Voix HD : un nouvel enjeu pour le traitement de la parole chez les personnes âgées

Anne Vanpé[1,2] *Hervé Provost*[1] *Nicolas Vuillerme*[2]
(1) Orange-Labs, BIZZ/VMC/SAM, Meylan
(2) AGIM, FRE 3405, CNRS-UJF-EPHE-UPMF
{anne.vanpe;herve.provost}@orange.com,Nicolas.Vuillerme@agim.eu

RESUME

L'analyse automatique de la parole représente un intérêt et un potentiel important pour le maintien à domicile des personnes âgées. Elle requiert actuellement l'utilisation de microphones permettant l'enregistrement d'un signal sonore de qualité, nécessaire à la gestion complexe des spécificités acoustiques de la voix des personnes âgées. Toutefois, cette méthodologie pose le problème du passage à l'échelle lors d'expérimentations. Nous pensons que l'utilisation du téléphone pourrait favoriser cette extension: plus de personnes touchées, des coûts réduits, l'automatisation des enregistrements et analyses possible… Cependant, s'il semble intéressant, le téléphone présente l'inconvénient de dégrader le signal audio.
Nous présentons ici la technologie *Voix HD*, nouvellement déployée par les principaux opérateurs de télécommunication, qui permet de lever ce verrou technologique. Assurant la transmission d'un signal audio de qualité, ce label pourrait constituer un outil efficace et approprié pour le traitement de la parole chez les personnes âgées.

ABSTRACT

***HD Voice* : a new issue for voice processing in elderly**

Use of automatic speech analysis is an interest and a great comer for home care of elderly. At the moment, it needs the use of microphone for recording a quality sound signal that is necessary for the complex processing of acoustic specificities of elderly voice. However, this methodology prevents experiments spreading. We think that the use of phone would be able to help this scaling up: more tested persons, less costs, possible automation of recordings and analysis… Phone seems to be interesting, but decreases audio signal quality.
We present here the HD Voice technology, newly supported by the main telcos, that allows to remove this technological bottle-neck. Ensuring a phone transmission of high quality audio signal, this seal of approval could constitute an efficient and suitable tool for speech processing in elderly people.

MOTS-CLES : téléphone, *Voix HD*, traitement de la parole, personnes âgées, maintien à domicile, gérontechnologie.
KEYWORDS: telephone, *HD Voice*, automatic speech processing, elderly people, home care, gerontechnology.

1 Introduction

L'analyse automatique de la parole représente un intérêt et un potentiel important pour le maintien à domicile des personnes âgées. En effet, la raison même de la proposition de

cet atelier réside dans le vieillissement croissant de la population, couplé au manque de place dans les institutions spécialisées destinées aux séniors. Le maintien à domicile représente ainsi un enjeu sociétal actuel majeur, qui nécessite le développement de nouvelles technologies améliorant le confort, le bien-être et la sécurité des personnes âgées vivant à domicile. Parmi ces technologies, certaines cherchent à exploiter la parole de ces personnes, à des fins d'amélioration du confort d'interaction au sein d'habitats intelligents, ou encore pour évaluer leur santé et éventuellement détecter certains symptômes annonciateurs d'une maladie de type dégénérative (Vacher, 2011).

Ce type d'étude en traitement de la parole nécessite actuellement l'utilisation de microphones permettant l'enregistrement des signaux vocaux de bonne qualité. Cette qualité de signal est d'autant plus indispensable que la voix des personnes âgées possède des spécificités acoustiques qui rendent les post-traitements des enregistrements vocaux (analyse de certains paramètres ou reconnaissance vocale) intrinsèquement complexes. Toutefois, le microphone *per se* est une contrainte qui gène les expérimentations pour passer à l'échelle.

Nous rappellerons d'abord en quoi le traitement de la parole représente l'un des enjeux du maintien à domicile (partie 2), avant de présenter en quoi un outil tel que le téléphone peut être intéressant méthodologiquement pour le domaine du traitement de la parole chez les personnes âgées (partie 3), en particulier s'il utilise la technologie *Voix HD*. Nous présenterons ainsi cette technologie (partie 4), puis en mentionnerons certaines perspectives (partie 5).

2 Le traitement de la parole : un des enjeux du maintien à domicile

Face à une population vieillissante qui exprime sa large préférence pour rester à domicile, le traitement du son, et particulièrement de la parole, sont devenus de nouveaux enjeux pour les technologies destinées à favoriser le maintien à domicile. Entre autres, « assurer une assistance domotique par une interaction naturelle (avec commandes vocales et tactile) » et « apporter plus de sécurité par la détection de situations de détresse ou d'effraction » sont par exemple deux des objectifs du projet Sweet Home[1] (mené par le laboratoire LIG, en collaboration avec d'autres partenaires).

Au sein de ce projet, des études ont mis en évidence les difficultés à surmonter dans ce domaine. En effet, le traitement de la parole peut permettre de mettre en place une interaction facile naturelle entre personnes âgées et technologies (*e.g.* Kumiko et al., 2004), reconnaître des appels de détresse (Vacher, 2011) ou encore détecter certains symptômes annonciateurs d'une maladie de type dégénérative (Lee et al., 2011).

Toutefois, certaines contraintes technologiques, ainsi que les spécificités acoustiques de la voix des personnes âgées, complexifient considérablement cette tâche (Vacher et al., 2010). D'une part, la nature des technologies utilisant la voix pour le maintien à domicile, notamment dans le cadre des habitats intelligents, nécessite souvent un traitement de la parole distante, bruitée et multi-source, et dans la majorité des cas, un

[1] Citations issues du site officiel du projet : http://sweet-home.imag.fr/index.php?choix=projet.

système robuste et fiable (notamment pour les systèmes liés à la santé ou la sécurité de la personne).

D'autre part, la voix des personnes âgées et, plus globalement, leur manière de parler, présente un certain nombre de particularités, liées aux changements physiologiques progressifs liés à la vieillesse, ou à leur perte de capacités cognitives et de contrôle moteur (*e.g.* Wilpon et Jacobsen, 1996 ; Linville, 1996 et 2002 ; Zellner-Keller, 2006 ; Gorham-Rowan et Laures-Gore, 2006 ; Hooper et Craidis, 2009 ; Vipperla, 2009) : hypo-articulation, taux de parole plus lent que chez les adultes actifs, F0 plus basse chez les femmes, *jitter* et *shimmer* plus élevés, intensité plus faible, diminution globale de l'énergie, augmentation du bruit, organisation temporelle de la parole différente, ou encore syntaxe et vocabulaire plus simples.

Lors des études portant sur la parole des personnes âgées, l'utilisation de microphones de bonne qualité est ainsi indispensable dans la majorité des cas (*e.g.* Vacher, 2011). Cette méthodologie requière l'acquisition de données vocales dans les meilleures conditions d'enregistrement possibles. Cela passe souvent par des interviews en face à face, de manière à contrôler au maximum le contenu et les modalités de l'enregistrement. Leur inconvénient est en particulier d'être coûteux en temps et en argent, ce qui rend difficile le passage à l'échelle de ces études, que cela soit lié à l'effectif ou à une répartition géographique large, de manière à obtenir de gros corpus de voix.

3 L'utilisation du téléphone pour passer à l'échelle lors des expérimentations

Nous pensons que l'utilisation du téléphone pourrait être une alternative intéressante à l'utilisation des microphones. D'un point de vue méthodologique, l'utilisation de cet outil qu'est le téléphone pourrait permettre aux études de passer à l'échelle : plus de personnes touchées, des coûts réduits, l'automatisation des enregistrements et des analyses possible, etc. Cependant, il a l'inconvénient de dégrader le signal audio.

Morano et Stern (1994) et Reynolds et al. (1995) ont testé des systèmes de reconnaissance de la parole et d'identification du locuteur sur des signaux vocaux téléphoniques. Il en est ressorti que la performance de ces derniers diminue avec les enregistrements téléphoniques (vs. enregistrements de haute qualité). Ils précisent que les principales pertes d'information sont dues à la bande passante limitée, à la fréquence d'échantillonnage moins élevée et au bruit supplémentaire.

En parallèle, l'intérêt de cette méthodologie et de ses limites a également été mis en évidence dans le domaine connexe de la détection de pathologies à travers la voix : certaines études ont ainsi utilisé le téléphone pour leurs expérimentations ou comme finalité technologique. Par exemple, Moran et al. (2006) ont évalué les dégradations acoustiques dues au téléphone dans un système de classification automatique des pathologies de la voix, et en particulier du larynx *(e.g.* dysphonies, lésions, nodules, etc.), cela à partir de vocalisations maintenues de [a]. Ils ont montré que 14% de la diminution de performance de leur système de classification était due aux mêmes paramètres que ceux relevés par les études précédentes de Morano et Stern (1994) et Reynolds et al. (1995).

Quant à Mundt et al. (2007), dans le cadre de *Healthcare Technology Systems* (Inc)[2], ils ont relevé l'influence de l'utilisation du téléphone pour l'enregistrement de données à analyser dans le cadre d'une technologie destinée à la détection automatique de la gravité de la dépression. Ils ont trouvé une différence significative entre les données obtenues avec l'utilisation d'un téléphone standard (RTC ou GSM au choix du sujet), par rapport à l'utilisation d'un téléphone fixe RNIS -Réseau Numérique à Intégration de Services- (téléphone numérique, signal codé à 64ko/s). En effet, avec le téléphone standard, les temps de vocalisations, les durées d'enregistrement total et les mesures des pauses sont significativement plus variables, et les intensités du signal sont plus faibles et plus variables. Cela semble affecter la qualité des données vocales recueillies et, en conséquence, la fiabilité et la validité de leurs analyses.

Nous présentons dans la partie suivante la technologie *Voix HD*, nouvellement déployée par les principaux opérateurs de télécommunication, qui pourrait permettre de lever ce verrou technologique.

4 La technologie *Voix HD*

D'un point de vue fonctionnel, la technologie *Voix HD* (voix Haute Définition, ou « voix en bande élargie ») augmente le confort et l'efficacité de la communication par la transmission d'un signal audio de qualité. En téléphonie, la qualité des signaux de parole transportés sur les réseaux de télécommunication est liée :

- au terminal téléphonique lui-même (qualité des écouteurs et du microphone) ;
- aux codecs qui numérisent les signaux et aux réseaux entre l'émetteur et le récepteur de l'appel, influant par exemple sur la fréquence d'échantillonnage, la bande passante et le débit ;
- aux traitements éventuels de correction des défauts (notamment contre le bruit et l'écho).

Dans le cas de cette technologie, la transmission d'une « voix Haute Définition » est possible par la combinaison :

- d'un ensemble de contraintes sur les caractéristiques acoustiques des téléphones concernés (concernant écouteur et microphone, ainsi que la compatibilité avec le codage/décodage d'un signal de bonne qualité) ;
- de l'utilisation du Codeur AMR-WB (*Adaptative Multi-Rate – Wide Band*[3]) ;
- de l'utilisation d'un réseau offrant une QoS (*Quality of Service*) garantie en termes de performance du transport et de disponibilité du service ; et
- de l'utilisation de technologies telles que les systèmes anti-écho et d'atténuation du bruit[4].

Cela implique, à l'heure actuelle, que l'émetteur comme le récepteur de l'appel possèdent un terminal mobile compatible avec la *Voix HD* et utilise le réseau mobile 3G pour avoir une qualité de signal optimale.

[2] http://www.healthtechsys.com/

[3] C'est-à-dire codeur adaptatif multi-débits (ici à large bande).

[4] Ces systèmes sont généralement connus en tant que VQE (*Voice Quality Enhancement*).

La technologie *Voix HD* est plus précisément une implémentation de protocoles de communication (qui nécessite actuellement la disponibilité du réseau 3G), qui correspondent à la norme de compression audio ITU-T G.722.2[5] (également normalisé par l'ETSI sous le nom « Codeur AMR-WB »- voir ci-dessus).

Concernant le traitement acoustique de la parole, l'amélioration des valeurs de paramètres acoustiques susceptibles d'être les plus intéressants sont la fréquence d'échantillonnage et la bande passante (Table 1), d'autant plus s'ils sont couplés à un système anti-écho et à une atténuation du bruit (Rodman, 2003 ; GSMAssociation, 2011).

Paramètres	**Téléphone classique**	**Téléphone avec *Voix HD***
Échantillonnage	8 000 Hz	16 000 Hz
Bande Passante	300 à 3400 Hz	50 à 7000 Hz

TABLE 1 – Comparaison de certains paramètres du signal, avec ou sans *Voix HD*.

Cette technologie, développée depuis de nombreuses années, est de plus en plus intégrée aux terminaux téléphoniques. Elle est de surcroît appuyée par les principaux opérateurs de télécommunication, ce qui permet un large déploiement.

Des études clients d'Orange France ont montré un taux de satisfaction de 96% concernant l'utilisation de cette technologie (les trois-quarts des testeurs étant prêts à changer de téléphone pour bénéficier de *Voix HD* (GSMAssociation, 2011)). Si le confort de communication est déjà apprécié par les utilisateurs, la qualité du signal audio pourrait également permettre aux chercheurs en traitement de la parole de bénéficier de cet apport technologique.

5 *Voix HD* : des perspectives prometteuses

Dans le cadre du traitement de la parole pour le maintien à domicile, nous avons identifié une difficulté des expérimentations concernant leur passage à l'échelle. Elle est entre autres liée à la nécessaire utilisation de microphones de qualité. L'alternative du téléphone pour ce passage à l'échelle n'était jusqu'alors pas satisfaisant dans ce cadre, en raison de la forte dégradation du signal acoustique alors enregistré.

La technologie *Voix HD*, en pleine expansion actuellement grâce notamment au soutien des principaux opérateurs de télécommunication, pourrait permettre de lever ce verrou technologique. Elle assure la transmission d'un signal audio de qualité, grâce notamment à une bande de fréquence élargie, un système anti-écho et une atténuation du bruit.

Ainsi, cette technologie pourrait constituer un outil efficace et approprié pour le traitement de la parole chez les personnes âgées.

[5] Cf. Page officielle concernant la norme : http://www.itu.int/rec/T-REC-G.722.2/fr.

Références

GORHAM-ROWAN, M. et LAURES-GORE, J. (2006). Acoustic-perceptual correlates of voice quality in elderly men and women. *In Journal of Communication Disorders*, 39, pages 171–184.

HOOPER, C. R. et CRAIDIS, A. (2009). Normal Changes in the Speech of Older Adults : You've still got what it takes ; it just takes a little longer! *In Perspectives on Gerontology*, 14.

KUMIKO, O., MITSUHIRO, M., ATSUSHI, E., SHOHEI, S. et REIKO, T. (2004). Input support for elderly people using speech recognition. *In IEIC Technical Report*, 104(139), pages 1–6.

LEE, H.R., GAYRAUD, F., HIRSCH, F., et BARKAT-DEFRADAS, M. (2011). Speech dysfluencies in normal and pathological aging : a comparison between Alzheimer patients and healthy eldery subjects. *In the 17th International Congress of Phonetic Sciences (ICPhS)*, Hong-Kong, pages 1174-1177.

LINVILLE, S.E. (1996). The sound of senescence. *In Journal of Voice*, 10(2), pages 190-200.

LINVILLE, S.E. (2002). Source characteristics of aged voice assessed from long-term average spectra. *In Journal of Voice*, 16(4), pages 472-479.

MORAN, R.J., REILLY, R.B. (2006). Telephony-Based Voice Pathology Assessment Using Automated Speech Analysis. *In IEEE Transactions on Biomedical Engineering*, 53(3), pages 468 – 477.

MORENO, P.J. et STERN, R.M. (1994). Sources of degradation of speech recognition in the telephone network, *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-94* , vol.1, Adelaide, Australia, Apr 1994, pages 109-112.

MUNDT, J.C., SNYDER, P.J., CANNIZZARO, M.S., CHAPPIE, K., et GERALTS, D.S. (2007). Voice acoustic measures of depression severity and treatment response collected via interactive voice response (IVR) technology. *In Journal of Neurolinguistics*, 20(1), pages 50-64.

GSMAssociation (2011). AMR-WB White Paper. By NTT DoCoMo, FT Group, DT, Ericsson, et Nokia.

PORTET, F., VACHER, M., GOLANSKI, C., ROUX, C. et MEILLON, B. (2011). Design and evaluation of a smart home voice interface for the elderly – Acceptability and objection aspects. *In Personal and Ubiquitous Computing Journal* (accepted).

REYNOLDS, D.A., ZISSMAN, M.A., QUATIERI, T.F., O'LEARY, G.C. et CARLSON, B.A. (1995). The effects of telephone transmission degradations on speaker recognition performance. *In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95,* vol.1, Detroit, MI, 9-12 May 1995, pages 329-332.

RODMAN, J. (2003). The effect of bandwidth on speech intelligibility. White paper, POLYCOM Inc., USA.

VACHER, M., FLEURY, A., PORTET, F., SERIGNAT, J.F., et NOURY, N. (2010). Complete Sound and Speech Recognition System for Health Smart Homes: Application to the Recognition of Activities of Daily Living. *In New Developments in Biomedical Engeneering*, Domenico

Campolo (Ed.), pages 645-673.

VACHER, M. (2011). Analyse sonore et multimodale dans le domaine de l'assistance à domicile. Mémoire d'HDR, Spécialité Informatique et Mathématiques Appliquées, Université de Grenoble.

VIPPERLA, R., WOLTERS, M., GEORGILA, K., AND RENALS, S. (2009). Speech input from older users in smart environments : challenges and perspectives. *In Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction. Part II: Intelligent and Ubiquitous Interaction Environments, UAHCI '09*, Berlin, pages 117–126.

WILPON, J. et JACOBSEN, C. (1996). A study of speech recognition for children and the elderly, *In IEEE Int. Conference on Acoustics, Speech and Signal Processing*, pages 349–352.

ZELLNER KELLER, B. (2006). Ageing and Speech Prosody. *In Speech Prosody 2006,* R. Hoffmann & H. Mixdorff (Eds.), pages 696-701.