vAssist : Le Majordome des personnes dépendantes.

Gérard Chollet, Daniel Régis Sarmento Caon, Thierry Simmonet, Jérôme Boudy

vAssist : Le Majordome des personnes dépendantes

Gérard Chollet[1] , Daniel R.S.Caon[1] , Thierry Simonnet[2] et Jérôme Boudy[3],

[1] CNRS-LTCI, TELECOM-ParisTech, 46 rue Barrault, 75634 Paris cedex 13, France
{gerard.chollet, daniel.caon}@telecom-paristech.fr
[2] ESIEE , 2 boulevard Blaise Pascal, Cité DESCARTES,BP 99,
93162 Noisy le Grand cedex, France
t.simonnet@esiee.fr
[3] TELECOM-SudParis, 9, rue Charles Fourier,
91011 Evry cedex, France

Résumé. vAssist est un projet du programme 'Ambient Assisted Living' de la communauté européenne. Il propose le développement et l'expérimentation d'un assistant personnel centralisé, le Majordome, pour les personnes dépendantes. Ces personnes utilisent un smartphone pour se connecter à un serveur. Elles dialoguent avec leurs Majordomes en Voix et Vidéo sur IP. Le Majordome détecte les situations de détresse et contacte les aidants et services médicaux si nécessaire. Cet article donne quelques détails sur l'architecture de télécommunication, sur le système de dialogue vocal et des résultats d'évaluation du système de reconnaissance automatique de la parole.

Keywords: Majordome, memex, MyLifeBits, dialogue vocal, VVoIP

1 Introduction

C'est en 1945 que V. Bush a décrit un appareil électronique (le memex) relié à une bibliothèque et capable d'afficher des livres et de projeter des films [5]. Cet article a inspiré les créateurs du 'World Wide Web' (le web). Le terme 'memex' est une contraction de 'memory extender'. Le web peut-être considéré comme la mémoire du monde; une partie est publique (tout utilisateur d'internet peut y accéder), une partie est privée (seuls certains utilisateurs, et parfois un seul, y ont accès).

Notre mémoire, souvent défaillante, est complétée par celle du web. Des moteurs de recherche nous facilitent l'accès aux informations disponibles sur le web. Il suffit d'utiliser un téléphone portable pour accéder à toutes ces informations.

Des informations personnelles peuvent aussi être collectées et indexées [8]. Cette mémoire peut être disponible sur un serveur personnel local ou distant. C'est notre prothèse mémorielle [11].

Une population visée est celle atteinte de la maladie d'Alzheimer, sous des formes précoces ('Mild Cognitive Impairment') [9], [1], [12], mais les personnes âgées et d'autres personnes pourront utiliser ce dispositif. Notre hypothèse est qu'un système Majordome adapté aux difficultés des personnes et contrôlé par elles-mêmes pourrait contribuer à leur soutien au domicile en leur fournissant une stimulation cognitive et en jouant le rôle de prothèse cognitive. Le Majordome, conçu pour s'inscrire dans un écosystème d'objets communicants, comporte deux parties:

- Une plate-forme fixe de divertissement qui regroupe les exercices de stimulation cognitive, la téléphonie, l'accès à internet, la télévision, la vidéo-conférence (relation avec l'aidant, la famille et le personnel soignant), la gestion des documents administratifs, médicaux, bancaires, impôts, factures,...
- Un système portable et communicant, le Majordome qui permet au patient d'alerter l'aidant en cas de chute [4] ou de comportement anormal, de connaître sa localisation en dehors de son domicile, de rester en contact téléphonique avec son aidant, de se faire rappeler ses tâches,...

Cet article est structuré comme suit :

- Quelques scénarios d'usage sont proposés en section 2 ;
- l'architecture de Voix et Vidéo sur IP est détaillée en section 3 ;
- le système de dialogue vocal en décrit en section 4 ;
- les résultats d'évaluation du système de reconnaissance automatique de la parole sont fournis en section 5 ;
- suivi par des conclusions et perspectives.

2 Scénarios d'usage

Supposons que l'on dispose en permanence d'un accès à une prothèse mémorielle personnelle et collective relayé par un Majordome audio-visuel. Quels sont les usages que nous pourrions en faire ? En voici une liste non exhaustive :

- retrouver son chemin car le Majordome matérialisé par un SmartPhone, relié au GPS, connaît notre localisation et peut nous guider,
- Enregistrer de courtes vidéo ou des photos pour faire un album de son proche passé,
- se souvenir du nom et autres infos sur une personne que l'on rencontre; le Majordome est équipé d'une caméra, prend une photo et retrouve ces informations,
- faire des achats; liste des courses, prix,...

- fournir des recettes de cuisine, se souvenir des plats que l'on a préparés pour ses amis, sa famille,...
- consulter et mettre à jour son agenda, ses rendez-vous, les factures à régler, les fêtes à souhaiter,
- répondre au téléphone, à la messagerie,...
- rechercher des infos sur le web,
- détecter des situations de détresse, les comportements anormaux,…

Certaines de ces fonctionnalités sont déjà disponibles sur des SmartPhones, d'autres sont en cours de développement comme le projet MyLifeBits de Microsoft [8].

3 Voix et Vidéo sur IP

Le Majordome (smartphone) communique avec le serveur en conformité avec le protocole SIP.

3.1 SIP (Session Initiation Protocol)

L'utilisation d'une infrastructure de VoIP implique la mise en place d'un PABX (auto-commutateur) logiciel. Nous avons retenu le produit de DIGIUM [2]. Celui-ci est configuré pour gérer les communications vocales, mais également pour prendre en charge la vidéo. Le réseau local des patients étant en adressage IP privé, un auto-commutateur local doit être installé chez chaque patient et raccordé au PABX central. Cette architecture utilisera la topologie en trapèze SIP:

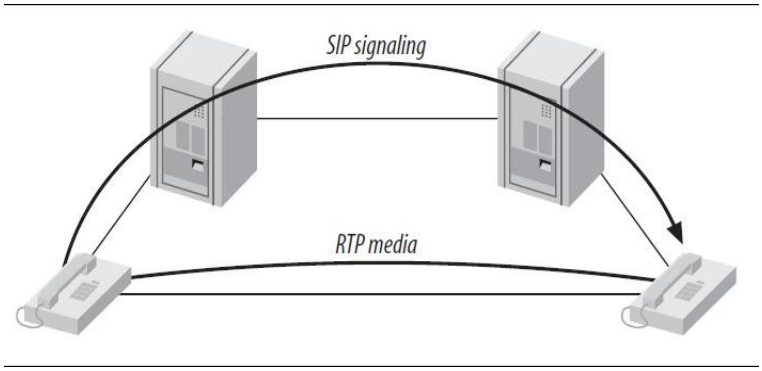

Fig. 1. Architecture en trapèze SIP.

Lors de l'appel d'un correspondant, une requête SIP est envoyé au PABX, qui la transmet, en fonction du plan de numérotation, directement à son correspondant ou dans notre cas au PABX enregistré du correspondant. Une fois ce dialogue effectué, une communication directe entre les 2 partenaires est établie en RTP (Real Time protocol) (Cf Fig 2). Ce protocole, sur UDP, permet de conserver l'ordre des paquets et d'abandonner ceux trop anciens.

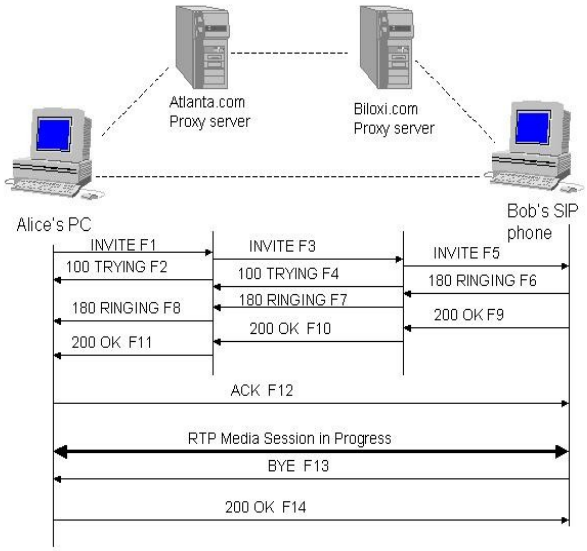

Fig. 2. Déroulement d'un appel.

La figure 3 illustre le positionnement des différentes composantes dans le système de couches OSI.

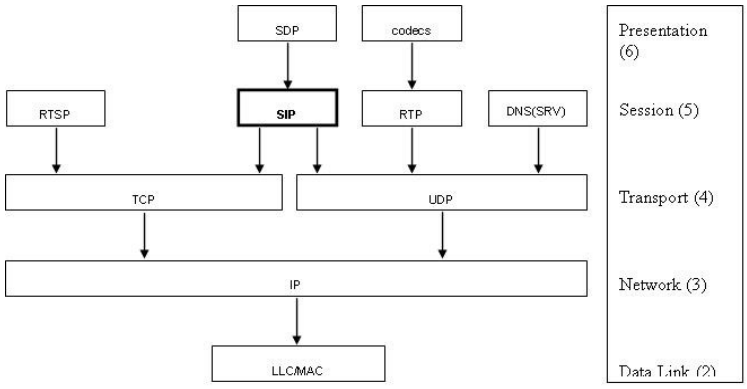

Fig. 3. SIP et les couches OSI.

3.2 PABX

Le PABX Asterisk est configuré de manière à offrir toutes les fonctionnalités de téléphonie classique et d'offrir les services de transmission vidéo.

Nous avons également mis en place les services de boite vocale, de manière à banaliser l'architecture et à offrir un service le plus proche possible du téléphone classique.

Un composant essentiel pour la transmission vidéo est le CODEC. Un codec (Code-DECode) est un procédé capable de coder un signal analogique et de décoder un signal numérique. Dans le cas de la VoIP, cela désigne aussi bien la norme que le module chargé d'encoder et de décoder la voix ou l'image. Le codec H264 encode et décode les flux respectant la nome MPEG-4 AVC/H264.

Les PABX n'ont pas la charge de traduire les flux en provenance d'un codec d'un client A pour le transmettre vers un autre codec du client B. Il faut donc que l'ensemble des codecs clients **ET** PABX soient homogènes et inter-opérables.

A ce jour les codecs utilisable avec Asterisk sont:

- Pour le son : µlaw, alaw, gsm, ilbc, speex, g726, adpcm, lpc10, g729, g723
- Pour la vidéo : h261, h263, h263+, h264

Il faut trouver un équilibre entre les facteurs « compression », « délai » et « qualité d'image ». La compression et le délai sont liés. Un facteur de compression élevé entraînera un délai de bufferisation élevé. Les codecs haute définition comme H264 sont actuellement peu adaptés aux transmissions de type visiophonie. Ils sont utilisés pour la télévision HD, application ou le délai de transmission n'est pas critique (flux unidirectionnel).

Nous nous sommes limités à l'utilisation de l'encodage H261 qui offre un ratio performance/délai correspondant à notre besoin.

3.3 Client SIP

Un client SIP doit fournir, pour notre application de visiophonie, un service contraint: faible latence et bonne qualité d'image. Un système de QoS doit être établi pour garantir une fluidité constante. Le client SIP doit être en mesure d'équilibrer dynamiquement le taux de compression et la qualité des différents canaux (vidéo et audio) pour s'adapter à la bande passante. Il doit également assurer l'arbitrage entre les flux audio et vidéo. Une application de la surveillance par exemple n'a pas besoin d'une grande qualité de son, mais d'une latence faible, alors que la priorité pour une application de visiophonie est la voix.

Le PABX Asterisk n'assure pas de fonction de traduction de codecs. C'est pourquoi le choix des codecs utilisés et le soft-phone est important. Le protocole SIP permet aux deux clients (appelant et appelé) une auto-négociation afin d'obtenir un couple de codecs pour l'encodage audio et vidéo compatible et optimal.

Nous avons choisi Ekiga comme client SIP pour plusieurs raisons:

- Il est Open-source et peut donc être modifié.
- Il est portable (Linux, Windows).

- Il offre nativement un bon nombre de codecs (H261, H264, Theora, PCMU, GSM ,...)
- De par son architecture, l'utilisation de la *libglade* (bibliothèque de codes) nous permet de modifier l'interface graphique à partir de fichiers XML qui sont crées avec Glade (le constructeur Gnome GUI).

Nous avons travaillé sur le client Ekiga, en intégrant le projet OpenSource et en prenant en charge la version Windows de ce projet.

Nous avons donc contribué en ajoutant une fonction de réponse automatique, qui peut être intéressante pour l'appel de personnes âgées. Cette fonction est débrayable pour garder la possibilité de ne pas répondre à un appel.

Nous avons également modifié l'interface graphique pour la simplifier (l'interface présente 4 gros boutons, sur lequel il est possible d'afficher la photo du correspondant. Les numéro de téléphone sont également pré-programmés évitant ainsi toute possibilité d'erreur dues à la manipulation du clavier et/ou de la souris.

Les premières expérimentations ont montré un délai de transmission sur réseau local de l'ordre de 400ms. Ces délais ont été ramenés a 80ms en utilisant internet, sans QoS.

3.4 Limitations et Evolutions

- Nous avons constaté une incompatibilité quant à l'utilisation de Windows Vista. Nous nous sommes concentré sur une cible Windows XP.
- Un gros travail a été effectué pour réduire les délais de transmissions (entre 30ms et 80ms)
- Le PABX Asterisk est public (à condition d'avoir un compte) et opérationnel avec tout client SIP (Ekiga, X-Lite, linphone, ...): wagram.esiee.fr.
- Les services proposés sont les suivants:
  - Conférence : Changement de canal vidéo par DTMF (appui de la touche du numéroteur)
  - Messagerie instantanée en cours de conversation.
  - Interfaçage en cours avec gtalk, msn, aim et yahoo messenger.
  - Interfaçage avec Julius (ASR) pour la reconnaissance vocale.
  - IVR vidéo

4 Dialogue vocal

La communication entre le client et son majordome se fait de façon naturelle par la voix. Pour cela un moteur de dialogue est installé sur le serveur. Ce moteur utilise un système de reconnaissance open source Julius disponible sur la plate forme Asterisk.

4.1 Julius et HTK

Le module de reconnaissance vocale est fondé sur l'usage des Modèles de Markov Cachés classiques (MMC) permettant de modéliser de manière statistique les modèles acoustiques des phonèmes ou/et mots utilisés dans le vocabulaire visé à l'aide d'outils logiciels comme HTK [14] et Julius [10]. Les modèles de langage (probabilités linguistiques complémentaires aux probabilités acoustiques) sont implicitement abordés dans l'usage de tels modèles statistiques afin de rendre robuste la reconnaissance des mots dans une phrase donnée (utilisation des statistiques N-grams et des règles de grammaire).

4.1 Données et Adaptation

Des stratégies d'adaptation seront mises en œuvre, en particulier basée sur l'adaptation multi-langue croisée entre langues ayant des matériaux phonétiques riches.

Les travaux de recherche de Tania Schultz [13], Rania Bayeh [3] et Gérard Chollet [7], concernant la reconnaissance de parole multi-linguale et indépendante de la langue, servent de point de départ.

5 Évaluation

Une première validation des logiciels de reconnaissance automatique de la parole a été réalisée sur des données enregistrées dans le cadre du projet européen CompanionAble (www.companionable.net). Chacun des 22 locuteurs hollandais ont été enregistrés pendant une heure dans une maison expérimentale (SmartHomes) à Eindhoven. Ils ont répétés des phrases prononcées par un 'souffleur'. Le tableau suivant (contenant 37 phrases différentes) donne des résultats obtenus pour un de ces locuteurs, après adaptation des modèles acoustiques à sa voix par MLLR (*Maximum Likelihood Linear Regression*, technique classique d'adaptation aussi étudiée dans [6]).

Tableau 1. Résultats pour un locuteur en parlant 37 phrases spécifiques, 10 fois chacune.

Phrase (prononcée 10 fois par le même locuteur)	**Phrases correctes (%)**	**Sémantique correcte (%)**
hellep	100	100
help me	50	90
kom naar de keuken	100	100
kom eens naar de keuken	100	100
wil je naar de keuken komen	60	60
...		

hector ik ga lunchen met kennisen	100	100
hector ik ga lunchen met buren	100	100
wolly ik ga uit eten	100	100
wolly ik ga uit eten met vrienden	100	100
wolly ik ga uit eten met kennisen	100	100
ja graag	40	80
Pourcentage moyenne	**86.39**	**94.44**
Pourcentage plus basse	**40**	**60**
Pourcentage plus haute	**100**	**100**

Le taux “sémantiquement correct” est une manière de décrire que deux phrases sont au même niveau de sens (e.g. “help me” et “hellep”), donc si “help me” est reconnu au lieu de “hellep”, la phrase est 100% correcte (sémantiquement) et le dialogue vocal peut se dérouler sans problème.

La deuxième validation des logiciels a été fait sur 20 des locuteurs (tous âgés), sans répétition de phrases. Une technique d'adaptation classique MAP (*Maximum a Posteriori*, aussi étudiée dans [6]) a été appliquée à partir d'un set de 10 phrases d'adaptation pour chaque locuteur. La figure 4 donne les taux de reconnaissance de mots par des modèles de langage 2-gram et 6-gram, avec et sans adaptation des modèles acoustiques. On constate l'amélioration obtenue grâce à l'adaptation et à la précision du modèle de langage. On voit aussi que les locuteurs n'utilisent pas tous le système de reconnaissance avec les mêmes taux de succès.

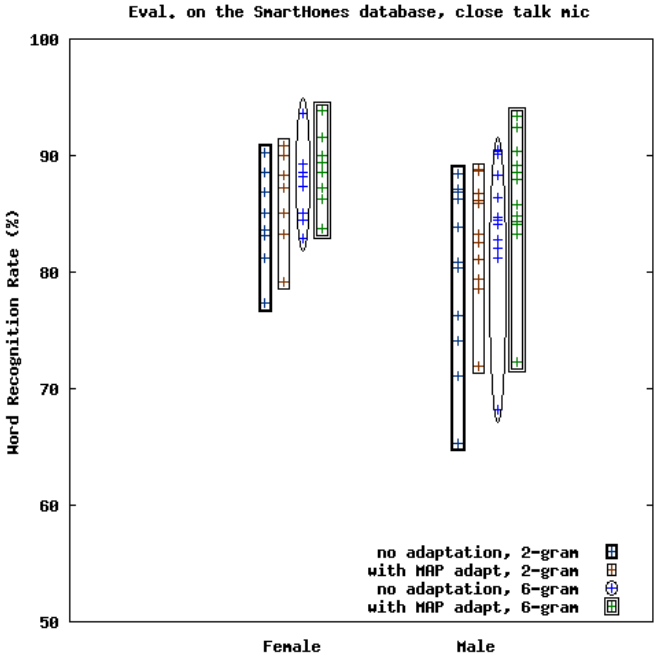

Fig 4. Évaluations de 20 locuteurs âgées (micro cravate).

6 Conclusions et Perspectives

L'infrastructure nécessaire à l'expérimentation d'un majordome mobile est en place. Elle utilise des briques logicielles libres aussi bien pour les télécommunications (PABX – Asterisk) et le traitement automatique de la parole (Julius).

Des résultats expérimentaux ont été obtenus en reconnaissance automatique de la parole sur des donnés enregistrées dans le cadre du projet CompanionAble.

Dans le cadre de vAssist, un *smartphone* (sous Android) sera utilisé. Le serveur Asterisk est opérationnel pour l'expérimentation de services liés aux scénarios d'usage listées en section 2.

Acknowledgments

La thèse de Daniel R.S.Caon est en partie financée par le projet européen CompanionAble (FP7/2007-2013, convention de subvention n° 216487).

References

1. Armstrong N., Nugent C., Moore G. and Finlay D., Using smartphones to address the needs of persons with Alzheimer's disease, Annales des Télécommunications, vol. 65, pp. 485-495 (2010);
2. https://www.asterisk.org/ ;
3. Bayeh, R. Reconnaissance de la Parole Multilingue: Adaptation de Modeles Acoustiques Multilingues vers une langue cible. Thése (Doctorat) — TELECOM Paristech, (2009);
4. Baldinger, J.-L., Boudy J., Dorizzi B., Levrey J.-P., Andreao R.V., Perpère C., Delavault F., Rocaries F., Dietrich C. and Lacombe A., Tele-surveillance system for patient at home: The mediville system. In: Miesenberger, K. et al. (Ed.). Computers Helping People with Special Needs. Springer Berlin / Heidelberg, Lecture Notes in Computer Science, v. 3118 p. 623–623. (2004) <http://dx.doi.org/10.1007/978-3-540-27817-7_59>.
5. Bush, Vannevar (1945). As We May Think. The Atlantic Monthly. Volume 176, No. 1; pages 101-108. (July, 1945);
6. Caon D.R.S., Amehraye A., Razik J., Chollet G., Andreao R.V., Mokbel C., Experiments on acoustic model supervised adaptation and evaluation by k-fold cross validation technique. In: ISIVC. 5th International Symposium on I/V Communications and Mobile Networks. Rabat, Morocco: IEEE, (2010);
7. Constantinescu, A.; Chollet, G. On cross-language experiments and data-driven units for alisp (automatic language independent speech processing). In: IEEE Workshop on Automatic Speech Recognition and Understanding. Santa Barbara, CA, USA: p. 606–613, (1997);
8. Jim Gemmell, Gordon Bell and Roger Lueder, MyLifeBits: a personal database for everything, Communications of the ACM, vol. 49, Issue 1 (Jan 2006), pp. 88.95. http://research.microsoft.com/en-us/projects/mylifebits/
9. Gitlin LN, Vause Earland T.. Améliorer la qualité de vie des personnes atteintes de démence: le rôle de l'approche non pharmacologique en réadaptation. In: JH Stone, M Blouin, editors.

International Encyclopedia of Rehabilitation, (2011). Available online: http://cirrie.buffalo.edu/encyclopedia/fr/article/28/

10. Lee, A.; Kawahara, T.; Shikano, K. In: EUROSPEECH. Julius - an open source real-time large vocabulary recognition engine. p. 1691–1694, (2001);
11. http://nomemoryspace.wordpress.com/2008/03/17/la-prothese-memorielle-limpact-de-la-publication-des-historiques-sur-la-societe-de-linternet/
12. Rigaud A.S. , Simonnet T. , Rialle V. , Rumenau P. , Vallet C. , Balginger J.-L., Belfeki I. , Boudy J. , Rotrou J. de , Sant' Anna M. de, Extra J., Faucounau V., Labouree F., Lacombe A., Orvoen G., Riguet M., Vella F., Vigouroux N., Wu Y.H. "Un exemple d'aide informatisée à domicile pour l'accompagnement de la maladie d'Alzheimer : le projet TANDEM", *NPG Neurologie - Psychiatrie - Gériatrie.* N°6, Vol.10, pp. 71-76, ISSN :1627-4830, LDAM Édition/Elsevier, ScienceDirect, (Avril 2010).
13. Schultz, T.; Katrin, K. Multilingual Speech Processing. Elsevier, (2006);
14. Young S., Evermann G., Gales M., Hain T., Kershaw D., Liu X. A., Moore G. , Odell J., Ollason D., Povey D., Valtchev V., and Woodland P. ,“The HTK Book (version 3.4)” Cambridge, UK, (2006).