Interactions sonores et vocales dans l'habitat

Pierrick Milhorat[1], Dan Istrate[3], Jérôme Boudy[2], Gérard Chollet[1]

(1) Télécom ParisTech, 37-39 rue Dareau, 75014, Paris, France

(2) Télécom SudParis, 9 rue Charles Fourier, 91011 Evry Cedex, France

(3) ESIGETEL, 1 Rue du Port de Valvins, 77210 Avon Cedex, France

milhorat@telecom-paristech.fr, dan.istrate@esigetel.fr,
jerome.boudy@it-sudparis.eu, chollet@telecom-paristech.fr

RÉSUMÉ

Cet article présente le système de reconnaissance son/parole en continu développé et évalué dans le cadre du projet européen CompanionAble. Ce système analyse le flux sonore en continu, détecte et reconnaît des sons de la vie quotidienne et des commandes vocales grâce à un microphone. L'architecture et la description de chaque module sont détaillées. Des contraintes ont été imposées à l'utilisateur et au concepteur, telle que la limitation du vocabulaire, dans le but d'obtenir des taux de reconnaissance et de rejet acceptables. Les premiers résultats sont présentés, les essais finaux sur le terrain du projet sont en cours dans une maison intelligente à Eindhoven.

ABSTRACT

Acoustic Interaction At Home

This paper describes a hands-free speech/sound recognition system developed and evaluated in the framework of the European Project CompanionAble. The system is able to work continuously on a distant microphone and detect not only vocal commands but also everyday life sounds. The proposed architecture and the description of each module are outlined. In order to have good recognition rate some constraints were defined for the user and the vocabulary was limited. First results are presented; currently project trials are underway.

MOTS-CLÉS: reconnaissance vocale, traitement du son, reconnaissance des sons, domotique.

KEYWORDS: hands-free speech recognition, sound processing, sound recognition, domotics.

JEP-TALN-RECITAL 2012, Atelier ILADI 2012: Interactions Langagières pour personnes Agées Dans les habitats Intelligents, pages 17–30,
Grenoble, 4 au 8 juin 2012.

1. Introduction

Le projet européen CompanionAble a pour objectif l'intégration d'un robot compagnon dans une maison intelligente à destination de seniors dépendants. Le robot sert d'interface entre les fonctionnalités de la maison (allumage/extinction des lumières, ouverture/fermeture des rideaux, lecture/pause de la chaîne Hi-Fi…) ainsi que d'assistant à la vie quotidienne. A l'aide de capteurs disséminés dans l'habitat (capteurs infra-rouges, capteurs d'ouverture de porte…) et de ses propres informations (caméra, ultrasons…), il assiste les résidents suivant des scénarios prédéfinis (entrée dans la maison, sortie, appel en visioconférence…) ou définis par l'utilisateur (rappel de l'agenda, alerte de prise de médicaments, objets placés dans les paniers du robot…).
Dans ce contexte, le robot, porteur d'un écran tactile, un écran interactif installé dans la cuisine et une tablette portative sont équipés d'une application graphique identique présentant toutes les fonctionnalités disponibles.
L'Institut Mines-Télécom est en charge de porter l'interaction vers une communication vocale. Un ensemble de commandes domotiques dérivées d'expérimentation pratiques a été établis auxquelles le système d'analyse acoustique doit réagir.
L'interaction avec les applications autres telles que l'agenda, les exercices cognitifs ou le contrôle du robot a également fait l'objet de définitions de commandes. Dans les deux cas, les commandes ne sont pas uniquement des mots mais des phrases complètes.
De nombreux travaux ont porté sur la reconnaissance vocale et les performances des systèmes commerciaux actuels démontrent leur généralisation à venir. La spécificité de nos travaux inclus dans ce projet réside dans la résolution du problème de la distance du microphone au locuteur. Celui-ci, unique, se trouve intégré au robot, soit à environ un mètre du sol et mobile. La distance au locuteur et le bruit environnant sont incontrôlables et variables. Les travaux de soustraction du bruit à l'aide de microphones enregistrant les sources de bruits ne s'appliquent pas aux conditions variables rencontrés lors des études d'usage préliminaires.
La deuxième section de cet article décrit le projet CompanionAble. Les sections 3 et 4 sont consacrées à la reconnaissance des sons. Vient ensuite la description du système de reconnaissance de la parole utilisé et adapté à ce contexte, en section 5. La sections 6 présente les évaluations du système. Les futurs axes de recherche et conclusions tirés de ce projet seront exprimés dans la dernière partie.

2. CompanionAble

CompanionAble est l'acronyme de Integrated Cognitive Assistive & Domotic Companion Robotic Systems for Ability & Security.
C'est un projet financé par la commission européenne qui réunit 18 partenaires académiques et industriels.
Les objectifs sont les suivants :

- combiner les capacités d'un robot « compagnon » mobile avec les fonctionnalités statiques d'un environnement intelligent.
- intégrer les données des capteurs de l'habitat à celles du robot
- créer un lien social entre les séniors et leurs proches et/ou leurs entourage médical
- améliorer la qualité de vie et l'autonomie des personnes dépendantes

Les partenaires sont localisés dans 7 pays, à savoir : la France, l'Allemagne, l'Espagne, l'Autriche, la Belgique, les Pays-Bas et le Royaume-Unis.
L'Institut Mines-Télécom (anciennement Groupe des Ecoles des Télécommunications) a pour rôle l'addition d'une interface vocale, d'un module multimodal de détection des situations de détresse et participe également à la localisation des personnes dans l'habitat. Cet article se concentre sur la partie acoustique des travaux.
Actuellement, le projet est entré dans une phase d'expérimentation pratique en situation réelles. Cela se déroulera à Eindhoven (Pays-Bas) et à Gits (Belgique) un panel d'utilisateurs potentiels sera amené à tester le système complet.

3. Architecture de traitement du son

Le son est acquis en continu par deux systèmes parallèles : l'un attribue un type au son (parole/son et type de son) tandis que l'autre transcrit la parole en texte. La figure 1 montre la communication entre les modules sonores qui utilise un protocole TCP/IP. Les sons reconnus ou les commandes vocales sont transmises au serveur du projet via un protocole SOAP. La reconnaissance vocale est filtrée par le module de reconnaissance des sons pour éviter les fausses alarmes (faux-positifs).

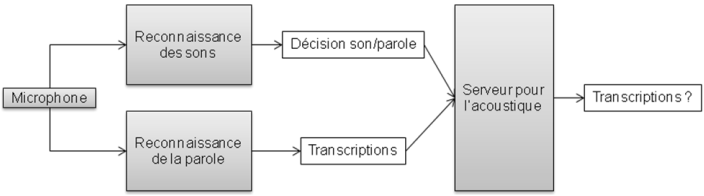

Figure 1 – Architecture de traitement des sons

4. Reconnaissance des sons

Le système de reconnaissance sonore consiste dans notre application en une séquence de deux procédés : un module de détection basé sur unetransformée en ondelettes et un module de reconnaissance hiérarchique (son/parole et classification des sons) basé sur des GMM (Rougui, 2009).

Les classes de sons utilisées lors des essais CompanionAble ont été apprises avec des enregistrements effectués avec un CMT (microphone produit par AKG) (Rougui, 2009) dans la maison SmartHomes à Eindhoven. Actuellement, le système dispose de 5 classes de son : chute d'objets, sonnette, clés, toux et applaudissements. Les classes de son ont été choisies pour la détection de situations de détresse et d'actions non parlées utiles pour le système domotique.

La sortie de la reconnaissance vocale est filtrée par le module de reconnaissance des sons pour empêcher une fausse commande associée à un son non parlé. Comme les deux modules tournent en parallèle, une synchronisation est requise. Le module sonore enregistre constamment les trois dernières décisions d'étiquetage sons/parole et décide de valider ou rejeter la reconnaissance vocale en fonction de la corrélation entre les deux.

Chaque module a été initialement évalué sur des bases de données. Les résultats de la classification des sons selon 15 classes sont de 80% de bonne reconnaissance . La reconnaissance parole/son a été et sera évaluée dans la maison SmartHomes ; les premiers résultats ont montré un taux avoisinant les 95%

5. Reconnaissance de la parole

Concernant le volet reconnaissance vocale, il se justifie, au sein du projet par la difficulté, voire l'incapacité d'une grande partie des utilisateurs cibles d'interagir avec un système informatique par le biais de menus sur des écrans tactiles. Les troubles cognitifs

ou des problèmes de mobilité trop importants rendraient le système obsolète s'il n'était pas doté d'une interface vocale à distance. Les commandes interprétables portent sur l'interaction « face au robot » combiné avec l'affichage graphique et sur les interactions à distance. Les trois problématiques auxquelles nous proposons une solution sont :

- la reconnaissance de commandes vocales dans un environnement bruité pour lequel les sources de bruit sont inconnues
- la reconnaissance de commandes vocales à distance variable
- la reconnaissance de commandes vocales toujours active

Les centres d'expérimentation basés aux Pays-Bas et en Belgique flamande contraignent le projet à réaliser l'interface vocale du robot en hollandais.

Julius, développé par le Kawahara lab de Tokyo, a été sélectionné comme étant le décodeur le plus approprié pour une application basique état de l'art (Lee, 2008). Il permet une reconnaissance sur un large vocabulaire (60 000 mots) en temps quasi-réel grâce à un algorithme à deux passes. Il s 'appuie sur des modèles de langage N-grams et des modèles acoustiques encodées sous forme de modèles de Markov cachés. La modularité de ce moteur de reconnaissance permet de traiter une même entrée (audio) avec plusieurs modèles (acoustiques et de langage) différenciés selon les besoins.

Les modèles de Markov cachés des phonèmes composant le modèle acoustique hollandais ont été appris sur le Corpus Gesproken Nederlands (CGN). Cela représente 800 heures d'enregistrements audio transcrits dans lesquelles presque 9 millions de mots sont prononcés, faisant du CGN le plus grand corpus pour le hollandais contemporain. Les sources sont réparties entre des sources monolocuteurs et multilocuteurs, promptées ou spontanées.

Français	**Fenêtre**	**Hollandais**
Non	GoodBye Frame	Nee
Hector	Main Frame	Hector
Je reviens tout de suite	GoodBye Frame	Ik ben zo terug/Ik ga niet weg/Zo terug
Quelques jours	GoodBye Frame	Een paar dagen/Paar dagen
Environ une heure	GoodBye Frame	Een uurtje/Een uur/Uur
Affiche les appels manqués	Greeting Frame	Gemiste oproepen/Laat gemiste oproepen
Affiche la liste des choses à faire	Greeting Frame	Taken/Laat taken zien/Start taken

Table 1 – Exemple de commandes vocales

Étant données les conditions imposées (toujours actif, distance au microphone variable, variété des fonds sonores, etc...), des solutions ont été proposées pour améliorer la robustesse du système.

« L'attention »

Le gestionnaire du dialogue propose un moyen de limiter le nombre de faux-positifs avec l'utilisation d'un mot « d'attention ». Ce mot, quand il est détecté, accroît le niveau d'attention qui décroît avec le temps. Un niveau d'attention non nul déclenche le traitement et l'analyse des données de la reconnaissance. Par exemple, à l'état initial, le niveau d'attention est nulle : le module de reconnaissance vocale est toujours actif et transmet ces résultats, tant que le mot clé n'est pas détecté, les transcriptions sont ignorées par le gestionnaire de dialogue. Dès lors que le niveau d'attention est supérieur à 0, le dialogue s'engage, et le gestionnaire interprète toutes les commandes reçues. Tant que le dialogue est soutenu (soit par répétition du mot d'attention, soit par une évolution du dialogue), le niveau d'attention s'accroît alors que les silences, du point de vue du gestionnaire de dialogue diminuent la variable d'attention qui, si elle atteint sa valeur plancher (nulle) coupe l'interprétation. Le choix d'un mot d'attention le plus discriminatoire possible augmente l'efficacité d'un tel mécanisme.

La classification sons/parole

Un moteur de reconnaissance vocale tel que Julius cherche la séquence de mots qui correspond au mieux à la séquence de vecteurs acoustiques présentée selon les probabilités contenues dans la combinaison des modèles acoustiques et de langage. Il est possible de créer un mot « poubelle » qui remplacerait l'ensemble des mots dont le score de reconnaissance serait trop faible. Dans notre application, nous avons choisi de décoder systématiquement les sons de l'habitat. Ainsi, les bruits qui se distinguent de la parole sont mis en correspondance avec une séquence de mots erronée. La classification des sons en deux catégories (parole/non-parole) permet un filtrage des données acoustiques. Ce filtrage est effectué en parallèle du processus de reconnaissance pour conserver les aspects temps réel inhérents au projet.

L'adaptation

Les techniques d'adaptation qui auraient pu être utilisées ont été les premières à être implémentées et testées. Un modèle de langage (N-grams) a été élaboré sur un corpus de

plus de 57500 phrases dérivées d'expériences pratiques et de paraphrases.
Une comparaison entre deux procédures d'adaptation, Maximum A Posteriori (MAP) et Maximum Likelihood Linear Regression (MLLR), a été faite (Caon, 2011).
Le locuteur est le même pour toute l'expérience. Il a été enregistré et les fichiers audio sont joués par un haut-parleur. Comme prévu, étant donné le peu de donnée d'adaptation disponibles (10 phrases par locuteur), l'adaptation par MLLR donne donc les meilleurs résultats. Sans adaptation, 60% des allocutions ont été correctement retranscrites par Julius. Ce taux s'élève à 70% avec l'adaptation par MAP et 73% avec l'adaptation MLLR. De fait, MLLR a été confirmé comme la technique d'adaptation la plus idoine.

La combinaison de modèles de langage

Dans une première version de l'application, un N-gram unique, appris sur un corpus de 57 658 phrases a été utilisé. La voix de chaque utilisateur était adaptée avant l'utilisation du système : la reconnaissance est ciblée pour un utilisateur déterminé par l'adaptation préalable de sa voix. Cette première version présentait trop de « faux-positifs », i.e. de commandes non désirées lors de tests pratiques.
Dans le but d'améliorer à la fois les taux de rejet et de reconnaissance, un filtre, décris ci-dessous, a été implémenté.
Le gestionnaire de dialogue modélise le dialogue comme un ensemble de fenêtres (Müller, 2010). Celles-ci contiennent chacune un graphe du sous-dialogue pour lequel les transitions sont déclenchées par l'état de variables internes au robot ou par des actions de l'utilisateur (commande vocale, pression de bouton, excitation de capteur...). Une fenêtre devient active lorsque l'une de ses conditions suffisantes d'activation est remplie, ce sont les même type de variable que celles associées aux transitions intra-fenêtre. De fait, il est possible de construire une hiérarchie du dialogue. La fenêtre principale ou racine, initialement active, contient (uniquement) tous les déclencheurs des sous-dialogues. Les fenêtres contiennent des noeuds terminaux, ce qui permet une auto-désactivation et un retour à la fenêtre principale.

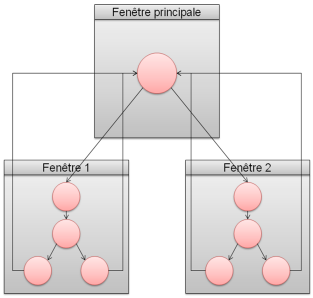

Figure 2 – Système de fenêtre du gestionnaire de dialogue

Les différentes sous-fenêtres ont été regroupées en 8 catégories. A chaque catégorie correspond un ensemble des commandes vocables possibles dans les sous-dialogues qui la compose. Pour chaque catégorie, un modèle de langage a été appris sur l'ensemble des commandes correspondantes.
Un 9ème modèle est appris sur l'ensemble des commandes vocales qui activent les sous-fenêtres, il est associé à la fenêtre principale.
Bien que le module de reconnaissance vocale ne connaisse pas l'état du dialogue (la fenêtre active) puisque la communication est uni-directionnelle pour garantir au mieux la synchronisation et parce que Julius le permet, 9 instances du décodeur, paramétrisées avec un modèle acoustique identique et un modèle de langage spécifique, analyse en parallèle les sons de l'habitat.
Cette méthode permet de favoriser en grande partie les commandes autorisées selon l'état du dialogue, cependant, elle aggrave les problèmes de rejet des allocutions en dehors de l'application.

Le test de similarité
La similarité entre deux hypothèses est mesurée en terme de distance de Levenshtein sur les mots. Elle cumule le nombre de substitution, d'ajout et de retrait de mots. De plus elle est ensuite divisée par la longueur des phrases, rendant alors une moyenne du nombre de différences par mots. La valeur de cette variable, relative à un seuil, définit la validation ou le rejet de l'hypothèse de commande vocale reconnu par l'instance du décodeur basée sur un vocabulaire restreint. Ce test permet de :

- confirmer les hypothèses correctes : une commande reconnue correctement (vrai-positif) par un décodeur spécialisé et reconnue correctement par le

décodeur général est validé par le test, l'hypothèse fournie par le décodeur spécialisé est transmise.

- rejeter les hypothèses incorrectes : une commande reconnue incorrectement (faux-positif) par un décodeur spécialisé et reconnue correctement ou similairement par le décodeur général est rejetée par le test, l'hypothèse fournie par le décodeur spécialisé est ignorée.

- corriger les hypothèses partiellement incorrectes : une commande reconnue correctement par un décodeur spécialisé et reconnue similairement par le décodeur général est validée par le test, l'hypothèse fournie par le décodeur spécialisé est transmise.

Le modèle de langage général doit, dans cette configuration, pouvoir modéliser les séquences de mots définies dans les modèles de langage spécialisés. Il est donc nécessaire d'ajouter les commandes vocales dans le corpus d'apprentissage du modèle général, de plus, nous introduisons un poids à ces commandes. Le poids optimal a été défini expérimentalement comme étant 1000, i.e. les commandes vocales ont été ajoutées mille fois au corpus CGN avant l'apprentissage.

Finalement, le test de similarité ne s'effectue pas sur une transcription par décodeur, il a été découvert, expérimentalement, que l'utilisation des n meilleures hypothèses améliorait le taux de reconnaissance sans impacter sensiblement le taux de rejet :

- au vu de la taille des modèles de langage restreints, seule la meilleure hypothèse est comparée.

- plusieurs hypothèses (les 3 meilleures dans notre application) fournies par le décodeur général passent le test de similarité.

L'ensemble de ces améliorations ont été implémentées, pour certaines directement dans le code de Julius ou du gestionnaire de dialogue. Elles sont évaluées dans la section suivante.

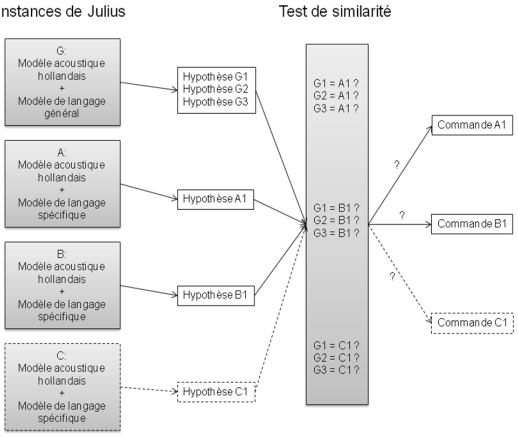

Figure 3 – Système de reconnaissance de la parole incluant le test de similarité

6. Évaluations

Une batterie de test pour éprouver la robustesse du système a été effectuée, nous présentons ici les résultats les plus probants et significatifs.

Pour l'ensemble des évaluations, un corpus de test a été préalablement enregistré auprès de 5 résidents hollandais. Chacun d'eux a enregistré 58 phrases : 10 phrases d'adaptation, 20 commandes de l'application, 22 allocutions hors application et 6 commandes dérivées. Une commande dérivée est une commande composée du vocabulaire de l'application mais dont la grammaire est inexacte.

L'installation de l'expérience est présenté schématiquement sur la figure 4. Les séquences sonores en hollandais sont produites par un haut-parleur et enregistrées par un microphone à distance variable. Un second haut-parleur, placé au-dessus du premier simule des bruits ambiants dans la seconde phase de l'expérience. Le volume sonore des locuteurs hollandais est ajusté aux situations réelles (environ 60 dBA).

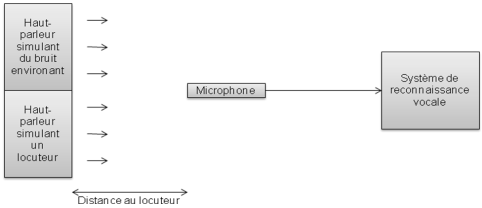

Figure 4 – Installation pour l'expérience

Lors d'une première phase, l'addition du test de similarité, le poids des commandes vocales dans le corpus d'apprentissage du modèle général et le nombre d'hypothèses présentées à la comparaison par le décodeur général sont évalué.

Dans le cas de commandes vocales autorisées par l'application, le décodeur de base, qui utilise un modèle unique de langage général comprenant les commandes vocales sans poids associé, obtient des résultats de reconnaissance de 15%. L'ajout du test de similarité sans modifier ce modèle de langage général porte le taux à 20% mais laisse apparaître des faux-positifs. Le système le plus évolué obtient des performances de 85% de reconnaissance pour un taux de faux-positifs nul.

Toutes les allocutions hors de l'application sont rejetées par le système.

En ce qui concerne les commandes dérivées, elles sont peu rejetées car proche des commandes réelles.

Système	Taux de reconnaissance	Taux de faux-positifs
Base + adaptation	15	0
Base + adaptation + test de similarité (poids des commandes : 1 ; hypothèses du décodeur général : 1)	20	10
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 1)	55	0
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 3)	85	0

Table 2 – Taux de reconnaissance correcte et de faux-positifs pour les commandes de l'application

Système	Taux de reconnaissance	Taux de faux-positifs
Base + adaptation	9.09	0
Base + adaptation + test de similarité (poids des commandes : 1 ; hypothèses du décodeur général : 1)	0	0
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 1)	0	0
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 3)	0	0

Table 3 – Taux de reconnaissance correcte et de faux-positifs pour des allocutions hors application

Système	Taux de reconnaissance	Taux de faux-positifs
Base + adaptation	16.67	0
Base + adaptation + test de similarité (poids des commandes : 1 ; hypothèses du décodeur général : 1)	33.33	0
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 1)	66.67	0
Base + adaptation + test de similarité (poids des commandes : 1000 ; hypothèses du décodeur général : 3)	66.67	0

Table 4 – Taux de reconnaissance correcte et de faux-positifs pour des commandes de l'application dérivées

La deuxième phase de l'expérience consistait en un test de résistance au bruit. Un second haut-parleur, simulant du bruit ambiant (non-stationnaire) est ajouté au dispositif. Cela a pour conséquence de diminuer les performances du système, autant du point de vue de la reconnaissance, que de celui du rejet.

Système	Taux de reconnaissance	Taux de faux-positifs
Machine à laver	74	11
Locuteur hollandais	53	11
Musique	47	5
Foule	42	11

Table 5 – Taux de reconnaissance correcte et de faux-positifs pour les commandes de l'application

Système	Taux de reconnaissance	Taux de faux-positifs
Machine à laver	0	0
Locuteur hollandais	0	0
Musique	0	0
Foule	0	3.64

Table 6 – Taux de reconnaissance correcte et de faux-positifs pour des allocutions hors application

Système	Taux de reconnaissance	Taux de faux-positifs
Machine à laver	40	0
Locuteur hollandais	60	0
Musique	20	0
Foule	60	0

Table 7 – Taux de reconnaissance correcte et de faux-positifs pour des commandes de l'application dérivées

7.Conclusion et perspectives

Le système présenté dans cet article propose de compléter l'interaction entre un robot et un humain par ajout de commandes vocales dans une maison intelligente. Le robot est toujours actif, tout comme doit l'être l'analyse des commandes vocales accessibles à tout moment. De par ces contraintes, la caractéristique la plus importante dont on doit tenir compte est la robustesse d'un tel système. Cela combine à la fois un taux de reconnaissance correct et un taux de rejet acceptable.

Un équilibre entre ces deux aspects doit être trouvé. Accepte-t-on de reconnaître des commandes erronées ? Peut-on demander à l'utilisateur de répéter plusieurs fois les commandes ? Pendant les tests pratiques précédents, il a été démontré que les faux-positifs perturbaient l'utilisateur et généraient des comportements inattendus. Pour parer à ce problème, les possibilités de commandes ont été restreintes, créant deux nouveaux obstacles. Les utilisateurs cibles sont des séniors qui pourraient avoir des difficultés à se rappeler les commandes précises. De plus, ils pourraient rapidement se désintéresser des fonctionnalités vocales s'ils perçoivent une fiabilité faible dans l'exécution des ordres qu'ils émettent.

Nous avons proposé, dans ce travail, d'expérimenter une combinaison de modèles de langage associée à un test de similarité pour améliorer la précision du système.

Un nouveau modèle de langage général a été construit à partir de la base néerlandaise CGN, supposons qu'il est capable de reconnaître n'importe quelle phrases en néerlandais ou une phrase proche. Une passe de la reconnaissance utilise un modèle de langage spécifique à l'application, voire à une partie de l'application. La similarité entre les deux sorties, i.e. la distance de Levenshtein entre les deux phrases, agit comme un filtre pour valider ou rejeter les sorties.

Ce système plus élaboré a démontré être plus robuste, autorisant un taux de reconnaissance correct ainsi que des cas limité de faux-positifs. Cependant, l'expérience a montré ses faiblesses dans le traitement et le rejet des allocutions courtes, i.e. phrases composées d'un seul mot. L'utilisation du mot-clé « d'attention » empêche la plupart du temps ce genre de situation de se produire.

Courant avril et mai de l'année 2012, des essais en situation réelle auront lieu à Eindhoven et Gits. Des couples de seniors sont invités à vivre dans une maison intelligente dans laquelle un robot compagnon interagira avec eux. Jusqu'à présent, le

test en conditions réelles le plus significatif eu lieu dans cette même maison (Eindhoven) dans un environnement sonore stationnaire. Par stationnaire, il est entendu que des personnes parlaient dans les pièces voisines et que leur voix parvenaient dans la pièce de test sans qu'à elles seules elles ne déclenchent le processus de reconnaissance programmé pour être effectif à partir d'un certain niveau sonore perçu. Un locuteur néerlandais, qui avait précédemment adapté le modèle acoustique à sa voix, prononce dès lors les 168 commandes définies à ce moment. Il était autorisé à prononcer une seconde fois les commandes mal ou non reconnues au premier essai. Le taux de reconnaissances correctes constatées s'élève alors à 89%, constituant le seuil bas pour l'évaluation de l'évolution du système. Pour des raisons de respect de la vie privée, les essais pratiques à venir ne seront pas enregistrés, un protocole d'évaluation devrait être établi pour reporter les résultats significatifs et scientifiques.

8. Remerciements

Ce travail a été soutenu par le projet européen CompanionAble. Nous remercions AKG (Vienne) et SmartHome (Eindhoven) pour leur appui. Nous remercions également Daniel Caon et Pierre Sendorek pour leur aide dans les premières implémentations du système de reconnaissance.

9. Références

LEE, A. (2008). *The Julius Book*.

ROUGUI, J. E., ISTRATE, D. et SOUIDENE, W. (2009). Audio Sound Event Identification for distress situations and context awareness. *In EMBC2009*, September 2-6, Minneapolis, USA, pp. 3501-3504.

ROUGUI, J. E., ISTRATE, D., SOUIDENE, W., OPITZ, M. et RIEMANN, M. (2009). Audio based surveillance for cognitive assistance using a CMT microphone within socially assistive technology. *In EMBC2009*, September 2-6, Minneapolis, USA, pp.2547-2550.

CAON, D., SIMMONET, T., BOUDY, J. et CHOLLET, G. (2011). vAssist: The Virtual Interactive assistant for Daily Home-care. *In pHealth conference, 8th International Conference on Wearable Nano and Macro Technologies for Personalized Health*, Lyon, France.

MÜLLER, S., SCHROETER, C. et GROSS, H.-M. (2010). Aspects of user specific dialog adaptation for an autonomous robot. *In International Scientific Colloquium*, Ilmenau, Allemagne.