MAP-i 2010-2011 Thesis proposal

Silent Speech for Human-Computer Interface in European Portuguese

Supervisor

Prof. António Joaquim da Silva Teixeira (DETI/IEETA, Universidade de Aveiro)

Email: ajst@ua.pt, Tel. 234370526

Co-supervisor

Prof. Miguel Sales Dias (Microsoft and ISCTE – Instituto Universitário de Lisboa)

Email: Miguel.Dias@microsoft.com

Research Unit: IEETA, University of Aveiro

Background

Automatic Speech Recognition (ASR) in the presence of environmental noise is still a problem in speech science [1]. To tackle this problem in the context of Automatic Speech Recognition for Human-Computer Interaction, we propose a novel Silent Speech Interface (SSI) in European Portuguese. An SSI performs ASR in the absence of an intelligible acoustic signal and can be used as a human-computer interface (HCI) modality in high-background-noise environments such as living rooms, or in aiding speech-impaired individuals such as elderly persons [2]. By acquiring sensor data from elements of the human speech production process – from glottal and articulators activity, their neural pathways or the brain itself – an SSI produces an alternative digital representation of speech, which can be recognized and interpreted as data, synthesized directly or routed into a communications network. The existent experimental SSI systems described in the literature are based on the following approaches: capture of the movement of fixed points on the articulators using Electromagnetic Articulography (EMA) sensors [3]; Real-time characterization of the vocal tract using ultra-sound (US) and optical imaging of the tongue and lips [4][5]; Capture movements of a talker's face through ultrasonic sensing devices [6][7]; Digital transformation of signals from a Non-Audible Murmur (NAM) microphone (a type of stethoscopic microphone) [8]; Analysis of glottal activity using electromagnetic [1][9], or vibration [10] sensors; surface electromyography (sEMG) of the articulator muscles or the larynx [11][12]; Interpretation of signals from electro-encephalographic (EEG) sensors [13][14]; Interpretation of signals from implants in the speech-motor cortex [15]; processing of signals from low power radar devices [16].

The existent SSIs have been mainly developed by investigation groups from EUA [4], Germany [17], France [18] and Japan [8], and focused on their respective languages. There is no published work for European Portuguese in the area of SSIs, although there are previous investigations on related areas, such as: use of EMA [19], Electroglotograph and MRI [20] for speech production studies, articulatory synthesis [21] and multimodal interfaces involving speech [22][23].

Objectives

This thesis aims at investigating and developing a successful and natural SSI HCI modality which targets all users (universal HCI) including elderly, for European Portuguese (EP). This SSI will be then used as a HCI modality to interact with computing systems and smartphones. The research will be specifically interested in studying the benefit brought by this novel HCI modality for EP elderly speakers, for whom speaking might require a substantial effort. This research will adopt the following approaches:

a. Multi-sensor Analysis

An SSI can be implemented using several types of sensors or a combination of them in order to achieve better results. For this thesis we will preferably adopt the less invasive approaches and sensors that are able to work both in silent and noisy environments such as, video, NAM microphones, sEMG or Ultra-Sound, following recommendations found in the literature, regarding multi-sensor devices. Further investigation will also be conducted on feature extraction, data acquisition and silent speech processing (silent speech recognition with Hidden Markov Models or other machine learning techniques), of data collected from these sensors, as well as on combining techniques through multi-sensor devices and data fusion in order to complement and overcome the inherent shortcomings of some devices without decreasing the usability.

b. European Portuguese Adoption

The existing SSI research projects do not contemplate any languages other than English and Japanese. With this work we will address the challenges of developing an SSI for European Portuguese, the first approach for this language in the Portuguese and international academia. One of the areas of research to address is the problem of recognizing nasal sounds.

c. Targeting Universal Interface and Elderly Speakers

After determining the different possibilities for each type of SSI, a hybrid and minimally invasive solution will be envisioned, specified, developed and tested, including existing hardware components and new software solutions, targeting a universal interface including elderly people. The specific limitations and requirements imposed by an elderly speaker need to be stipulated based on a pre-defined user profile in order to provide an efficient use of the interface.

d. User Requirements and Usability Evaluation

During the full span of the project duration, close contact with end-users, including elderly, will be sought, starting from user requirements capture to the adoption of a full usability evaluation methodology, which will collect feedback and draw conclusions based on real subjects while interacting (using SSI) with computing systems and smartphones, respectively, in real case indoor home scenarios and in mobility environments.

References

[1] Ng, L., Burnett, G., Holzrichter, J., Gable, T., 2000. Denoising of human speech using combined acoustic and EM sensor signal processing. In: Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, 5–9 June 2000, Vol. 1, pp. 229–232.

[2] Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J.M., and Brumberg, J.S., "Silent speech interfaces", Speech Communication, 52, pp. 270-287, 2009.

[3] Fagan, M.J., Ell, S.R., Gilbert, J.M., Sarrazin, E., Chapman, P.M., 2008. Development of a (silent) speech recognition system for patients following laryngectomy. Med. Eng. Phys. 30 (4), 419–425.

[4] Thomas Hueber, Elie-Laurent Benaroya , Gérard Chollet, Bruce Denby, Gérard Dreyfus, Maureen Stone, Visuo-Phonetic Decoding using Multi-Stream and Context-Dependent Models for an Ultrasound-based Silent Speech Interface, in Proceedings of Interspeech 2009, Brighton, UK; September 2009.

[5] Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., Stone, M., 2007a. Eigentongue feature extraction for an ultrasound-based silent speech interface. In: IEEE Internat. Conf. on Acoustic, Speech, and Signal Processing, ICASSP07, Honolulu, Vol. 1, pp. 1245–1248.

[6] Srinivasan S., Raj B., and Ezzat T., "Ultrasonic sensing for robust speech recognition," in ICASSP, 2010.

[7] Kalgaonkar K. and Raj B., "Ultrasonic doppler sensor for speaker recognition," in *ICASSP*, 2008.

[8] Tomoki Toda, Keigo Nakamura, Takayuki Nagai,Tomomi Kaino, Yoshitaka Nakajima, Kiyohiro Shikano, Technologies for Processing Body-Conducted Speech Detected with Non-Audible Murmur Microphone, in Proceedings of Interspeech 2009, Brighton, UK; September 2009.

[9] Quatieri, T.F., Messing, D., Brady, K., Campbell, W.B., Campbell, J.P., Brandstein, M., Weinstein, C.J., Tardelli, J.D., Gatewood, P.D., 2006. Exploiting non-acoustic sensors for speech enhancement. IEEE Trans. Audio Speech Lang. Process. 14 (2), 533–544.

[10] Patil, S.A., Hansen, J.H.L., this issue. A competitive alternative for speaker assessment: physiological Microphone (PMIC). Speech Comm.

[11] Michael Wand, Szu-Chen Stan Jou, Arthur R. Toth, Tanja Schultz, Synthesizing Speech from Electromyography using Voice Transformation Techniques, in Proceedings of Interspeech 2009, Brighton, UK; September 2009.

[12] Maier-Hein, L., Metze, F., Schultz, T., Waibel, A., 2005. Session independent non-audible speech recognition using surface electromyography. In: IEEE Workshop on Automatic Speech Recognition and Understanding, San Juan, Puerto Rico, pp. 331–336.

[13] Dornhege, G., del R. Millan, J., Hinterberger, T., McFarland, D., Mu¨ ller, K.-R. (Eds.), 2007. Towards Brain–Computer Interfacing. MIT Press.

[14] Wolpaw, J.R., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T., 2002. Brain–computer interfaces for communication and control. Clin. Neurophysiol. 113 (6), 767–791.

[15] Brumberg, J.S, Nieto-Castanon, A., Kennedy, P.R., Guenther, F.H., this issue. Brain–computer interfaces for speech communication. Speech Comm.

[16] John F. Holzrichter, Characterizing Silent and Pseudo-Silent Speech using Radar-like Sensors, in Proceedings of Interspeech 2009, Brighton, UK; September 2009.

[17] Calliess, J.-P., Schultz, T., 2006. Further Investigations on Unspoken Speech. Studienarbeit, Universita¨ t Karlsruhe (TH), Karlsruhe, Germany.

[18] Viet-Anh Tran, Gérard Bailly, Hélène Loevenbruck , Tomoki Toda, Multimodal HMM-based NAM-to-speech conversion, in Proceedings of Interspeech 2009, Brighton, UK; September 2009.

[19] ROSSATO, Solange; TEIXEIRA, António; FERREIRA, Liliana - Les Nasales du Portugais et du Français : une étude comparative sur les données EMMA . In XXVI Journées d'Études de la Parole. Dinard, FR, Jun. 2006.

[20] MARTINS, Paula; CARBONE, Inês; PINTO, Alda; SILVA, Augusto; TEIXEIRA, António - European Portuguese MRI based speech production studies. Speech Communication. NL: Elsevier. ISSN: 0167-6393, vol. 50, nº 11/12 (12. 2008). p. 925 – 952.

[21] António Teixeira and Francisco Vaz, Síntese Articulatória dos Sons Nasais do Português, Anais do V Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR) 2000, pp. 183-193, ICMC-USP, Atibaia, São Paulo, Brasil.

[22] TEIXEIRA, António J. S.; MARTINEZ, Roberto; SILVA, Luís Nuno; JESUS, Luís M. T.; PRÍNCIPE, José Carlos; VAZ, Francisco A. C. - Simulation of Human Speech Production Applied to the Study and Synthesis of European Portuguese. Eurasip Journal on Applied Signal Processing: Hindawi Publishing Corporation, vol. 2005, nº 9 (Jun. 2005). p. 1435-1448.

[23] M. Sales Dias et al., Using Hand Gesture and Speech in a Multimodal Augmented Reality Environment, GW2007, LNAI 5085, pp.175-180, 2009.