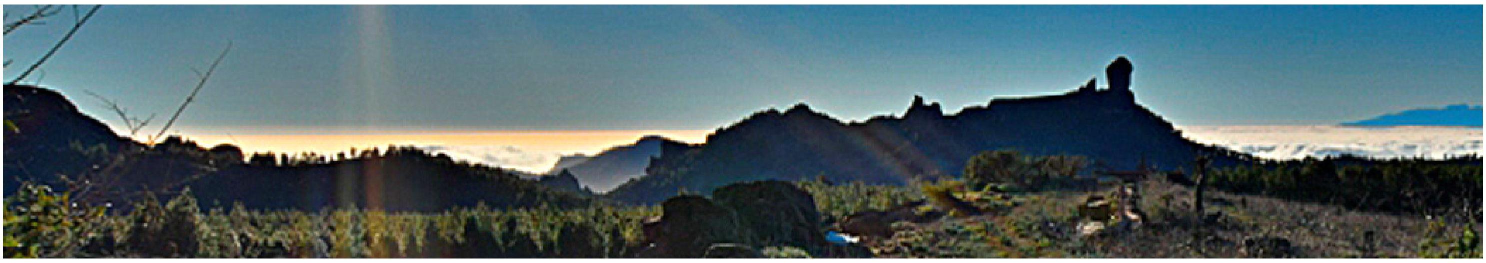
iberSPEECH2014

November 19-21 2014, Las Palmas de Gran Canaria



CONFERENCE PROCEEDINGS

iberSPEECH 2014

VIII Jornadas en Tecnologías del Habla

and IV Iberian SLTech Workshop

Escuela de Ingeniería en Telecomunicación y Electrónica
Universidad de Las Palmas de Gran Canaria
SPAIN

Organized by:

IDeTIC

División de Procesado Digital de Señal

RTTH

Spanish Thematic Network on Speech Technology (RTTH)

ISCA SIG

Supported by:

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

ULPGC 25 ANIVERSARIO 1989-2014

Selected papers published by:

Springer
Lecture Notes in Artificial Intelligence
Lecture Notes in Computer Science

LNAI Vol. 8854

PROCEEDINGS

iberSPEECH 2014

VII Jornadas en Tecnología del Habla
and
III Iberian SLTech Workshop

November 19-21, 2014
Escuela de Ingeniería enTelecomunicación y Electrónica
Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, Spain

Editors: Juan Luis Navarro Mesa
Pedro Quintana Morales
Alfonso Ortega
António Teixeira
Eduardo Hernández Pérez
Antonio Ravelo García
Iván Guerra Moreno
Manuel Medina Molina
Sofía Martín González

Organized by:

IDeTIC

División de Procesado Digital de Señal

RTTH

Spanish Thematic Network on Speech Technology (RTTH)

Supported by:

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA

ULPGC 25 ANIVERSARIO 1989-2014

ISBN: 978-84-617-2862-6

http://iberspeech2014.ulpgc.es

WELCOME MESSAGE

The Spanish Thematic Network on Speech Technology (RTTH) and the ISCA-Special Interest Group on Iberian Languages (SIG-IL) are pleased to present the selected papers of IberSpeech 2014, Joint VIII Jornadas en Tecnologías del Habla and IV Iberian SLTech Workshop, held in Las Palmas de Gran Canaria, Spain, on November 19-21. The Organizing Committee of IberSpeech believes and trusts that we have achieved the quality that the researchers in Advances in Speech and Language Technologies for Iberian Languages value. To ensure this quality, each article has been reviewed at least by three members of the Scientific Review Committee, who have provided feedback to improve the final version of the articles in this book.

The conference has become mature as different editions have been organized, starting in Vigo 2010 with FALA and continuing in Madrid 2012 with the new denomination: Iberspeech. This new edition becomes a step further in the support of researchers in Iberian Languages. IberSpeech is a joint event resulting from the merging of two conferences, the ”Jornadas en Tecnología del Habla” and the Iberian SLTech Workshop. The first has been organized by the “Red Temática en Tecnologías del Habla” (Spanish Speech Technology Thematic Network, http://www.rthabla.es) since 2000. This network was created in 1999 and currently includes over 200 researchers and 30 research groups in speech technology all over Spain. And the second, has been organized by the Special Interest Group on Iberian Languages (SIG-IL, http://www.il-sig.org/) of the International Speech Communication Association (ISCA), The Iberian SLTech Workshop had its first edition in Porto Salvo, Portugal, in 2009.

As a result, IberSpeech is one of the most important research meetings in the field of speech and language processing focusing on Iberian Languages, attracting many researchers (about 140 in the 2014 edition), mainly from Spain, Portugal, and from other Iberian-speaking countries in Latin America. We have also attracted the interest of several research groups from all around the world, including China, United Kingdom, France, Japan, Hungary, Israel, Norway, Czech Republic, and Germany.

Although the main focus is on Iberian Languages and the Iberian region, the conference is not restricted to them. Proof of this are the ALBAYZIN Technology Competitive Evaluations, organized in conjunction with the conference, which in this edition have attracted the interest of several research groups. The ALBAYZIN Technology Competitive Evaluations have been organized alongside with the conference since 2006, promoting the fair and transparent comparison of technology in different fields related to speech and language technology. In this edition we have two different evaluations: Audio Segmentation and Search on Speech. The organization of each one of these evaluations requires preparing development and test data, providing data along with a clear set of rules to the participants, and gathering and comparing results from participants. This organization has been carried out by different groups of researchers and is crucial for the success in participation that we are envisaging. Although results from the evaluations cannot be included in this volume due to timing restrictions, we would like to express our gratitude to the organizers and also to the participants in the evaluations.

We have had 60 submitted papers and, after a strict peer-reviewing process, only 29 have been selected for publication in this volume of the Springer Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence. This selection has been based on the scores and comments provided by our Scientific Review Committee, which includes over 79 researchers from different institutions mainly from Spain, Portugal, Latin America, USA, UK, Hungary and Czech Republic to which we also would like to express our deepest gratitude. Each article has been reviewed by at least three different reviewers and authors have had time to address the comments before submitting the camera-ready paper. The articles have been organized into four different topics:

- Speech Production, Analysis, Coding and Synthesis
- Speaker and Language Characterization
- Automatic Speech Recognition
- Speech and Language Technologies in Different Application Fields

Besides the excellent research articles included in this volume, the conference had the pleasure of having two extraordinary keynote speakers: Dr. Pedro Gómez Vilda (Departamento de Arquitectura y Tecnología de Sistemas Informáticos de la Universidad Politécnica de Madrid, Spain) and Dr. Roger K. Moore (Department of Computer Science University of Sheffield, UK).

We would also like to thank Springer, and in particular to Alfred Hoffmann, for the possibility of publishing this volume, his suggestions in order to increase the spread of the international scope of IberSpeech 2014, his help and great work in preparing it.

Finally, we would like to thank all those whose effort has made possible this conference, the members of the local organizing committee, the technical and program chairs, the reviewers and so many people who put their best to achieve a successful conference.

Juan Luis Navarro Mesa
Alfonso Ortega Giménez
António Teixeira

IberSpeech 2014 Co-Chairs

ORGANIZATION

General Chairs

Juan Luis Navarro Mesa	Universidad de Las Palmas de Gran Canaria
Alfonso Ortega Giménez	Universidad de Zaragoza
António Teixeira	Universidade de Aveiro

Technical Chairs

Carmen García Mateo	Universidade de Vigo
Doroteo Torre Toledano	Universidad Autónoma de Madrid
Rubén San-Segundo Hernández	Universidad Politécnica de Madrid

Program Chairs

Eduardo Hernández Pérez	Universidad de Las Palmas de Gran Canaria
Pedro Quintana Morales	Universidad de Las Palmas de Gran Canaria
Jesús Alonso Hernández	Universidad de Las Palmas de Gran Canaria

Publication Chair

Sofía Martín González	Universidad de Las Palmas de Gran Canaria

Demos Chairs

Javier Hernando Pericas	Universidad Politécnica de Cataluña
Rubén San-Segundo Hernández	Universidad Politécnica de Madrid
Daniela Braga	Voicebox Technologies

Awards Chairs

Javier Hernando Pericas	Universidad Politécnica de Cataluña
Inma Hernáez Rioja	Universidad del País Vasco
Daniela Braga	Voicebox Technologies

Evaluation Chairs

Antonio Ravelo García	Universidad de Las Palmas de Gran Canaria
Carlos Travieso González	Universidad de Las Palmas de Gran Canaria

Local Organizing Committee

Universidad de Las Palmas de Gran Canaria

Jesús B. Alonso Hernández
Juan M. Caballero Suárez
Fidel Cabrera Quintero
Elena García Quevedo
Iván D. Guerra Moreno
Eduardo Hernández Pérez
Sofía I. Martín González
Manuel M. Medina Molina
Juan Luis Navarro Mesa
Pedro J. Quintana Morales
Antonio G. Ravelo García
Carlos Travieso González

Albayzin Commitee

Juan Luis Navarro Mesa	Universidad de Las Palmas de Gran Canaria
Alfonso Ortega Giménez	Universidad de Zaragoza
António Teixeira	Universidade de Aveiro
Rubén San-Segundo Hernández	Universidad Politécnica de Madrid

Scientific Review Committee

Alberto Abad	L2F/Spoken Language Systems Laboratory
Jesús B. Alonso Hernández	Universidad de Las Palmas de Gran Canaria
Olatz Arregi Uriarte	Euskal Herriko Unibertsitatea
Plínio Barbosa	Universidade Estadual de Campinas
Daniela Braga	Voicebox
Juan M. Caballero Suárez	Universidad de Las Palmas de Gran Canaria
Fidel Cabrera Quintero	Universidad de Las Palmas de Gran Canaria
Valentín Cardeñoso Payo	Universidad de Valladolid
Paula Carvalho	Universidade de Lisboa
María José Castro Bleda	Universitat Politécnica de Valencia
Jan Cernocky	BTU (Brno)
Chin Hui Lee	Georgia Tech
Ricardo de Córdoba Herralde	Universidad Politécnica de Madrid
Carmen de la Mota Górriz	Universitat Autònoma de Barcelona
Laura Docío Fernández	Universidade de Vigo
Daniel Erro Eslava	Euskal Herriko Unibertsitatea
David Escudero Mancebo	Universidad de Valladolid
Rubén Fernández Pozo	Universidad Politécnica de Madrid
Javier Ferreiros López	Universidad Politécnica de Madrid
Julián Fierrez Aguilar	Universidad Autónoma de Madrid
Javier Franco Pedroso	Universidad Autónoma de Madrid
Ascensión Gallardo Antolín	Universidad Carlos III de Madrid
Carmen García Mateo	Universidade de Vigo
Elena García Quevedo	Universidad de Las Palmas de Gran Canaria
Juan Ignacio Godino Llorente	Universidad Politécnica de Madrid
Pedro Gómez Vilda	Universidad Politécnica de Madrid
Iván D. Guerra Moreno	Universidad de Las Palmas de Gran Canaria
Annika Hamalainen	Microsoft (MLDC)
Inma Hernáez Rioja	Euskal Herriko Unibertsitatea
Eduardo Hernández Pérez	Universidad de Las Palmas de Gran Canaria
Francisco Javier Hernando Pericas	Universitat Politècnica de Catalunya
Lluís Felip Hurtado Oliver	Universitat Politécnica de Valencia
Eduardo Lleida Solano	Universidad de Zaragoza
María Teresa López Soto	Universidad de Sevilla
Ramón López-Cózar Delgado	Universidad de Granada
Jordi Luque Serrano	Telefónica I+D
José B. Mariño Acebal	Universitat Politècnica de Catalunya
Sofía I. Martín González	Universidad de Las Palmas de Gran Canaria
Carlos David Martínez Hinarejos	Universitat Politécnica de Valencia
Manuel M. Medina Molina	Universidad de Las Palmas de Gran Canaria
Hugo Meinedo	Microsoft (MLDC)
Helena Moniz	INESC
Juan Manuel Montero Martínez	Universidad Politécnica de Madrid

Nicolas Morales Mombiela	Nuance
Antonio Moreno Sandoval	Universidad Autónoma de Madrid
Climent Nadeu Camprubi	Universitat Politècnica de Catalunya
Juan Luis Navarro Mesa	Universidad de Las Palmas de Gran Canaria
Eva Navas Cordón	Euskal Herriko Unibertsitatea
Géza Nemeth	BME, Hungary
Juan Nolazco Flores	Tecnológico de Monterrey, Mexico
Alfonso Ortega Giménez	Universidad de Zaragoza
Antonio Miguel Peinado Herreros	Universidad de Granada
Carmen Peláez Moreno	Universidad Carlos III de Madrid
José Luis Pérez Córdoba	Universidad de Granada
Ferrán Pla Santamaría	Universitat Politécnica de Valencia
Paulo Quaresma	Universidade de Évora
Pedro J. Quintana Morales	Universidad de Las Palmas de Gran Canaria
Daniel Ramos Castro	Universidad Autónoma de Madrid
Andreia Rauber	Universidade Católica de Pelotas
Antonio G. Ravelo García	Universidad de Las Palmas de Gran Canaria
José Adrián Rodríguez Fonollosa	Universitat Politècnica de Catalunya
Eduardo Rodríguez Banga	Universidade de Vigo
Luis Javier Rodríguez Fuentes	Microsoft (MLDC)
Joan Andreu Sánchez Peiró	Universitat Politécnica de Valencia
Emilio Sanchís Arnal	Universitat Politécnica de Valencia
Rubén San-Segundo Hernández	Universidad Politécnica de Madrid
Diana Santos	University of Oslo
Kepa Sarasola Gabiola	Euskal Herriko Unibertsitatea
Encarnación Segarra Soriano	Universitat Politécnica de Valencia
Mário Silva	Universidade de Lisboa
Alberto Simões	ESEIG/IPP
Richard Stern	Carnegie Mellon University
António Teixeira	Universidade de Aveiro
Javier Tejedor Noguerales	Universidad de Alcalá de Henares
Doroteo Torre Toledano	Universidad Autónoma de Madrid
Isabel Trancoso	IST/University of Lisbon & INESC-ID
Carlos Travieso González	Universidad de Las Palmas de Gran Canaria
María Amparo Varona Fernández	Euskal Herriko Unibertsitatea
Aline Villavicencio	Federal University of Rio Grande do Sul

The IRIS Project: A Liaison between Industry and Academia towards Natural Multimodal Communication

João Freitas[1,2], Sara Candeias[1], Miguel Sales Dias[1,3], Eduardo Lleida[4], Alfonso Ortega[4], António Teixeira[2], Samuel Silva[2], Cengiz Acarturk[5], Verónica Orvalho[6]

[1] Microsoft Language Development Center, Lisboa, Portugal
[2] Dep. Electronics Telecommunications & Informatics/IEETA, University of Aveiro, Portugal
[3] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal
[4] Universidad de Zaragoza, Spain
[5] Middle East Technical University, Turkey
[6] Faculdade de Ciências da Universidade do Porto, Portugal
[7] Instituto de Telecomunicações, Portugal

{t-joaof, t-sacand, midias}@microsoft.com,
{lleida,ortega}@unizar.es, {ajst,sss}@ua.pt,
acarturk@metu.edu.tr, veronica.orvalho@dcc.fc.up.pt

Abstract. This paper describes a project with the overall goal of providing a natural interaction communication platform accessible and adapted for all users, especially for people with speech impairments and elderly, by sharing knowledge between Industry and Academia. The platform will adopt the principles of natural user interfaces such as speech, silent speech, gestures, pictograms, among others, and will provide a set of services that allow easy access to social networks, friends and remote family members, thus contributing to overcome social-exclusion of people with special needs or impairments. Application of these features will be performed in the context of serious games, virtual reality environments and assisted living scenarios. The project will be executed in the scope of the Marie Curie Action Industry-Academia Partnerships and Pathways and will bring together the knowledge of five partners, from three different countries, Portugal, Spain and Turkey. This synergy will be based on transfer of knowledge mechanisms such as regular assignments and seminars, and is expected to have repercussions in scientific, technological, social and economic domains.

Keywords: Communication, Multimodal, Natural Interaction, Speech impairments, Knowledge Transfer.

1 Introduction

Communication is a process that facilitates integration into society, allowing humans to bond and to express their needs or desires. Thus, failing to do so may lead to social exclusion and, consequently, to psychological issues.

Speech is the main way for humans to communicate and in the last decades we have seen the rise of speech technologies, mainly in desktop and mobile devices, providing a more natural interaction with this type of machines. However, this kind of technologies is not yet adapted for users bearing speech impairments or users with more dissimilar speech patterns such as children or elderly.

In this context, focusing on the human communication process, we have created a project called IRIS, inspired in the ancient Greek goddess, who was a messenger of the gods. The overall goal of IRIS is to provide a natural interaction communication platform accessible and adapted for all users, especially for people with speech impairments and elderly in indoor scenarios, by sharing knowledge between Industry and Academia. Human-Computer interaction with this platform will adopt the principles of universal design and natural user interfaces such as speech, silent speech, gestures, gaze, tactile devices, pictograms, animated characters and personalized synthetic voices. The platform will provide a set of services that allow easy access to social networks, friends and remote family members, fighting social-exclusion of people with special needs or impairments. Application of these features will be performed in the context of serious games, virtual reality environments and assisted living scenarios. We will also explore the use of personalized avatars (that resemble the user) in asynchronous human-human and human-machine communications, in situations where the user is deprived of his/her voice and in scenarios where it is not possible to have a video signal transmission, due to low bandwidth or privacy reasons. Biometrics will complement the platform, in the sense that authentication and authorization are fundamental aspects for assuring secure access to personal information in a natural way. Figure 1 summarizes the research disciplines and potential outcomes from IRIS.

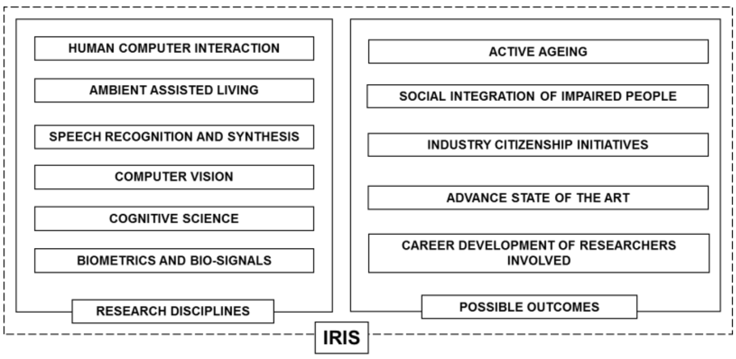


Fig. 1. Research disciplines and potential outcomes for IRIS.

This project will be carried out in the scope of the Marie Curie Actions framework under the Industry-Academia Partnerships and Pathways (IAPP) category. This research action aims at improving industry-academia collaboration in terms of research training, career development and knowledge sharing [1]. Thus, the proposed goals will need to rely on a successful transfer of knowledge between 5 partners, 2 from Industry and 3 from Academia. More details about the program can be found in [2].

The remains of this paper is structured as follows: Section 2 describes some of the target research topics and how they relate to IRIS; Section 3 presents the methodology adopted for this project, including how to leverage the mechanisms inherent to this type of projects (i.e. IAPP) to reach the designated goals; and finally, Section 4 describes the potential project impact from several perspectives.

2 Research Areas

IRIS is a multidisciplinary project that will leverage and advance the knowledge from multiple research fields. Overall, the main innovative aspects of this project are the following: a) provide an easy and natural way of accessing daily communication tasks through a multimodal interface that works for a diverse and distinct range of pathologies with particular emphasis on speech impairments and elderly people; b) analyze the impact of end-to-end communication scenarios with these interfaces, in virtual and real world scenarios, with users that otherwise were not able to conduct such tasks due to disabilities; c) create a solution and a set of resources that span and apply to multiple countries.

We aim at going beyond the state-of-the-art in a wide range of topics.These include elderly speech, silent speech interfaces, pictograms, voice based and multimodal biometrics, speech pathologies, and biological and psychophysical measures in the context of human-computer interaction (HCI) and cognitive science studies. These topics will be addressed on an individual basis and as a whole, with the ultimate goal of synchronizing all the advances into a single solution. In the sections below, we describe some of the related work in these areas and how they relate with IRIS.

2.1 Elderly-Computer Interaction

Elderly population individuals have developed resistance to conventional forms of HCI [3], like the keyboard and mouse, therefore making it necessary to test new natural forms of interaction such as speech, silent speech, touch and gestures. In addition, elder people often have difficulties with motor skills due to health problems such as arthritis. Therefore, proposing solutions that avoid small equipment, difficult to handle, may be an advantage over current solutions. It is also known that due to ageing, senses like vision become less accurate, hence difficulties in the perception of details or important information in conventional graphical interfaces may arise since current interfaces, most notably in the mobility area, are often not designed with these difficulties in mind.

There is also evidence that the European Union (EU) population is ageing rapidly [4]. This means that it is necessary to create solutions that allow overcoming the difficulties age brings to people who want to use new technologies in order to remain socially active. Elderly people who are connected to the world through the internet are less likely to become depressed and have greater probability of becoming socially integrated [5].

2.2 Speech Pathologies

Recent advances in machine learning and signal processing allow the detection of laryngeal pathologies through an automatic voice analysis. This is a promising field for speech therapists, mainly due to its non-invasive nature and objectivity for making decisions. Even the performance of automatic systems is still not perfect, thus they can be used as an additional source of "IRIS" information for other laryngoscopial exams [6]. Researching activity in this area is focused on: finding new features [7, 8], assessing quality of classification results [9], and researching newer classifier approaches [10].

In IRIS, a new vector space for voice pathology detection will be proposed, known as iVector space [11] that has become the state of the art in the fields of speaker and language recognition. This space is created as a factorization of the acoustic space in two terms. The first term accounts for the common information of the whole training database. In the second term, all sources of variability are taken into account and it is the one referred as iVector space or total variability space. Models for the normal and pathological classes will be trained by grouping iVectors from each of them separately and assuming they follow a Gaussian distribution. Speech pathology detection and classification will also be used on the IRIS project to gather speech information to build the user profile.

2.3 Silent Speech Interfaces

A Silent Speech Interface (SSI) performs Automatic Speech Recognition (ASR) in the absence of an intelligible acoustic signal and can be used as an HCI modality in high-background-noise environments such as living rooms, or in aiding speech-impaired individuals such as elderly persons [12]. By acquiring data from elements of the human speech production process – from glottal and articulators activity, their neural pathways or the central nervous system – 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 . Recently, conventional ASR systems rely only on acoustic information, making them susceptible to problems like environmental noise, privacy, information disclosure and also excluding users with speech impairments (e.g. persons who have undergone a laryngectomy) or elderly citizens for whom speaking requires a substantial effort. To tackle this problem in the context of ASR for Human-Computer Interaction, we envisage for IRIS a novel multimodal SSI.

In view of the objectives established for IRIS, we have chosen a set of modalities that combine less invasive with the some of the most promising approaches found in the state-of-the-art. As such, it is our aim to build a multimodal SSI that uses Video and Depth information (i.e. RGB-D), Ultrasonic Doppler sensing, and Surface Electromyography. By combining multiple modalities we expect to address some of the challenges listed by Denby et al. [1], such as robustness, sensor positioning and nasality.

2.4 Pictograms

The IRIS platform is being designed by the principle of "design-for-all" which includes the population with spoken language difficulties. Communication and language are essential to every human being, to relate to others, to learn, to enjoy and to participate in society. For this reason, all people, whether children, youth, adults and the elderly, who for whatever reason have not acquired or have lost a level sufficient to communicate satisfactorily speech need to use an augmentative and alternative communication system.

The Augmentative and Alternative Communication Systems (AAC) are different forms of expression to the spoken language, which aim to increase (augmentative) and/or compensate (alternative) communication difficulties of many people with disabilities. AAC includes various symbol systems, both graphics (photographs, drawings, pictograms, words or letters) and gestural (mime, gesture or sign language).

IRIS makes use of pictograms for non-verbal communication. Pictograms are used as both, input and output modalities, in the way that a user can write sentences using pictograms and the system can communicate with users by means of pictograms. Pictograms are one of the preferred means to obtain information of the environment for those people that take special benefits from the visual channel, such as people with Autism Spectrum Disorders (ASD). IRIS will make use of the pictogram set provided by the Aragonese Portal of Augmentative and Alternative Communication (http://www.arassac.es).

2.5 Voice based and multimodal biometrics

Voice based biometrics or speaker identification begun in the 60s, with the study of several pattern recognition and feature selection techniques and has been receiving an increasing amount of attention from the research community [11].

The most widely used speech features are inspired in the human speech production system, which can be decomposed in the vocal tract and the excitation source (glottal pulse or noise generator). Recently, non-segmental acoustic features have been combined with spectral ones such as Mel Frequency Cepstral Coefficients (MFCC) and Linear Prediction Coding (LPC), providing improved performance. Regarding the modelling side, the GMM-UBM (Gaussian Mixture Model-Universal Background Model) approach [13], has been recently substituted by the new techniques based on Joint Factor Analysis (JFA) [14] [15] solutions due to the limitations of the former for modelling channel variability in an appropriate way. Simpler and robust systems can be obtained with improved performance with respect to the JFA approach. Nevertheless, the accuracy of these systems can be considerably degraded if the quality of the speech signal is below the expected level or if the speech signal under analysis differs substantially from the speech samples used in the development process. Speech signal quality can be degraded due to several factors: low signal to noise ratio, saturation, acoustic artefacts, or presence of pathological speech. Therefore, the use of unimodal biometric systems presents limitations if the incoming data is noisy or if the variability is very high. Some other reasons to move to multimodal approaches are the potential non-universality of the unimodal variety (i.e. problems of speech production for some individuals, mobility impairments, etc.) or the robustness against spoof attacks.

In a multimodal biometric system, such as the one targeted in IRIS, the combination of the information coming from different sources of knowledge can be performed at different levels: Fusion at data or feature level, fusion at the match score level or fusion at the decision level. The multimodal approaches that integrate information at an early stage are usually preferred since they can offer improved performance and effectiveness [16]. Among the different modalities that can be used to build a multimodal biometric system, IRIS project will focus on the fusion of gesture biometrics with voice biometrics.

2.6 Multimodal Interfaces

Multimodal interfaces try to combine different modes of communicating, such as speech and gesture, in order to attain a more natural interaction and provide applications that are more engaging, exhibit more flexibility and contribute to improved accessibility, by different user groups in different scenarios, to tackle, for example, age-related impairments (e.g., loss of mobility, vision and hearing). Important aspects of multimodal interface development include: a) Modality Selection, guided by a full specification of requirements including a user profile, as complete as possible (addressing cognitive abilities, experience, physical characteristics, etc.), and characterizing the application domain, tasks and potential contexts of use [17]; b) Modality Mapping to different tasks and data types should be performed considering the characteristics of the human perceptual system and how each sense can respond to different task demands; c) Modality Combination is an important issue since assigning modalities to different tasks and data should not forget how modalities might interact, influencing how the user perceives information (e.g. to avoid sensory overload [18]).

The task of combining input modalities, to attain a single action is supported by d) Modality fusion engines [19]; e) System Adaptability also plays an important role for two main reasons: adapting to different user profiles (e.g., age, preferences and physical skills) in order to improve performance and satisfaction and adapting to changes in task and context [20]; f) Usability Evaluation of multimodal interfaces [21] should not forget some specific issues such as the importance of testing using context changes [22] and cognitive workload [23] conditions. The social impact [24] of the proposed systems must also be carefully assessed: the user might feel ashamed to use it in public if it somehow makes their difficulties or impairments more noticeable.

The main contributions of IRIS to the field of multimodal interfaces research can be described at three different levels: 1) Multimodal System, providing adequate/meaningful combinations of input and output modalities to support users in different communication tasks, adaptable to different usage contexts and user profiles, namely different pathologies and usage patterns; 2) Multidisciplinary approach provided by the different partners and their different areas of expertise allows approaching multimodal interfaces at different levels: software engineers, human computer interaction and usability experts, both the industry and academia points-of-view, and a set of infrastructures, namely a CAVE and an usability lab, allowing a strong emphasis on technically innovative approaches, usable/useful systems and market oriented solutions; 3) Design Options Validation: even though some empirical studies have been performed, this field clearly needs further contributions to validate design options on different application scenarios and, by doing that, contribute to increased knowledge to support further research and feed the industry with clear/proven guidelines to apply to new products. The conducted usability evaluations will provide valuable data to support or discard different design options leading to new or more specific guidelines for multimodal interfaces.

2.7 Biological and Psychophysical Measures

The broad range of recent research methods in HCI and cognitive science cuts across different research domains, such as cognitive psychology, educational psychology, psychophysics, and linguistics. These methods involve controlled experiments, questionnaires, in-depth interviews, focus groups, and usability inspection methods and cognitive modeling [25, 26]. Controlled experiments, being used widely to study human cognition and to evaluate human-computer interfaces for usability, measure different aspects of perceptual and cognitive processes by means of biological and psychophysical measures, such as response time, Electromyography (EMG: measures electrical activity produced by skeletal muscles) and Electro-dermal Activity (EDA: measures changes in the skin's ability to conduct electricity), eye/gaze tracking, and optical brain imaging (fNIR).

The basic motivation for studying eye movements is the relationship between eye movements and cognitive processes; in particular, the relationship between eye movements and visual attention. Eye tracking provides researchers with robust experimental data about online comprehension processes [27, 28]. Therefore, it has been widely used in HCI research since the past two decades. Eye movements, being indicators of cognitive processes, are used for the analysis of processing difficulties in visual displays, which in turn lead to guidelines and principles for design.

IRIS has the potential to contribute to the findings on the interdependency between the recent state-of-the-art biological and psychophysical methods, as well as on the investigation of the contribution of the complementary methods, i.e. interaction analysis and Quality of Experience (QoE) modelling, in evaluating multimodal interfaces.

3 Methodology

To support the framework behind this project it is necessary to define clear processes for knowledge transfer. Industry and Academia have different methods and different goals, thus a common ground between institutions needs to be found. Hence, more than defining which tasks lead to the stated objectives, we need to know how to execute these tasks using the resources available at IAPP projects and encompassing any restriction inherent to the institutions.

At an initial stage, it is important to have a clear idea of the knowledge held by each partner and how that knowledge is going to be integrated in the project. Thus, the first secondments are organized in way that there is an intersection between secondments from different institutions at the same destination. This method allows for researchers to share their experience and adequately plan future work. When the secondment ends, these researchers can take the acquired knowledge back to their institution and have a detailed overview of the project. For some areas, there is also the need to organize more specific courses given by experienced researchers, also open to the research community.

In terms of research plan, the work will be split into individual experiments/tasks that evolve towards the final objectives. The experiments will have multiple researchers assigned to them and should result in collaborative publications in an international conference, journal or book from the respective area.

Due to the multiple research disciplines encompassed by the project, a single research method (e.g. quantitative, qualitative) cannot be established. Therefore, these methods will be specified in a collaborative manner, by the researchers, before work on a specific task begins.

On a more generic level, the experiments will follow an iterative approach composed by the following stages: 1) problem identification; 2) problem assessment and state-of-the-art analysis; 3) development of a hypothesis/prototype; 4) application of test suite; 5) analysis and results comparison; 6) validation and conclusion formulation. To facilitate technological integration of all contributions, careful design and interoperability considerations are of paramount importance from the start of the project.

4 Impact

IRIS has potential repercussions in scientific, technological, social and economic domains, as depicted in Figure 2. From a scientific perspective IRIS aims to build the grounds for new ways of human computer-interaction and mediated human-human interaction. By combining multiple research disciplines and considering the defined objectives we expect not only to present novel ideas and solutions but also to make these applicable in the real world. As such, additional effort must be made towards realist but also innovative solutions. We also expect to create a strong impact in society, especially among those whose imparities somehow cause info-exclusion or avoid them from being more active in nowadays society. This aspect will be explored through existent partnerships with institutions and organizations that focus on social integration of people with impairments, senior universities that have daily contact with elderly citizens and special education schools that focus on the education of young students with special needs.

From an economical perspective IRIS has the potential to indirectly make the difference. By creating the mechanisms for a more informed and linked society, we expect to enable access not only to other people but also to products and goods available online. Likewise, we are opening the door for people that so far were not able to expose their ideas and skills to the community by lack of communication means. The involvement of both academia and industry in this project also allows establishing a bridge between research and concrete solutions that may strive into market.

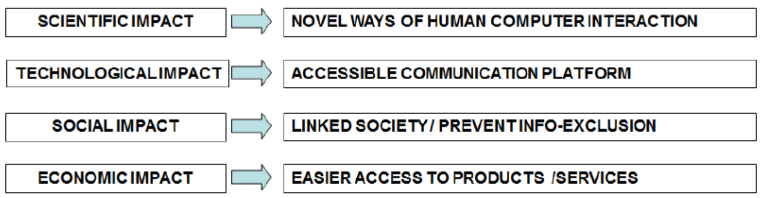


Fig. 2. Impact areas for IRIS

Acknowledgements. This work was funded by Marie Curie Actions IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP).

References

1. Commission, E.: Work Programme: Marie Curie Actions. (2013).
2. Commission, E.: Guide For Applicants: IAPP Call Specific. (2013).
3. Phang, C.W., Sutanto, J., Kankanhalli, A., Li, Y., Tan, B.C.Y., Teo, H.-H.: Senior citizens' acceptance of information systems: A study in the context of e-government services. IEEE Trans. Eng. Manag. 53, 555–569 (2006).
4. Commission, E.: Europe in figures – Eurostat yearbook 2010.
5. Cisek, E., Triche, K.: Depression and Social Support Among Older Adult Computer Users. 113th Annual Convention of the American Psychological Association (2005).
6. Godino-Llorente, J.I., Gomez-Vilda, P., Blanco-Velasco, M.: Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. Biomed. Eng. IEEE Trans. 53, 1943–1953 (2006).
7. Parsa, V., Jamieson, D.G.: Identification of pathological voices using glottal noise measures. J. speech, Lang. Hear. Res. 43, 469–485 (2000).
8. Gavidia-Ceballos, L., Hansen, J.H.L.: Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection. Biomed. Eng. IEEE Trans. 43, 373–383 (1996).
9. Tadeusiewicz, R., Wszolek, W., Modrzejewski, M.: The evaluation of speech deformation treated for larynx cancer using neural network and pattern recognition methods. Proc. EANN (1998).
10. Markaki, M., Stylianou, Y.: Using modulation spectra for voice pathology detection and classification. Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE. pp. 2514–2517. IEEE (2009).

11. Dehak, N., Kenny, P., Dehak, R., Dumouchel, P., Ouellet, P.: Front-end factor analysis for speaker verification. Audio, Speech, Lang. Process. IEEE Trans. 19, 788–798 (2011).
12. Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J.M., Brumberg, J.S.: Silent speech interfaces. Speech Commun. 52, 270–287 (2010).
13. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker models. Speech Audio Process. IEEE Trans. 3, 72–83 (1995).
14. Kenny, P., Ouellet, P., Dehak, N., Gupta, V., Dumouchel, P.: A study of interspeaker variability in speaker verification. Audio, Speech, Lang. Process. IEEE Trans. 16, 980–988 (2008).
15. Dehak, N., Dehak, R., Kenny, P., Brümmer, N., Ouellet, P., Dumouchel, P.: Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification. Interspeech. 9, 1559–1562 (2009).
16. Ross, A., Jain, A.: Multimodal biometrics: An overview. na (2004).
17. Obrenovic, Z., Abascal, J., Starcevic, D.: Universal accessibility as a multimodal design issue. Commun. ACM. 50, 83–88 (2007).
18. Hollender, N., Hofmann, C., Deneke, M., Schmitz, B.: Integrating cognitive load theory and concepts of human–computer interaction. Comput. Human Behav. 26, 1278–1288 (2010).
19. Atrey, P.K., Hossain, M.A., El Saddik, A., Kankanhalli, M.S.: Multimodal fusion for multimedia analysis: a survey. Multimed. Syst. 16, 345–379 (2010).
20. Kong, J., Zhang, W.Y., Yu, N., Xia, X.J.: Design of human-centric adaptive multimodal interfaces. Int. J. Hum. Comput. Stud. 69, 854–869 (2011).
21. Wechsung, I., Engelbrecht, K.-P., Kühnel, C., Möller, S., Weiss, B.: Measuring the Quality of Service and Quality of Experience of multimodal human–machine interaction. J. Multimodal User Interfaces. 6, 73–85 (2012).
22. Naumann, A.B., Wechsung, I., Hurtienne, J.: Multimodal interaction: A suitable strategy for including older users? Interact. Comput. 22, 465–474 (2010).
23. Leung, R., MacLean, K., Bertelsen, M.B., Saubhasik, M.: Evaluation of haptically augmented touchscreen gui elements under cognitive load. Proceedings of the 9th international conference on Multimodal interfaces. pp. 374–381. ACM (2007).
24. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 887–896. ACM (2010).
25. Cairns, P., Cox, A.L.: Research methods for human-computer interaction. Cambridge University Press (2008).
26. Sears, A., Jacko, J.A.: Human-computer interaction: Designing for diverse users and domains. CRC Press (2009).
27. Liversedge, S., Gilchrist, I., Everling, S.: The Oxford handbook of eye movements. Oxford University Press (2011).
28. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: A comprehensive guide to methods and measures. Oxford University Press (2011).