Development of Voice-Based Tools for Accessibility to Computer Services

Desarrollo de herramientas de accesibilidad al ordenador basadas en la voz

Oscar Saz Torralba, William Ricardo Rodríguez Dueñas, and Eduardo Lleida Solano

Communications Technology Group (GTC), Aragón Institute for Engineering Research (I3A),
University of Zaragoza, Zaragoza,
Spain
{oskarsaz, wricardo, lleida}@unizar.es

Abstract. This work presents the development of two tools which intend to provide accessibility to people with different handicaps in their access to computer applications using speech technologies. The basis of the work is to use voice emissions from a severely handicapped user to substitute mouse and key strokes in one of the tools; and the movement of the cursor in the other. The speech technologies required for these tasks are robust energy estimation and robust formant calculation and normalization. The paper will also provide a comprehensive view of all the process required for a successful development of these tools, starting with the contact with assistive and education institutions, understanding the difficulties of these collectives in their everyday life, defining how technology can help in these cases, the actual development of the tools and, finally, the deployment of the tools with real users to assess their usability and functionality.

Keywords. Voice I/O, user-centered design, signal analysis synthesis and processing, speech processing, universal accessibility and handicapped aids.

Resumen. Este trabajo presenta el desarrollo de dos herramientas que buscan proveer accesibilidad a personas con diferentes discapacidades en su acceso a aplicaciones informáticas a través de las tecnologías del habla. La base del trabajo es usar emisiones de voz por parte de un usuario con discapacidad severa para sustituir las pulsaciones de ratón o teclado en una de las herramientas; y el movimiento del cursor en la otra. Las tecnologías del habla requeridas para ello son estimación de energía robusta y cálculo robusto de formantes con normalización. El artículo también buscar dar una visión global del proceso requerido para el desarrollo exitoso de estas herramientas, empezando por el contacto con instituciones asistivas y educativas, entender las dificultades de estos colectivos, definir qué soluciones puede dar la tecnología, el propio desarrollo de las herramientas y, finalmente, el despliegue de las mismas ante usuarios reales para conocer su usabilidad y funcionalidad.

Palabras clave. Voz E/S, diseño centrado en el usuario, análisis síntesis y procesado de señal, procesado de voz, accesibilidad universal y ayudas a discapacidad.

1 Introduction

Recent advances in Information and Communication Technologies (ICTs) have made the access to knowledge immediate and universal. The possibilities that computers and Internet open for a better education via e-learning, effective work inclusion and e-government strongly facilitate our lives nowadays. However, new barriers appear for sensorial, physical and cognitive impaired individuals within these novel technological elements. This way, elements which were initially supposed to promote universal access to information are actually increasing the gap between impaired and non-impaired users.

For this reason, the so called e-inclusion is becoming a major issue for governments, industry and also academia. Technical solutions which can help to avoid technological barriers are coming out from research institutions and technological companies, accompanied by an effort from

funding agencies and institutions [Cucchiarini et al., 2008]. However, certain difficulties arise when creating technical aids for the handicapped, which limit the development of these devices. On one hand, every user is unique in his/her specific impairments and needs making that the tools or systems developed have to be especially designed for each one of them or have a high capability of adaptation. Furthermore, developers require the knowledge and help from experts in different fields like sociology, education or assistance to fully understand the needs of the handicapped community and, hence, develop tools which can be really useful for them.

The work presented in this paper is part of a long running interest in new software and systems based on speech technologies applied to enhance the quality of life of the handicapped community, which has lead previously to the set of tools for automated speech therapy in "Comunica" [Saz et al., 2009]. The basis of this work is the collaboration with different associations with interest in the education and assistance of the handicapped: the Coordinator of Associations for the Handicapped (CADIS) in Huesca, Spain and the Spanish Confederation of Associations for the Assistance to People with Cerebral Palsy (ASPACE) in Huesca, Spain. This collaboration has led to two different voice and speech based solutions to provide accessibility to computers that are presented and explained in this work.

This paper is organized as follows: Section 2 will first provide an introduction to these barriers which limit handicapped individuals accessing information. Then, Sections 3 and 4 will present the two approaches to use voice as assistive interface in the access to computers by physically handicapped individuals. In the end, Sections 5 and 6 will give the discussion and conclusions to the work carried out.

2 Barriers in the Access to Information

There are several elements which limit the possibilities for handicapped individuals to access information and knowledge. This distinction relies on the specific impairment of the user and the way it affects his/her motor function, cognitive capabilities or sensorial faculties.

For persons who suffer severe physical impairments, the control of the peripherals of a computer can become impossible due to the affections at the nervous system which control the muscles and their movements; for instance, a patient of cerebral palsy with spastic movements. For this reason, computer interaction for these people has to be fully adapted via the use of devices which substitute or emulate these peripherals. Virtual mice and virtual keyboards simplify and reduce the workload of the physically impaired user: They present sequentially in screen the different options of the device (directions of the mouse, keys in the keyboard) and the user activates an external switch (or push button) to select the option currently highlighted in screen. The speed of the interaction when working with these devices is very slow compared to the normal interaction of an unimpaired user, but the big difference that they make for a handicapped user who, otherwise, could not access the computer without them is so huge that it quickly overcomes any difficulty in their use.

2.1 Human Computer Interfaces for the Disabled

The need for new types of interfaces is, hence, important for all these communities. These novel proposals for interfaces include speech (which will be reviewed later), movement tracking [Sánchez Ortega et al., 2009] [Varona et al., 2008], eye tracking [Palomo et al., 2009], brain computer interface [Iturrate et al., 2009], or any other possible interface which permits for a more natural human machine interaction despite the impairments of the user. Several research groups worldwide are developing new proposals of these enhanced interfaces, which, in some cases lead to devices which can be finally applied to a certain part of the handicapped population.

Among these possibilities, it is usually considered that speech (or voice) can provide these users with an enhanced interface that overcomes their existing limitations. Speech is a very natural way of communication and the most common interface in human communication, so it
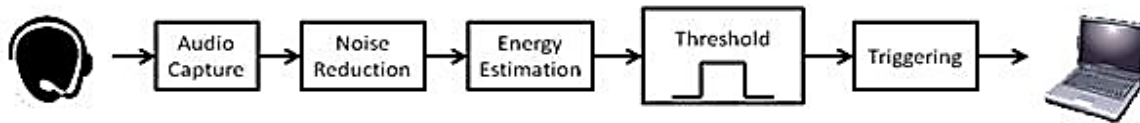
Fig. 1. Signal processing for voice-based triggering of mouse strokes

has very often been proposed for natural human machine interaction.

Different approaches have been studied through research for creating speech interfaces for control of external appliances through Automatic Speech Recognition (ASR) [Hawley et al., 2003], [Sanders et al., 2002], [Sharma **&** Hasegawa-Johnson, 2009] which can be easily translated to computer control or many other applications. Also, emulation of joystick or mouse peripherals has been proved to be possible by the detection of vowels produced by the user [Bilmes et al., 2006], [Harada et al., 2008]. Devices which translate difficult speech from a handicapped user to a clearer and more intelligible speech have also been developed to provide accessibility [Creer et al., 2009].

However, more work is required in this area. The specificity of every user and the difficulty to translate cases of success from previous works to different people and conditions make necessary to continue on the research of speech and voice as alternative interaction. This research has to present two edges: First, increase the theoretic knowledge on the different disorders and its influence on automated systems like ASR; and, second, the rapid development of tools based on these systems which can actually provide a solution to people.

3 Voice-Based Triggering of Keyboard and Mouse Events

As mentioned, voice can, in many cases, suppose an easier interface for users with a limited control of movements, overcoming the limitations of physical devices. However, motor disabilities very often produce also a deterioration of the articulatory skills of the patient in the form of dysarthria, dysphonia and other speech impairments. In these cases, speech can be extremely unintelligible and becomes a serious challenge for automated speech-based systems.

"Vozclick" was developed with the aim of substituting the use of physical triggering of keyboard and mouse events. It directly relies on the voice emission of the user, independently of the content of the utterance, to avoid limitations in usability due to the possible speech impairments of the user. It performs a signal processing of the input audio signal to detect pulses of voice in the user's emission, as it can be seen in Figure 1.

The system captures the audio from the user and performs noise reduction by the use of a spectral subtraction technique to avoid false alarms due to external noises in the environment or to electrical noise in the audio equipment. The energy of the audio frame is calculated and, when a certain intensity threshold ($E_{threshold}$) is reached and maintained during a predefined amount of time t ($t_{minimum} < t < t_{maximum}$) the software automatically triggers the desired event (mouse click or key stroke).

The user interface of "VozClick" in Figure2 was designed to allow a quick configuration of all the key elements in the application. The person who assists the disabled user can set up the values of the intensity and duration of the voice emission to be recognized as a voluntary action of the user ($E_{threshold}$, $t_{minimum}$ and $t_{maximum}$). It can be decided also whether to apply noise subtraction or not, to consider only voiced or unvoiced segments of speech; and, finally, the event to be triggered by voice (left button mouse click or specific key stroke).
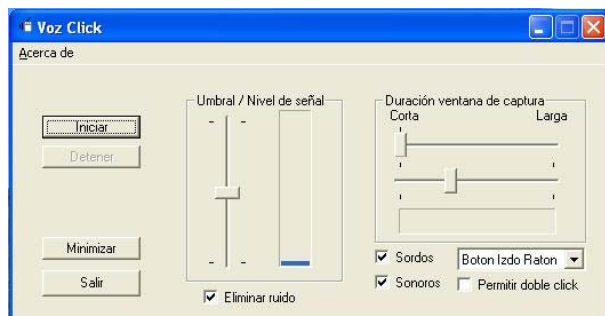
Fig. 2. "VozClick" user interface

3.1 Energy Estimation and Noise Reduction

The estimation of the energy is performed after windowing the input signal in frames of 25 milliseconds with a 50% overlap. The instant energy for the n-th frame is calculated in E_{frame} as in Equation (1).

$$E_{frame}(n) = \sum_{i=1}^{framesize} x(i)^2 \qquad (1)$$

To reduce the high variability of the input energy between frames, which may lead to inaccurate detection of the rising and descending slopes in the oral production of the user, an smoothed calculation of the logarithm of the energy (logE) is used for each frame n as in Equation (2).

$$\log E(n) = 0.1 * \log_{10}\left(E_{frame}(n)\right) + \\ 0.9 * \log E(n-1) \qquad (2)$$

Although this method for calculating the energy is very simple, it works accurately with signals with a low presence of noise, but it can be extremely dependent on the input background energy due to the many sources of sound which are not originated by the user. These background noises, which include external sounds as well as internal electronic noises, may limit the functionality of the tool, as they would produce a large amount of false alarms in the means of triggering of events not desired by the user.

For this reason, a spectral subtraction method based on Minimum Mean Square Error Log Spectral Amplitude (MMSE-LSA) estimation is applied [Ephraim **&** Malah, 1984][Ephraim **&** Malah, 1985] to the captured speech signal. This method estimates a filter G in the spectral domain for each frame n which is a function of the estimation of the noise present on the spectrum and the estimated Signal to Noise Ratio (SNR) for the frame. The estimation of the power spectral density of the noise, required to estimate the SNR, is readjusted every unvoiced frame to track dynamically changes in the noise spectrum due to possible variations in the environment.

4 Vowel-Controlled Emulation of Mouse

Users with more capabilities in their control of phonation and articulation might find "VozClick" too simple for their interests and possibilities. Even when an ASR system would still not be suitable for these users, vowels can be used to build a whole system which allows to fully emulate a device like a mouse as it has been proved before for English [Bilmes et al., 2006].

"VocalClick" made use of the knowledge acquired in robust formant estimation and vowel detection for different users via formant normalization, as seen in Figure 3. The audio captured from the user is followed by a process for the estimation of the first two formant frequencies in the speech (F_1 and F_2). This process performs a Linear Predictive Coding (LPC) analysis of the signal, followed by a liftering in the homomorphic domain to avoid the influence of the fundamental frequency (or pitch) in the formant estimation. The calculated formant values are normalized according to the relationship between physical parameters of the user (height and vocal tract length) obtaining the corresponding normalized formant frequencies F_{1N} and F_{2N} [Rodríguez and Lleida, 2009]. The vowel corresponding to the normalized formant values is decoded as the closest vowel in the formant map and the corresponding mouse movement is performed. The use of formant normalization allows using the same formant map
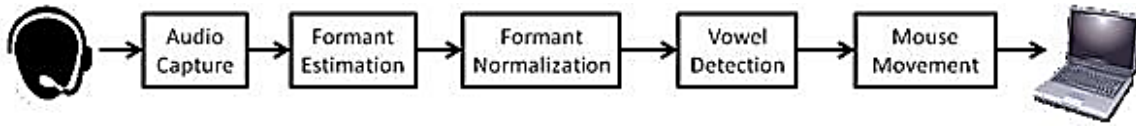
Fig. 3. Signal processing for vowel control of mouse

for all users, independently of their physical characteristics.

The user interface of "VocalClick" in Figure 4 indicates how different vowels act in different spatial directions (/a/ for right, /e/ for up, /i/ for left and /o/ or /u/ for down). Other elements can also be configured, especially the gender and height of the person; these two data allow for an accurate formant normalization which can map the formant map of a given user to the standard formant map. There are also thresholds for the intensity and duration of the vocalic emission in a similar as how "Vozclick" worked, to avoid unwanted noises triggering the system.
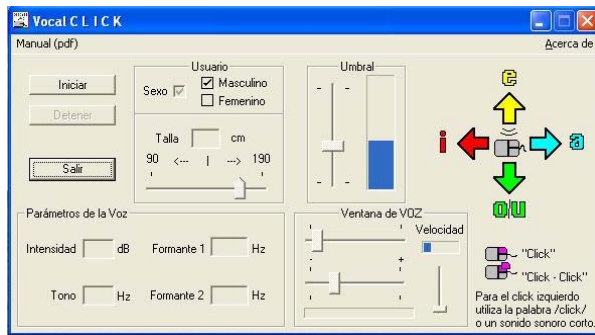

Fig. 4. "VocalClick" user interface

4.1 Robust Formant Estimation

The conventional autocorrelation method using LPC works well in signals with long pitch period (low-pitched). As the pitch period of high-pitched speech is small, the periodic replicas cause aliasing of the autocorrelation sequence. In other words, the accuracy of the LPC method decreases as the fundamental frequency F_0 of speech increases [Rahman **&** Shimamura, 2005]. In that case, it is required to separate these effects in order to obtain formants not contaminated by F_0 by means of homomorphic

analysis. The main idea within the homomorphic analysis is the deconvolution of a segment of speech x(n) into a component representing the vocal tract impulse response h(n), and a component representing the excitation source e(n) as in Equation (3).

$$x(n) = e(n) * h(n) \qquad (3)$$

The way in which such separation is achieved is through linear filtering of the cepstrum, defined as the inverse Fourier transform of the log spectrum of the signal. As the cepstrum in the complex domain is not suitable to be used because of its high sensitivity to phase [Rabiner & Schafer, 1978], the real-domain cepstrum c(n) defined by Equation (4) is used, where X(k) is the N-point Fourier transform of the input speech signal x(n).

$$c(n) = \frac{1}{N} \sum_{K=0}^{N-1} \ln|X(k)|\, e^{j\frac{2\pi}{N}kn} \quad 0 \le n \\ \le N-1 \qquad (4)$$

The values of c(n) around the origin correspond primarily to the vocal tract impulse information, while the farthest values are affected mostly by the excitation. Knowing previously the value of the pitch period (T_{pitch}), from the LPC analysis using the autocorrelation method, it is possible to filter the cepstrum signal (liftering) and use the liftered signal to find the formant frequencies. After the liftering process, the formant frequencies without the pitch influence are obtained using the original LPC method, increasing the formant estimation accuracy.

4.2 Formant Normalization

The main difficulty in decoding the uttered vowel by a speaker knowing only the formants

frequencies from the user's voice is that different speakers have different formant values for the same vowel. More precisely, speakers with shorter vocal tract lengths (children and women) tend to have higher F_1 and F_2 values than adult males, producing an uncertainty in the vowel decoding. To avoid this, once the current user's vocal tract length is estimated, formant normalization allows translating the formant frequencies of a given speaker to a standard formant map which can be used for decoding the uttered vowel.

If no a priori information is known from the speaker, like gender or age, the estimation of the vocal tract length can be done from the formant frequencies calculated directly from speech. Considering the vocal tract as a uniform lossless acoustic tube, its resonant frequencies are given by Equation (5) and they are uniformly spaced, where v = 35300 cm/s is the speed of sound at 35º C, and l is the length of the uniform tube in cm.

$$F_k = \frac{v}{4l}\,(2k-1)\ \ k = 1.2.3\ ... \tag{5}$$

The estimation of the length can be reduced to fitting the set of resonance frequencies of the uniform tube [Necioglu et al., 2000], which are determined solely by its length l. Therefore, the problem can be approximated to minimizing Equation (6), as the sum of the difference between the measured formants and the resonances of the uniform tube. From [Burhan et al., 2000], this error measure can be simplified in the second part of Equation (6) using the distance function between the M measured formants and the odd resonances of a uniform tube, $(2k-1)F_1$.

$$\varepsilon = \sum_{k=1}^{M} D\left(F_k, (2k-1)\frac{v}{4l}\right) \tag{6}$$

$$= \sum_{k=1}^{M} \frac{\left(\frac{F_k}{2k-1} - F_1\right)^2}{F_1}$$

Finally, the vocal tract length can be obtained with the expression in Equation (7) which makes use of the estimated resonance frequency of the

uniform tube (F_1), calculated from previous equations as in Equation (8).

$$VTL = \frac{v}{4\,F_1} \tag{7}$$

$$F_1 = \sqrt{\frac{1}{M}\sum_{k=1}^{M}\left(\frac{F_k}{2k-1}\right)^2} \tag{8}$$

A more straightforward possibility, if the speaker provides information on gender and height is to directly estimate the vocal tract length applying the known relationships between these parameters and the vocal tract. This is the process performed in "VocalClick", which makes use of previous work on the study of the relationship between physical parameters and the vocal tract [Rodríguez **&** Lleida, 2009]. If this could not be possible, the vocal tract length could be estimated through an enrollment session uttering the 5 Spanish vowels (/a/, /e/, /i/, /o/ and /u/) and performing the process depicted previously.

Once the speaker's vocal tract length is estimated, the formant frequencies F_k obtained can be normalized with it. The formant normalization used in this work is based on the hypothesis that the vocal tract configurations of different speakers are similar to each other and differ only in length [Wakita, 1977]. Based upon this hypothesis for normalization, it is necessary to compute the resonance frequencies of an acoustic tube when the length of the tube l is varied to a reference length l_R without altering its shape. Hence, the normalized formants F_{kN} are computed in Equation (9) by multiplying the unnormalized formants by the length factor, l/l_R, with l_R fixed at 17.5cm, and l the vocal tract length obtained previously.

$$F_{kN} = \frac{l}{l_R}\,\tilde{F}_k \tag{9}$$

This formant estimation and normalization methods were evaluated for providing Vocal Tract Length Normalization (VTLN) in an ASR task [Rodríguez et al., 2010]. The results show that the

technique is accurate in detecting differences among vocal tracts of different individuals in comparison with similar techniques for vocal tract normalization based either in parametric approaches [Gouvea **&** Stern, 1997] or in statistical approaches [Lee and Rose, 1998].

5 Deployment and Discussion

The main point of discussion in the development of assistive tools for the handicapped is always whether the creations which arise from technological developers are actually helpful or not for the intended users. It is not rare the case where technological solutions with complex engineering processing within them end up being forgotten due to the fact that they do not match the special requirements of the users.

The technical aids presented here counted with the collaboration of different associations and institutions for their development and posterior deployment. CADIS-Huesca and ASPACE-Huesca actively participated in the development of "VozClick" and "VocalClick", through several coordinating meetings and evaluation sessions with 3 users affected with cerebral palsy of the assistive center of ASPACE in Huesca. These meetings served to make new proposals in the tools, including the use of noise reduction methods to avoid the influence of environmental and electrical noises. "VozClick" and "VocalClick" are now available for the whole community through the website of "Comunica"[1], altogether with extra material and tutorials, in an effort to widespread every tool which can serve to improve the quality of life of part of the handicapped community.

One of the most difficult tasks when developing technical aids is to find a way to validate the usefulness of the developed technologies. Currently, a group of users of ASPACE-Huesca with severe cerebral palsy are using "VozClick" to complement virtual mice and keyboards for interacting with computers. These users have mid to high cognitive capabilities so they are able to perform several tasks like navigating the Internet, perform simple text

processing or work with presentation programs with the help of devices which help them to overcome their physical handicaps, as this is the case. "VozClick" has helped them to avoid the use of physical switches which, in some cases, produced difficulty of use and excessive physical load resulting in discomfort and pain.

6 Conclusions

As conclusion to this work, a set of technical aids for providing accessibility to computer and Internet services have been presented. These tools propose the use of different speech technologies as a way to solve the gap that sensorial, physical and cognitive disabled users suffer in their access to new technologies. These technologies have been voice detection and robust formant estimation. The maturity reached in these technologies at research and commercial levels made them feasible to be used in these special tasks, where the adaptability to the user and the robustness to different situations are key points in the success of these developed tools.

This work has been acknowledged by the foundation Physical Disabled of Aragón (DFA) as finalist in their XIII Awards for Universal Accessibility 2010 in the category of Accessibility in Information and Communication Technologies. This award recognizes the effort of individuals, institutions, companies and academic groups to reduce the barriers of handicapped people (not only physical handicapped) in their access to several areas of their life, including mobility, access to information and communication.

Future work has to deal with the use of more advance systems for oral interfaces. ASR can be applied to a large number of users with handicaps whose speech is still intelligible and recognizable. Adaptation and personalization are the key points to develop successful speech systems for users with special needs. This adaptation allows the system to learn the special acoustic properties in the user speech, the lexical variants in the speech and the special vocabulary and syntax of a specific user. All this can lead to personalized systems that can accompany the person through different tasks and activities.

---

[1] http://www.vocaliza.es/cadis.html

Acknowledgements

Referencias

1. **Bilmes, J.-A., Malkin, J., Li, X., Harada, S., Kilanski, K., Kirchhoff, K., Wright, R., Subramanya, A., Landay, J.-A., Dowden, P. & Chizeck, H. (2006).** The Vocal Joystick. *2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006)*, Toulouse, France, 1, 625-628.

2. **Necioglu, B.F., Clements, M.A., & Barnwell, T.P. (2000).** Unsupervised Estimation of the Human Vocal Tract Length over Sentence Level Utterances. *2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2000),* Istanbul, Turkey, 3, 1319-1322.

3. **Creer, S. M., Green, P. D., Cunningham, S. P. & Fatema, K. (2009).** Personalizing Synthetic Voices for People with Progressive Speech Disorders: Judging Voice Similarity. *10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009)*, Brighton, United Kingdom, 1427-1430.

4. **Cucchiarini, C., Lembrechts, D. & Strik, H. (2008).** HLT and Communicative Disabilities: The Need for Co-Operation between Government, Industry and Academia. *Language and Speech Technology Conference* (*LangTech 2008*), Rome, Italy, 125-128.

5. **Ephraim, Y. & Malah, D. (1984).** Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Transactions on Acoustic, Speech Signal Processing,* 32(6), 1109-1121.

6. **Ephraim, Y. & Malah, D. (1985)**. Speech Enhancement Using A Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Transactions on Acoustic, Speech Signal Processing,* 33(2), 443–445.

7. **Gouvea, E.-B., & Stern, R.-M. (1997).** Speaker Normalization Through Formant-Based Warping Of The Frequency Scale. *5th European Conference on Speech Communication and Technology (EUROSPEECH 1997)*, Rhodes, Greece, 1139–1142.

8. **Harada, S., Landay, J.-A., Malkin, J., Li, X. & Bilmes, J.A. (2008).** The Vocal Joystick: Evaluation of Voice-Based Cursor Control Techniques for Assistive Technology. *Disability and Rehabilitation: Assistive Technology*, 3(1), 22-34.

9. **Hawley, M., Enderby, P., Green, P., Brownsell, S., Hatzis, A., Parker, M., Carmichael, J., Cunningham, S., O'Neill, P. & Palmer, R. (2003).** STARDUST: Speech Training And Recognition for Dysarthic Users of Assistive Technology. *7th European Conference for the Advancement of Assistive Technology (AAATE 2003)*, Dublin, Ireland, 959-963.

10. **Iturrate, I., Antelis, J.-M., Kübler, A. & Mínguez, J. (2009).** A Non-Invasive Brain-Actuated Wheelchair based on a P300 Neurophysiological Protocol and Automated Navigation. *IEEE Transactions on Robotics*, 25(3), 614–627.

11. **Lee, L. & Rose, R. (1998).** A Frequency Warping Approach to Speaker Normalization. *IEEE Transactions on Speech and Audio Processing,* 6(1), 49–60.

12. **Palomo, P., González, T., Rivas, R., Irazabal I. & Ruiz, A. (2009).** IRISCOM. Proyecto IRIS. *IV Jornadas Iberoamericanas de Tecnologías de Apoyo a Discapacidad, Las tecnologías de apoyo en parálisis cerebral,* Madrid, Spain, 87-91.

13. **Rodríguez, W.-R. & Lleida, E. (2009)**. Formant Estimation in Children's Speech and its application for a Spanish Speech Therapy Tool. *2009 Workshop on Speech and Language Technologies in Education (SLaTE 2009)*, Wroxall Abbey Estates, Warwickshire, United Kingdom, 81-84.

14. **Rodríguez, W.-R., Saz, O., Miguel, A. & Lleida, E. (2010).** On Line Vocal Tract Length Estimation for Speaker Normalization in Speech Recognition. *VI Jornadas en Tecnología del Habla y II Iberian SLTech Workshop (Fala 2010)*, Vigo, Spain, 119-122.

15. **Sánchez Ortega, P.-L., Cámara Nebreda, J.-M. & Núñez Angulo, B. (2009)**. Interacción Con Ordenador Mediantes Dispositivos Inalámbricos Para Usuarios Con Movilidad Muy Reducida. Colaboración Universidad de Burgos – APACE. *IV Jornadas Iberoamericanas de Tecnologías de Apoyo a*

Discapacidad, Las tecnologías de apoyo en parálisis cerebral, Madrid, Spain, 45-49.

16. **Sanders, E., Ruiter, M., Beijer, L. & Strik, H. (2002).** Automatic Recognition of Dutch Dysarthric Speech: A Pilot Study. *7th International Conference on Spoken Language Processing (INTERSPEECH 2002),* Denver CO, USA, 661-664.

17. **Saz, O., Lleida, E., Abarca, L. & Mejuto, S. (2009).** Mouseclick: Acceso Al Ordenador A Través De La Voz. *IV Jornadas Iberoamericanas de Tecnologías de Apoyo a Discapacidad, Las tecnologías de apoyo en parálisis cerebral,* Madrid, Spain, 63-67.

18. **Saz, O., Yin, S.-C., Lleida, E., Rose, R., Vaquero, C. & Rodríguez, W.-R. (2009)**. Tools and Technologies for Computer-Aided Speech and Language Therapy. *Speech Communication*, 51(10), 948-96.

19. **Rabiner, L.-R. & Schafer, R.-W. (1978).** *Digital Processing of Speech Signals.* Englewood Cliffs NJ, USA: Prentice Hall.

20. **Rahman, M.-S. & Shimamura, T. (2005).** Formant Frequency Estimation of High-Pitched Speech by Homomorphic Prediction. *Acoustical Science and Technology,* 26(6), 502-510.

21. **Sharma, H.-V. & Hasegawa-Johnson, M., (2009).** Universal Access: Speech Recognition for Talkers with Spastic Dysarthria. *10th Annual Conference of the International Speech Communication Association (INTERSPEECH 2009)*, Brighton, United Kingdom, 1451–1454

22. **Varona, J., Manresa-Yee, C. & Perales, F.-J. (2008).** Hands-Free Vision-Based Interface for Computer Accessibility, *Journal of Network and Computer Applications,* 31(4), 357–374.

23. **Wakita, H. (1977).** Normalization of Vowels by Vocal Tract Length and its Application to Vowel Identification. *IEEE Transactions on Acoustic, Speech Signal Processing,* 25(2), 183-192.

Oscar Saz Torralba

He received a B.Sc. in Telecommunications Engineering by the University of Zaragoza in 2004 and a Ph.D. by the University of Zaragoza in 2009. He is currently a Fulbright scholar at the Language Technologies Institute at Carnegie Mellon University in Pittsburgh, USA. His research interests are new applications of speech technologies such as oral interfaces for handicapped individuals and computer-assisted language learning.

William Ricardo Rodríguez Dueñas

He received a B.Sc. in Biomedical Engineering by the University of Science and Development in 2002, a M.Sc. in Biomedical Engineering in 2008 by the University of Zaragoza and a Ph.D. by the University of Zaragoza in 2010. He currently is an assistant professor at the Universidad Antonio Nariño in Bogotá, Colombia. His research interests are technologies and interfaces for rehabilitation and speech systems for children.

Eduardo Lleida Solano

He received a B.Sc. in Telecommunications Engineering by the Polytechnic University of Catalonia in 1985 and a Ph.D. by the Polytechnic University of Catalonia in 1990. He currently is a full professor at the Department of electrical engineering and communications of the University of Zaragoza. His research interests are related to speech processing in different areas like acoustics, automatic speech recognition, speech biometrics and oral interfaces.