'utilisation avec de la musique. Ce résultat peut en partie s'expliquer par le fait que le système d'annulation de bruit est à la base prévu pour de l'annulation de voix et s'avère moins adapté à la musique. En contrepartie le SRAP et le DDA sont beaucoup moins perturbés par la musique que par la radio : le spectre musical est sans doute filtré par les modèles acoustiques. Nos prochains travaux s'articuleront autour du type de bruits (voix, musique...) afin de sélectionner à la volée la meilleure approche de décodage.

Références

Barker, J., Christensen, H., Ma, N., Green, P. et Vincent, E. (2011). The PASCAL 'CHiME' Speech Separation and Recognition Challenge. *In InterSpeech 2011*. (to appear).

Hamill, M., Young, V., Boger, J. et Mihailidis, A. (2009). Development of an automated speech recognition interface for personal emergency response systems. *Journal of NeuroEngineering and Rehabilitation*, 6.

Lecouteux, B., Linarès, G., Estève, Y. et Gravier, G. (2008). Generalized driven decoding for speech recognition system combination. *In Proc. IEEE ICASSP 2008*, pages 1549–1552.

Linarès, G., Nocéra, P., Massonié, D. et Matrouf, D. (2007). The LIA speech recognition system : from 10xRT to 1xRT. *In Proc. TSD'07*, pages 302–308.

Portet, F., Vacher, M., Golanski, C., Roux, C. et Meillon, B. (in press). Design and evaluation of a smart home voice interface for the elderly — Acceptability and objection aspects. *Personal and Ubiquitous Computing*.

VACHER, M., FLEURY, A., GUIRAND, N., SERIGNAT, J.-f. et NOURY, N. (2009). Speech recognition in a smart home : some experiments for telemonitoring. *In* CORNELIU BURILEANU, H.-N. T., éditeur : *From Speech Processing to Spoken Language Technology*, pages 171–179, Constanta (Romania). Publishing House of the Romanian Academy.

VACHER, M., PORTET, F., FLEURY, A. et 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.

VALIN, J.-M. et COLLINGS, I. B. (2007). A new robust frequency domain echo canceller with closed-loop learning rate adaptation. *In Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing, ICASSP'07*, volume 1, page 93–96, Honolulu, Hawaii, USA.

VIPPERLA, R. C., WOLTERS, M., GEORGILA, K. et RENALS, S. (2009). Speech input from older users in smart environments : Challenges and perspectives. *In HCI International : Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments*.

VOVOS, A., KLADIS, B. et FAKOTAKIS, N. (2005). Speech operated smart-home control system for users with special needs. *In Proc. InterSpeech 2005*, pages 193–196.

WÖLFEL, M. et MCDONOUGH, J. (2009). *Distant Speech Recognition*. Published by Wiley.