Reconnaissance automatique de la parole dans les habitats intelligents : application à l'assistance à domicile

Michel Vacher, François Portet, Frédéric Aman, Benjamin Lecouteux, Solange Rossato, Véronique Auberge

Reconnaissance automatique de la parole dans les habitats intelligents : application à l'assistance à domicile

Michel Vacher, François Portet, Frédéric Aman, Benjamin Lecouteux, Solange Rossato, Véronique Aubergé
CNRS, LIG, F-38000 Grenoble, France
Univ.Grenoble Alpes, LIG, F-38000 Grenoble, France
Laboratoire d'Informatique de Grenoble équipe GETALP
41, rue des Mathématiques, BP 53, 38041 Grenoble Cedex9
Email : Michel.Vacher@imag.fr

***Résumé*—Pour relever le défi du maintien à domicile de la population vieillissante, l'enjeu est d'améliorer la vie quotidienne des personnes âgées afin qu'elles soient toujours maitresses de leurs choix et qu'elles utilisent la technologie pour continuer à vivre de manière autonome, à apprendre et à s'investir dans la vie sociale. Les technologies du traitement de la parole peuvent apporter une contribution importante. Dans cet article, nous dressons un tour d'horizon des verrous et écueils techniques à résoudre pour les rendre pleinement opérationnelles avant de décrire les travaux que nous avons accomplis et qui ont permis d'appliquer ces techniques à la commande vocale dans un appartement intelligent équipé de matériel domotique et de les évaluer avec des utilisateurs âgés ou malvoyants.**

Keywords*—*reconnaissance automatique de la parole ; habitat intelligent ; voix âgée

I. Introduction

Les pays industrialisés sont confrontés aux défis imposés par le vieillissement de la population, qui est corrélé à l'augmentation du handicap, car il devient urgent de proposer des solutions pour améliorer la vie quotidienne des personnes âgées afin qu'elles soient toujours maitresses de leurs choix et qu'elles utilisent la technologie pour continuer à vivre de manière autonome, à apprendre et à participer à la vie sociale. Un des souhaits des personnes âgées est de continuer à vivre confortablement dans leur propre maison en toute sécurité même si leur autonomie diminue. L'assistance à la vie autonome ou *Ambient Assisted Living* (AAL) permet d'anticiper et de répondre aux besoins des personnes en perte d'autonomie avec les Technologies de l'Information et de la Communication.

Dans ce domaine, le développement des maisons intelligentes est considéré comme un moyen prometteur pour apporter une assistance à domicile [1]. Ce nouveau concept englobe la domotique qui est l'ensemble des techniques utilisées dans le bâtiment permettant de centraliser et piloter les différents systèmes nécessaires pour le confort, la sécurité ou les communications. De nos jours, l'une des meilleures interfaces est l'interface utilisateur vocale (VUI), dont la technologie est mature et qui permet une interaction en langage naturel afin que l'utilisateur n'ait pas besoin d'apprendre des procédures informatiques complexes [2]. Cette interface est aussi bien adaptée aux différents profils d'utilisateurs (personnes à mobilité réduite ou à faible compétence technique) et à des situations d'urgence (mains libre et interaction à distance). Actuellement, un nombre croissant de maisons intelligentes et de projets prennent en considération la Reconnaissance Automatique de la Parole (RAP) dans leur conception.

Dans cet article, après un bref état de l'art en section II, nous allons présenter en section III des résultats du projet SWEET-HOME en ce qui concerne la commande vocale de la domotique ainsi que ceux du projet CIRDO sur la détection des appels à l'aide dans la voix âgée en section IV avant de conclure.

II. Les études en cours et les défis à relever

Les projets consacrés à l'utilisation des technologies de reconnaissance de la parole sont assez nombreux, comme les projets COMPANIONS [3], CompanionAble [4], ALADIN [5] ou PIPIN [6]. Cependant, il subsiste beaucoup d'interrogations sur l'impact réel tant du point de vue assistance que coût d'installation malgré des évaluations réelles dans des appartements de test avec les projets PERS [7], DIRHA [8], car peu ont été évalués avec des personnes âgées comme dans SWEET-HOME [9] et à notre connaissance aucun n'a été mis en place dans leur habitat.

En effet, si les microphones sont positionnés de manière fixe dans l'appartement pour éviter l'usage d'une télécommande manipulée par la personne, la reconnaissance vocale sera rendue plus difficile à cause des conditions distantes [10] où la réverbération et l'acoustique de la pièce interviennent. Par ailleurs, l'environnement sonore sera lui-même perturbé par des sons non-langagiers [11]. Il faudra donc pour chaque enregistrement être en mesure de distinguer entre son et parole, seule cette dernière devant être traitée par le système de RAP. Dans le cas où la parole est mélangée à un bruit de l'appartement, il faut de plus séparer la parole du bruit [12] ce qui nécessite des traitements complexes.

Il convient par ailleurs d'adapter les systèmes de RAP à l'application visée. Cela concerne principalement le modèle acoustique (MA) utilisé. La voix âgée est affectée par une production moins précise des consonnes, des tremblements,

Les travaux présentés ont été financés par les projets SWEET-HOME (ANR-2009-VERS–011) et CIRDO (ANR-2010-TECS-012)

des hésitations et une articulation plus lente [13], certaines dégénérescences du conduit vocal, des cordes vocales et des muscles du larynx ont aussi une influence [14]. En conséquence, les performances des systèmes de RAP usuels indépendants du locuteur décroissent, ce qui a été constaté pour différentes langues [15] [16].

III. Commande vocale de la domotique

A. L'approche Sweet-Home

L'hypothèse de base est que le système sera utilisé lorsqu'une seule personne est présente dans l'appartement, son utilisation serait moins pertinente dans le cas contraire. Le système doit permettre d'améliorer le confort de la personne grâce à une commande distante très utile pour les personnes ayant des difficultés à se déplacer. Il doit aussi être capable d'agir en cas de situation à risque, soit lors d'un appel explicite à l'aide, soit lorsque le système détecte lui même une situation à risque.

La mise en place d'une analyse sonore permettant la commande vocale de la domotique dans un habitat intelligent nécessite d'être capable de discriminer entre son et parole, seules les paroles devant être traitées par le système de RAP. Le couplage entre le système de RAP et la domotique permettra une commande vocale en s'appuyant sur la connaissance du « contexte » nécessaire pour une prise de décision à partir de données incertaines pour comprendre l'implicite contenu dans le langage naturel. Citons comme exemple la commande « Allume la lumière » qui ne précise ni quelle lampe éclairer ni l'intensité souhaitée.

B. Le système mis en œuvre

Le système Sweet-Home vise à permettre une interaction en contexte par un contrôle intelligent de la domotique s'appuyant sur la commande vocale et la détection de situations spécifiques, il a été décrit dans [9].

Les différents constituants interagissent entre eux comme montré sur la figure 1. Le cœur du dispositif est le contrôleur intelligent qui est en charge de prendre des décisions à chaque instant à partir des informations qu'il reçoit du réseau domotique et du système d'analyse sonore PATSH et qui lui permettent d'inférer le contexte (localisation et activité de la personne dans l'appartement). Une décision peut être soit consécutive à un ordre détecté au niveau de la reconnaissance de la parole (par exemple : « Nestor allume la lumière »), soit suite à la détection par le contrôleur lui-même d'une situation critique (par exemple : la porte d'entrée est restée ouverte). Il peut alors envoyer un ordre à un actionneur domotique, émettre une alarme vocale ou mettre l'utilisateur en liaison avec un interlocuteur extérieur par exemple avec un dispositif comme *e-lio* qui permet notamment une communication de type vidéoconférence entre la personne et sa famille [1].

Un ensemble de 7 à 8 microphones placés au plafond dans les différentes pièces de l'appartement permet au système d'analyse sonore PATSH d'enregistrer au fil de l'eau les commandes prononcées. Le système sélectionne les phrases ayant le meilleur rapport signal sur bruit (RSB) pour les

1. http ://www.technosens.fr/

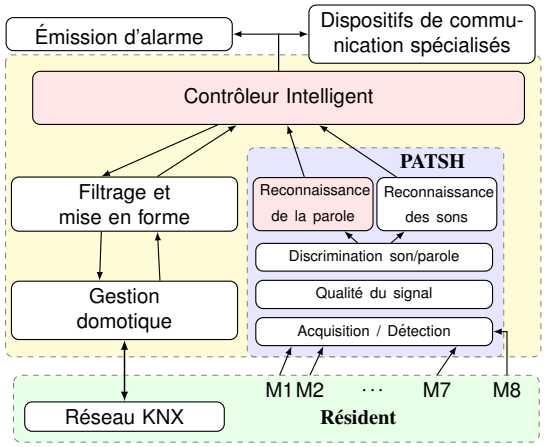

Figure 1. Le système Sweet-Home

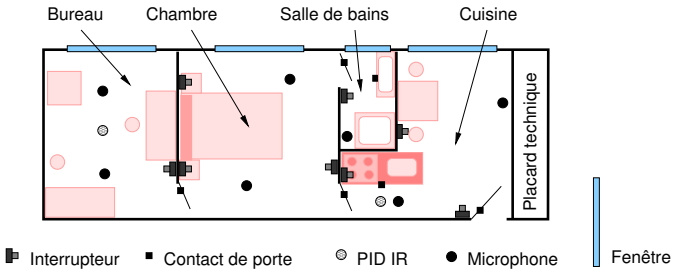

Figure 2. Implémentation des microphones et des capteurs dans le bâtiment intelligent Domus

envoyer au système de reconnaissance de la parole (RAP). Le contrôleur intelligent filtre ensuite l'hypothèse reconnue pour détecter la présence éventuelle d'un ordre domotique ou d'un appel de détresse.

L'appartement de test utilisé est l'habitat intelligent Domus du laboratoire LIG qui est un appartement complètement fonctionnel de 35m^2 composé d'une cuisine, d'une chambre, d'une salle de bain et d'un bureau. La figure 2 montre la position des microphones dans les différentes pièces. L'appartement est aussi équipé de plus de 150 capteurs domotiques reliés par un réseau domotique.

C. Résultats du projet

L'inclusion des utilisateurs finaux a révélé l'intérêt de l'approche et des perspectives de recherche (études d'usage et validation finale) [2]. Le logiciel d'analyse sonore en temps réel PATSH (Plateforme d'Accueil de Traitement Sweet-Home) a permis d'intégrer les modules d'analyse sonore sur une maquette fonctionnant dans un habitat réel. Le contrôleur intelligent s'est montré efficace pour le contrôle en contexte de la domotique. Les évaluations ont été réalisées dans l'appartement Domus avec des utilisateurs jouant des scénarios reproduisant des activités de la vie quotidienne (AVQ ou ADL), soit 25 personnes adultes jeunes et 11 personnes âgées et/ou malvoyantes. Enfin, les précieux corpus multimodaux (parole, son, capteurs domotique) permettront d'alimenter des recherches ultérieures [17]. Ils ont été utilisés récemment pour développer un nouveau système de reconnaissance qui a permis de baisser le taux d'erreur de détection des ordres domotiques de 41% à 13% [18].

IV. Etude de la reconnaissance automatique de la voix âgée

Le projet CIRDO [2], porté par le LIRIS, vise à la réalisation d'un « Compagnon Intelligent Réagissant au Doigt et à l'Œil » pour détecter les situations d'urgence et de détresse auxquelles peut être confrontée une personne âgée vivant seule à domicile [19]. Ce système utilise des données audio et vidéo, le traitement de ces dernières ne fait pas partie de cet article et nous ne traiterons ici que de ce qui concerne le traitement automatique de la parole. En ce qui concerne l'analyse et la reconnaissance de la parole dans l'application envisagée, il s'agit d'analyser de manière automatique la voix d'une personne qui non seulement sera âgée mais aussi fortement émue. Ce sont deux paramètres rarement abordés dans les études de reconnaissance automatique de la parole et dont on sait qu'ils perturbent fortement les systèmes de RAP. Dans ce projet, la reconnaissance de parole permettra d'extraire les appels à l'aide lorsque la personne ne peut plus se déplacer ou se sent en danger.

A. Les corpus enregistrés

L'étude nécessite l'utilisation de corpus de parole particuliers à ce contexte : type de phrases prononcées, âge et état émotionnel du locuteur. Ce genre de corpus étant inexistant, nous nous sommes heurtés à certains problèmes pour recueillir ce corpus : difficultés dans le recrutement de personnes âgées volontaires, contraintes dues à la fatigue rapide de ces personnes. Nous avons à l'heure actuelle enregistré un corpus de parole lue aussi bien par des personnes âgées que des personnes jeunes (*AD80*), un corpus de parole spontanée avec 22 personnes âgées (*ERES38*) et un corpus (*VD*) de voix émues.

1) Anodin-Détresse80: Le corpus *AD80* est composé d'un total de 95 locuteurs (59 femmes et 36 hommes). Le groupe *locuteurs âgés* est composé de 43 locuteurs (32 femmes et 11 hommes) âgés de 62 à 94 ans, qui ont lu au total 2663 phrases de détresse (52 minutes) et 434 phrases d'appels aux aidants (16 minutes), et 3006 phrases anodines (1 heure et 6 minutes). Le groupe *locuteurs jeunes* est composé de 52 locuteurs (27 femmes et 25 hommes) âgés de 18 à 64 ans, qui ont lu au total 3306 phrases de détresse (56 minutes), 961 phrases d'appels aux aidants (26 minutes), et 3897 phrases anodines (1 heure et 12 minutes). Un locuteur de 62 ans, bien qu'il soit sous notre seuil de personnes âgées fixé à 65 ans, a été placé dans le groupe de personnes âgées du fait de sa présence en maison de retraite, sa perte d'autonomie et son vieillissement physique avancé. Finalement, le corpus *AD80* est constitué de 14 267 phrases annotées, avec 4 heures et 49 minutes d'enregistrements.

Dix exemples de phrases de détresse, d'appels aux aidants et de phrases anodines sont donnés en table I.

2) ERES38: Le corpus *ERES38* (Entretiens RESidences 38) a pour but d'être utilisé pour l'adaptation des modèles acoustiques à la voix des personnes âgées et pour étudier les caractéristiques de la voix des personnes âgées. Ce corpus a été enregistré en 2011 dans le lieu de vie des personnes âgées volontaires, qui étaient résidentes de structures spécifiques pour personnes âgées : foyers logements ou maisons de retraite dans l'agglomération grenobloise.

Au final, le corpus *ERES38* a été acquis auprès de 23 personnes âgées (16 femmes et 7 hommes) de 68 à 98 ans. Le corpus inclut 48 minutes de lectures annotées et 16 heures et 56 minutes d'interviews.

3) Voix-Détresse: Ce corpus a été enregistré en 2013 et 2014 au laboratoire. Il a été demandé aux locuteurs de lire 20 phrases de détresse de façon neutre. Puis, nous avons associé à chaque phrase une photo représentant une situation montrant un personnage en état de détresse, et avons demandé aux locuteurs de se mettre dans la peau des personnages et d'énoncer les phrases de façon très expressive. Les émotions recherchées était principalement les émotions négatives telles que la peur, la colère et la tristesse.

Nous avons constitué deux groupes, le groupe *locuteurs jeunes* (20 participants) et le groupe *locuteurs âgés* (5 participants).

Au total, 1742 phrases ont été prononcées (521 phrases neutres et 1221 phrases émues), soit 28 minutes d'enregistrement (7 minutes et 30 secondes pour les phrases neutres et 20 minutes et 30 secondes pour les phrases émues).

TABLE II. *Taux d'erreur de mots moyens pour les différents groupes (29 locuteurs au total) en fonction des modèles acoustiques (MA) standards ou adaptés au locuteur*

Groupe	MA standard	MA adapté
Voix hommes âgés (8 loc.)	63,65%	22,80%
Voix femmes âgées (21 loc.)	36,74%	15,21%
Moyenne (29 loc.)	44,17%	17,30%

B. Les résultats

Les performances de reconnaissance automatique obtenues avec des systèmes utilisant un modèle acoustique standard, c'est à dire appris sur des voix adultes jeunes, montrent un dégradation forte des performances pour les locuteurs âgés (cf. table II). Une analyse plus complète des résultats montre par ailleurs que la dispersion des résultats et bien plus grande dans le cas des locuteurs âgés [20], que les performances décroissent lorsque le score GIR diminue. En effet, le niveau de dépendance reflète le niveau réel de vieillissement physique et psychique de la personne, ce vieillissement influençant la production de la parole. D'une manière générale, les phonèmes nécessitant plus d'énergie lors de leur production sont les plus mal reconnus. Cependant, l'adaptation des modèles acoustiques par des méthodes classiques comme MLLR permet d'améliorer les performances, sauf dans le cas des locuteurs obtenant les résultats de reconnaissance les plus faibles.

Nous avons mis en évidence une augmentation des taux d'erreur dans le cas des voix âgées [21]. Ceci a été confirmé par des études plus complètes et encore en cours utilisant le corpus *VD* qui ont aussi permis de conclure que l'adaptation des modèles acoustiques avec des données de paroles émues permettait d'atténuer la perte de performance.

V. Conclusion

Dans cet article nous avons présenté deux projets visant à utiliser la reconnaissance de parole pour assister les personnes

2. http ://liris.cnrs.fr/cirdo/

TABLE I. *Exemples de phrases du corpus AD80.*

Phrases de détresse	Appels aux aidants	Phrases anodines
Aidez-moi !	e-lio appelle le SAMU !	Bonjour madame !
Au secours !	e-lio appelle les pompiers !	Ça va très bien.
Je me sens mal !	e-lio appelle les secours !	Ce livre est intéressant.
Je suis tombé !	e-lio appelle un docteur !	Il fait soleil.
Du secours s'il vous plaît !	e-lio appelle une ambulance !	J'ai ouvert la porte.
Je ne peux plus bouger !	e-lio appelle une infirmière !	Je dois prendre mon médicament !
Je ne suis pas bien !	e-lio appelle ma fille !	J'allume la lumière !
Je suis blessé !	e-lio appelle mon fils !	Je me suis endormi tout de suite !
Je ne peux pas me relever !	e-lio tu peux téléphoner au samu ?	Le café est brûlant !
Ma jambe ne me porte plus !	e-lio il faut appeler les secours !	Où sont mes lunettes ?

à leur domicile. Le projet CIRDO a permis de confirmer que les système de reconnaissance classiques ne sont pas adaptés à la voix âgée et de montrer qu'il était possible de développer un système de reconnaissance adapté à la voix âgée.

Le projet SWEET-HOME a permis l'évaluation d'un système de commande vocale de la domotique dans un appartement par des personnes jeunes ou âgées dont certaines étaient malvoyantes qui effectuaient des activités de la vie quotidienne. Ces personnes ont apprécié l'utilisation du système malgré ses imperfections et son temps de réaction, elles ont en particulier regretté que le système ne donne pas de retour quand il est en train de traiter leur demande. Les personnes âgées veulent conserver leur autonomie et craignent de chuter tandis que les personnes malvoyantes souhaitent recevoir des informations sur l'état courant de leur environnement de vie.

REMERCIEMENTS

Les auteurs remercient toutes les personnes qui ont accepté de participer aux différents enregistrements et aux expérimentations.

RÉFÉRENCES

[1] M. Chan, D. Estève, C. Escriba, and E. Campo, "A review of smart homes- present state and future challenges," *Computer Methods and Programs in Biomedicine*, vol. 91, no. 1, pp. 55–81, 2008.

[2] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly – Acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 127–144, 2013, impact-F=1.13 estim. in 2012.

[3] M. Cavazza, R. S. de la Camara, and M. Turunen, "How was your day ? : a companion eca," in *AAMAS*, 2010, pp. 1629–1630.

[4] J. Rougui, D. Istrate, and W. Souidene, "Audio sound event identification for distress situations and context awareness," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, Minneapolis, USA, 2009, pp. 3501–3504.

[5] J. F. Gemmeke, B. Ons, N. Tessema, H. Van Hamme, J. Van De Loo, G. De Pauw, W. Daelemans, J. Huyghe, J. Derboven, L. Vuegen, B. Van Den Broeck, P. Karsmakers, and B. Vanrumste, "Self-taught assistive vocal interfaces : an overview of the aladin project," in *Interspeech 2013*, 2013, pp. 2039–2043.

[6] I. Casanueva, H. Christensen, T. Hain, and P. Green, "Adaptive speech recognition and dialogue management for users with speech disorders," in *Proceedings of Interspeech 2014*, Singapore, 2014, pp. 1033–1037.

[7] M. Hamill, V. Young, J. Boger, and A. Mihailidis, "Development of an automated speech recognition interface for personal emergency response systems," *Journal of NeuroEngineering and Rehabilitation*, vol. 6, no. 1, pp. 1–26, 2009.

[8] M. Ravanelli and M. Omologo, "On the selection of the impulse responses for distant-speech recognition based on contaminated speech training," in *Proceedings of Interspeech 2014*, Singapore, 2014, pp. 1028–1032.

[9] M. Vacher, P. Chahuara, B. Lecouteux, D. Istrate, F. Portet, T. Joubert, M. E. A. SEHILI, B. Meillon, N. Bonnefond, S. Fabre, C. Roux, and S. Caffiau, "The SWEET-HOME Project : Audio Technology in Smart Homes to improve Well-being and Reliance," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC'13)*, Osaka, Japan, Jul. 2013, pp. 7298–7301. [Online]. Available : http ://hal.archives-ouvertes.fr/hal-00962222

[10] M. Wölfel and J. W. McDonough, *Distant Speech Recognition*. New York : Wiley, 2009.

[11] M. Vacher, F. Portet, A. Fleury, and N. Noury, "Development of Audio Sensing Technology for Ambient Assisted Living : Applications and Challenges," *International journal of E-Health and medical communications*, vol. 2, no. 1, pp. 35–54, Mar. 2011.

[12] J. Barker, H. Christensen, N. Ma, P. Green, and E. Vincent, "The PASCAL 'CHiME' Speech Separation and Recognition Challenge," in *InterSpeech 2011*, 2011.

[13] W. Ryan and K. Burk, "Perceptual and acoustic correlates in the speech of males," *Journal of Communication Disorders*, vol. 7, pp. 181–192, 1974.

[14] N. Takeda, G. Thomas, and C. Ludlow, "Aging effects on motor units in the human thyroarytenoid muscle," *Laryngoscope*, vol. 110, pp. 1018–1025, 2000.

[15] A. Baba, S. Yoshizawa, M. Yamada, A. Lee, and K. Shikano, "Acoustic models of the elderly for large-vocabulary continuous speech recognition," *Electronics and Communications in Japan, Part 2, Vol. 87, No. 7, 2004*, vol. 87, no. 2, pp. 49–57, 2004.

[16] R. C. Vipperla, M. Wolters, K. Georgila, and S. Renals, "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*, ser. Lecture Notes in Computer Science, no. 5615. Springer, 2009, pp. 117–126.

[17] M. Vacher, B. Lecouteux, P. Chahuara, F. Portet, B. Meillon, and N. Bonnefond, "The Sweet-Home speech and multimodal corpus for home automation interaction," in *The 9th edition of the Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 2014, pp. 4499–4506.

[18] M. Vacher, B. Lecouteux, and F. Portet, "Multichannel Automatic Recognition of Voice Command in a Multi-Room Smart Home : an Experiment involving Seniors and Users with Visual Impairment," in *Interspeech 2014*, Singapour, Singapore, Sep. 2014, pp. 1–5.

[19] S. Bouakaz, M. Vacher, M.-E. Bobillier-Chaumon, F. Aman, S. Bekkadja, F. Portet, E. Guillou, S. Rossato, E. Desserée, P. Traineau, J.-P. Vimon, and T. Chevalier, "CIRDO : Smart companion for helping elderly to live at home for longer," *IRBM*, vol. 35, no. 2, pp. 101–108, Mar. 2014, 8 pages.

[20] F. Aman, M. Vacher, S. Rossato, and F. Portet, "Speech Recognition of Aged Voices in the AAL Context : Detection of Distress Sentences," in *The 7th International Conference on Speech Technology and Human-Computer Dialogue, SpeD 2013*. Cluj-Napoca, Romania : IEEE, 2013, pp. 177–184.

[21] F. Aman, V. Auberge, and M. Vacher, "How affects can perturbe the automatic speech recognition of domotic interactions," in *Workshop on Affective Social Speech Signals*, Grenoble, France, 2013, pp. 1–5.