



CONFERENCE GUIDE

IberSPEECH 2012

VII Jornadas en Tecnología del Habla and III Iberian SLTech Workshop

21-23 NOVEMBER 2012
ESCUELA POLITÉCNICA SUPERIOR
UNIVERSIDAD AUTÓNOMA DE MADRID (SPAIN)

Organized by:



ATVS Biometric Recognition Group



Spanish Thematic Network on Speech Technology (RTTH)



ISCA Special Interest Group on Iberian Languages (SIG-IL)



Supported by:



Consorcio MAVIR

Proceedings published by:



Springer Communications in Computer and Information Science



VII Jornadas en Tecnologías del Habla and III Iberian SLTech Workshop

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

WELCOME MESSAGE

On behalf of the IberSPEECH 2012 Organizing Committee, the Spanish Thematic Network on Speech Technology (RTTH) and the ISCA-Special Interest Group on Iberian Languages (SIG-IL), it is our pleasure to welcome you to the IberSPEECH 2012 Conference, hosted by the ATVS Biometric Research Group, Universidad Autónoma de Madrid and held from 21 to 23 November 2012 in Madrid, Spain.

This conference is the result of the merging of two conferences: the “*Jornadas en Tecnología del Habla*” (Spanish Speech Technology Workshop) and the Iberian SLTech Workshop. The first one has been organized by the “*Red Temática en Tecnología 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 in Spain. The first Iberian SLTech Workshop was organized in Porto Salvo, Portugal, in 2009, by the Special Interest Group on Iberian Languages (SIG-IL, <http://www.il-sig.org/>) of the International Speech Communication Association (ISCA, <http://www.isca-speech.org>) and has been organized in conjunction with the “*Jornadas en Tecnología del Habla*” since 2010.

As a result, *IberSPEECH 2012: Joint “VII Jornadas en Tecnología del Habla” and III Iberian SLTech Workshop*, is probably 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 this edition) mainly from Spain and Portugal, and being also a natural meeting for researchers from Latin America. However, 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 from all around the world, including USA, United Kingdom, France, Japan, China, and Switzerland, among others.

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 had five different evaluations: Language Recognition, Audio Segmentation, Speech Synthesis, Search on Speech and Handwriting Recognition. 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 of the evaluations. Results of these evaluations will be presented in a Special Session on Wednesday 21 November 2012, and are also included in the online conference proceedings.

In this edition Springer has accepted to publish a selection of the papers submitted to this conference in a volume of Communications in Computer and Information Science (CCIS), in particular in volume 328 entitled *Advances in Speech and Language Technologies for Iberian Languages*. Of the approximately one hundred articles accepted for presentation at the conference, only 29 of the papers in the

regular paper track have been selected for this publication. This selection has been based on the scores and comments provided by our Scientific Review Committee which includes over 75 researchers from different institutions mainly from Spain, Portugal and Latin America, and to whom we would like to express our deepest gratitude. Each article in the regular paper track has been reviewed by three different reviewers and authors have had some time to address the comments before submitting the camera-ready paper. The articles in the regular paper track published in Springer CCIS have been organized into six oral sessions. The other 24 articles accepted in the regular paper track have been organized into two poster sessions the first two days of the conference.

Besides the regular paper track and the articles related to the ALBAYZIN Evaluations, the conference has special tracks for presenting Projects, Demos and Thesis. In this edition we have had 10 articles presenting projects, 11 articles presenting demos, and 8 articles presenting Ph.D. Thesis. All these will be presented in a Special Poster and Demos session the second day of the conference.

We will also have the pleasure of having three extraordinary keynote speakers: Dr. Jan "Honza" Cernocky (Brno University of Technology, BUT, Czech Republic), Dr. Philip Rose (Australian National University, Australia) and Dr. Pedro Moreno (Google Research, NY, USA).

The conference is mainly organized and supported by the Spanish Thematic Network on Speech Technology ("*Red Temática en Tecnología del Habla*") and the ISCA Special Interest Group on Iberian Languages (SIG-IL). We have also received support from the Universidad Autónoma de Madrid (UAM) and the Campus Internacional Excelencia UAM+CSIC, which have not only provided a fantastic venue for organizing the conference (the Escuela Politécnica Superior), but also financial support. Also, several companies have provided financial support for the conference, including Google, Microsoft and Telefónica (through Cátedra UAM-Telefónica). Last but not least, we have had financial support from the MA2VICMR consortium. Without the financial support of all of them this conference would simply have not been possible.

We would also like to thank Springer, and in particular to Alfred Hoffmann and Leonie Kunz, for the possibility of publishing a selection of the articles in this conference in a volume of Communications in Computer and Information Science and their help and great work in preparing it. This will help increasing the international impact of this conference.

Finally we would like to thank all the people that have been putting the best of their efforts in making this conference a reality: the Organizing Committee, the Local Organizing Committee, the organizations and companies that have provided financial support, the reviewers, the authors, the organizers and participants in the ALBAYZIN evaluations. We join all of them in wishing all attendees a fruitful and very enjoyable conference.

Doroteo Torre Toledano
Alfonso Ortega Giménez
António Teixeira

IberSPEECH 2012 Co-Chairs

IberSPEECH 2012 PROGRAM AT A GLANCE

Wednesday 21 November	Thursday 22 November	Friday 23 November
8:00 On-site registration (Building C, Sala Grados)		
8:30 Opening Ceremony (Building C, Sala Grados)	8:45 On-site registration (Building C, Sala Grados)	8:45 On-site registration (Building C, Sala Grados)
9:00 O1 - Oral Session: Speaker Characterization and Recognition Chair: Luis Hernández (Building C, Sala Grados)	9:00 O3 - Oral Session: Pathology Detection and Speech Characterization Chair: Ignacio Godino (Building C, Sala Grados)	9:00 O5 - Oral Session: Robustness in Automatic Speech Recognition Chair: Ascensión Gallardo (Building C, Sala Grados)
10:40 Coffee break	10:40 Coffee break	10:40 Coffee break
11:00 Keynote Talk Jan 'Honza' Cernocky Chair: Carmen García Mateo (Building A, Salón Actos)	11:00 Keynote Talk Philip Rose Chair: Joaquín González (Building A, Salón Actos)	11:00 Keynote Talk Pedro Moreno Chair: Javier Ortega (Building A, Salón Actos)
12:00 P1 - Poster Session Chair: António Teixeira (Building C, Hall)	12:00 P2 - Poster Session Chair: Inma Hernaez (Building C, Hall)	12:00 O6 - Oral Session: Applications of Speech & Language Technologies Chair: Asunción Moreno (Building C, Sala Grados)
13:30 Lunch (Plaza Mayor UAM)	13:30 Lunch (Plaza Mayor UAM)	13:40 Closing Ceremony (Building C, Sala Grados)
15:00 O2 - Oral Session: Audio and Speech Segmentation Chair: Eduardo Lleida (Building C, Sala Grados)	15:00 O4 - Oral Session: Dialogue and Multimodal Systems Chair: Javier Ferreiros (Building C, Sala Grados)	
16:40 Coffee break	16:40 Coffee break	
17:00 Albayzin Evaluations Chair: Daniela Braga (Building C, Sala de Grados and Hall)	17:00 Projects, Demos and Thesis Session Chair: Alberto Abad (Building C, Hall)	
18:30 RTTH Assembly (Building C, Sala Grados)	18:30 Optional guided visit to Museo Reina Sofía	
19:30 Welcome Reception (Building C, Hall)		
	21.00 Gala Dinner at restaurant Samarkanda	



VII Jornadas en Tecnologías del Habla and III Iberian SLTech Workshop

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

PRACTICAL INFORMATION

Conference Venues:



- (1) **Escuela Politécnica Superior, Building C, Sala de Grados:** Registration, Secretariat, Oral Sessions, Opening and Closing Ceremonies, Presentation of General Results of Albayzin Evaluations, RTTH Assembly.
- (2) **Escuela Politécnica Superior, Building C, Hall:** Poster Sessions, Albayzin Evaluations Special Session, Projects, Demos and Thesis Special Session, Welcome Reception, Coffee Breaks.
- (3) **Escuela Politécnica Superior, Building A, Salón de Actos:** Keynote Talks.
- (4) **Plaza Mayor, Restaurant:** Lunch.

Asking for help:

Members of the Local Organizing Committee will wear an orange badge to be easily identified. Please, contact them for anything you think they may help with.

Instructions for presenters:

Oral Presentations: If you haven't sent your presentation in advance, please contact the technical assistant present at the session at least 15 minutes before the session starts so that your presentation can be prepared.

Poster Presentations: Please put up your poster at least 30 minutes before the session starts and remove it just after the session ends.

Demos: We will provide a table and a poster holder for each demo. If you need anything else, please contact any of the local organizers urgently.

Internet Access:

Users of EduRoam can use it for Internet access in all the campus. In other cases, you can connect using the following WiFi network that covers all the conference venues.

Name of WiFi network (SSID):	IberSpeech2012
Password:	Speech%20!2

Lunch:

Lunch will be at the self-service restaurant located at Plaza Mayor. You should receive in your personal documentation three tickets for lunch corresponding to the three days of the conference. You should take these tickets with you and give them as payment for your lunch. Please, take into account that each ticket is valid for a particular person and a particular date. If you are a vegetarian or have special requirements for lunch, please contact the conference secretariat.

Where to work individually:

If you need to work individually, a good place to do it is at the library of the Escuela Politécnica Superior. The library is located on the fourth floor of Building A.

Where to work in groups:

If you need to have a meeting, please contact our secretariat and we will try to reserve a meeting room for you. We cannot guarantee the availability of meeting rooms.

Cafeteria:

Coffee breaks will be served in the morning and afternoon, but if you need anything else, there is a cafeteria in Building A, at the other extreme of the building from the Salón de Actos.



**VII Jornadas en Tecnologías del Habla
and III Iberian SLTech Workshop**

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

TECHNICAL PROGRAM

O1 - Oral Session: Speaker Characterization and Recognition

Date: Wednesday 21 November

Location: Building C, Sala de Grados.

Hour: 9:00 - 10:40

Chairperson: Luis Hernández Gómez.

- O1-1: Reliability Estimation of the Speaker Verification Decisions Using Bayesian Networks to Combine Information from Multiple Speech Quality Measures, *Jesus Villalba, Eduardo Lleida Solano, Alfonso Ortega Giménez, and Antonio Miguel Artiaga.*
 - O1-2: On the use of Total Variability and Probabilistic Linear Discriminant Analysis for Speaker Verification on Short Utterances, *Javier González Domínguez, Rubén Zazo, and Joaquin González-Rodríguez.*
 - O1-3: Cepstral Trajectories in Linguistic Units for Text-Independent Speaker Recognition, *Javier Franco Pedroso, Fernando Espinoza, and Joaquin González-Rodríguez.*
 - O1-4: Improving the Quality of Standard GMM-based Voice Conversion Systems by Considering Physically Motivated Linear Transformations, *Tudor-Cătălin Zorilă, Daniel Erro Eslava, and Inma Hernaez Rioja.*
 - O1-5: Evaluation of a New Beam-Search Formant Tracking Algorithm in Noisy Environments, *Dayana Ribas Gonzalez, Jose Enrique Garcia Lainez, Antonio Miguel Artiaga, Alfonso Ortega Giménez, Eduardo Lleida Solano, and Jose Ramon Calvo de Lara.*
-

Keynote Talk: Jan "Honza" Černocký

Date: Wednesday 21 November

Location: Building A, Salón de Actos.

Hour: 11:00 - 12:00

Chairperson: Carmen García Mateo.

Title: Building speech recognizers with limited resources

Speaker: Jan Honza Černocký

With contributions of: Martin Karafiát, Miloš Janda, Mirko Hannemann, Karel Veselý, František Grézl, Ekaterina Egorova

Abstract:

Research groups and industry have reached excellent results in automatic speech recognition (ASR) of language with abundant speech and text resources. There is however a steady need for recognition, keyword spotting and spoken term detection systems in languages that lack these resources.

This talk will summarize the work Brno University of Technology (BUT) speech group has done so far in this field. Supported by a project of Czech Ministry of Trade and Commerce and by U.S. IARPA BABEL program, we have investigated approaches that allow for re-using data from well represented languages to obtain better results in the recognition of low-resource ones: in feature extraction, we have worked on language independency of nowadays very popular bottle-neck features generated by neural networks. We have also studied sub-space Gaussian Mixture Models (SGMMs) and Regional-Dependent Linear Transforms (RDLT) for acoustic modeling and adaptation, respectively. Finally, we will present several ideas in automatic building of pronunciation dictionaries.

Speaker Bio:

Jan "Honza" Černocký, <http://www.fit.vutbr.cz/~cernocky> (Ing. [MS] 1993 Brno University of Technology (BUT); Dr. [PhD] 1998 Université Paris XI and BUT) was with the Institute of Radio-electronics, BUT (Faculty of Electrical Engineering and Computer Science) as assistant professor from 1997. Since February 2002, he is with the Faculty of Information Technology (FIT), BUT as Associate Professor (Doc.) and Head of the Department of Computer Graphics and Multimedia. With Dr. Burget and Prof. Hermansky, he is leading the BUT Speech@FIT research group. Dr.

Cernocky supervises several PhD students. He has been involved with several European projects: SPEECHDAT-E (4th FP, technical coordination), SpeeCon, Multimodal meeting manager (M4, both 5th FP), Augmented Multiparty interaction (AMI, 6th FP), Augmented Multiparty Interaction with Distance Access (AMIDA, 6th FP), and CareTaker (6th FP). He was BUT's principal investigator in EC-sponsored MOBIO project (7th FP) and in projects supported by Czech Ministries of Industry and Interior. He is currently coordinating a project supported by Technology Agency of the Czech Republic (TACR) and supervises projects funded by U.S. Government.

His research interests include signal processing and speech processing (speaker and language recognition, keyword spotting and spoken term detection). He is author or co-author of more than 50 papers in journals and at conferences. He has served as reviewer for conferences and journals, including IEEE Transactions on Speech, Audio and Language Processing. He is on the scientific board of FIT, scientific board of Text-Speech-Dialogue conference, editorial board of the journal Radioengineering, on the board of Czechoslovak section of IEEE, and on the IEEE SLTC technical committee. He served as general co-chair of ICASSP 2011 in Prague.

As faculty member of FIT BUT, Dr. Cernocky is also active in teaching, he is responsible for signal-processing, pattern recognition, speech and natural language processing courses in all levels of studies (bachelor, master, doctoral).

Dr. Cernocky is a senior member of IEEE and member of ISCA.

P1 - Poster Session

Date: Wednesday, 21 November

Location: Building C, Hall.

Hour: 12:00 - 13:30

Chairperson: António Teixeira.

- P1-1: Enhanced Amharic Speech Recognition System, *Abraham Zewoudie and Sebsbie Hailemariam* (Retired from conference).
 - P1-2: Audio Encoding for Heart and Breath Sounds Acquired with Digital Stethoscope, *Ignacio Palacios Santos, Doroteo Torre Toledano, and Andrés Martínez Fernández*.
 - P1-3: Morphological Processing of a Dynamic Compressive Gammachirp Filterbank for Automatic Speech Recognition, *Joyner Cadore, Carmen Peláez-Moreno, and Ascensión Gallardo Antolín*.
 - P1-4: Manipulation of Unvoiced Fricatives for Improved Discrimination by Hearing Impaired, *Rami Cohen, Oren Ierushalmi, Yizhar Lavner, Michal Genussov, and Ami Steiner*.
 - P1-5: Multiclass Pitch Accent Classification for Assisting Manual Prosodic Labeling, *David Escudero Mancebo, Francisco Vicaíno-Ortega, César González-Ferreras, Carlos Vivaracho-Pascual, Mercedes Cabrera-Abreu, Eva Estebas-Vilaplana, and Valentín Cardeñoso Payo*.
 - P1-6: Language Identification based on a Discriminative Text Categorization Technique, *Miguel Caraballo, Ricardo de Cordoba Herralde, Luis D'Haro, Rubén San-Segundo and José Pardo*.
 - P1-7: Binary Position Assignment of Two Known Simultaneous Acoustic Sources, *Rupayan Chakraborty, Climent Nadeu Camprubi, and Taras Butko*.
 - P1-8: Emotions recognition using binary fingerprints, *Xavier Anguera, Esperança Movellan, and Miquel Ferrarons*.
 - P1-9: Measuring Acoustic Reduction in Feature Space, *Henrik Schulz, and José Adrián Rodríguez Fonollosa*.
 - P1-10: A preliminary approach to forensic speaker recognition using phonemes, *Pedro Univaso, Miguel Martínez Soler, Diego Evin, and Jorge Gurlekian*.
 - P1-11: Spectral Envelope Representation using Sums of Gaussians, *Anderson Machado, Antonio Bonafonte Cávez, and Marcelo Queiroz*.
 - P1-12: Towards a Robust Dynamic Frequency Warping Text-Independent Voice Conversion System, *Tudor-Cătălin Zorilă, Daniel Erro, Yannis Stylianou, and Inma Hernáez*.
-

O2 - Oral Session: Audio and Speech Segmentation

Date: Wednesday, 21 November

Location: Building C, Sala de Grados.

Hour: 15:00 - 16:40

Chairperson: Eduardo Lleida Solano.

- O2-1: On the Influence of Automatic Segmentation and Clustering in Automatic Speech Recognition, *Paula Lopez-Otero, Laura Docio-Fernandez, Carmen Garcia-Mateo, and Antonio Cardenal-Lopez.*
 - O2-2: Preliminary Results of Alignment of Text and Audio in News and Song, *Darwin Patricio Córdova Lucero and Doroteo Torre Toledano.*
 - O2-3: Aligning very long speech signals to bilingual transcriptions of parliamentary sessions, *German Bordel, Mikel Penagarikano, Luis Javier Rodríguez Fuentes, and María Amparo Varona Fernández.*
 - O2-4: Factor Analysis Segmentation and Classification in Broadcast News Domain, *Diego Castán, Alfonso Ortega Giménez, and Eduardo Lleida Solano.*
 - O2-5: Prosodic and Phonetic Features for Speaking Styles Classification and Detection, *Arlindo Veiga, Dirce Celorico, Jorge Proença, Sara Candeias, and Fernando Perdigão.*
-

ALBAYZIN Evaluation Special Session

Date: Wednesday, 21 November **Location:** Building C, Sala de Grados and Hall.

Hour: 17:00 - 18:30

Chairperson: Eduardo Lleida Solano.

(Presentation of general results at Sala de Grados, followed by poster presentation at Hall)

LANGUAGE RECOGNITION EVALUATION

- ANHYT, *Fusheng Ba, Xuefei Jia, Zhihua Jin, Rui Xing, and Xu Ji.*
- GTH, *Luis Fernando D'Haro and Ricardo de Córdoba.*
- L2F, *Alberto Abad.*
- MSMP-LaBRI, *Jean Luc Rouas.*
- THU-EE, *Jia Liu, Liang He, Jiaming Xu, Weiwei Liu, and Wei-Qiang Zhang.*
- UVigo-GTM, *Paula Lopez-Otero, Laura Docio-Fernandez, and Carmen Garcia-Mateo.*
- ViVoLab UZ, *Paula Lopez-Otero, Laura Docio-Fernandez, and Carmen Garcia-Mateo.*

AUDIO SEGMENTATION EVALUATION

- Aholab, *David Tavares, Eva Navas, Daniel Erro, and Ibon Saratxaga.*
- CAIAC-UAM, *Héctor Delgado and Javier Serrano.*
- GTTS, *Luis Javier Rodríguez-Fuentes, Mikel Penagarikano, Amparo Varona, Mireia Diez and Germán Bordel.*
- UVigo-GTM, *Paula Lopez-Otero, Laura Docio-Fernandez, and Carmen Garcia-Mateo.*
- PRHLT-UPV, *J. A. Silvestre-Cerdá, A. Giménez, J. Andrés-Ferrer, J. Civera and A. Juan.*

SPEECH SYNTHESIS EVALUATION

- Aholab, *David Tavares, Eva Navas, Daniel Erro, and Ibon Saratxaga.*
- Simple4All, *Jaime Lorenzo-Trueba, Oliver Watts, Roberto Barra-Chicote, Junichi Yamagishi, Simon King and Juan M. Montero.*

SEARCH-ON-SPEECH EVALUATION

- EliRF, *Emilio Sanchis, Lluís-F. Hurtado, Jon A. Gómez, Marcos Calvo and Raúl Fabra.*
 - GTTS, *Amparo Varona, Mikel Penagarikano, Luis Javier Rodríguez-Fuentes, Germán Bordel and Mireia Diez.*
 - Telefonica Research, *Xavier Anguera.*
 - ViVoLab UZ, *Jaime Lorenzo-Trueba, Oliver Watts, Roberto Barra-Chicote, Junichi Yamagishi, Simon King and Juan M. Montero.*
-

O3 - Oral Session: Pathology Detection and Speech Characterization

Date: Thursday, 22 November

Location: Building C, Sala de Grados.

Hour: 9:00 - 10:40

Chairperson: Juan Ignacio Godino Llorente.

- O3-1: Voice Pathology Detection on the Saarbrücken Voice Database with Calibration and Fusion of Scores Using MultiFocal Toolkit, *David Martínez González, Eduardo Lleida Solano, Alfonso Ortega Giménez, Antonio Miguel Artiaga, and Jesus Villalba.*
 - O3-2: Score Level versus Audio Level Fusion for Voice Pathology Detection on the Saarbrücken Voice Database, *David Martínez González, Eduardo Lleida Solano, Alfonso Ortega Giménez, and Antonio Miguel Artiaga.*
 - O3-3: Using HMM to detect speakers with severe Obstructive Sleep Apnoea Syndrome, *Ana Montero Benavides, José Luis Blanco, Alejandra Fernández, Rubén Fernandez Pozo, Doroteo Torre Toledano, and Luis Hernández Gómez.*
 - O3-4: Acoustic analysis of European Portuguese oral vowels produced by children, *Catarina Oliveira, Maria Manuel Cunha, Samuel Silva, António Teixeira, and Pedro Couto.*
 - O3-5: Impact of age in ASR for the elderly: preliminary experiments in European Portuguese, *Thomas Pellegrini, Isabel Trancoso, Annika Hamalainen, António Calado, Miguel Dias, and Daniela Braga.*
-

Keynote Talk: Philip Rose

Date: Thursday 22 November

Location: Building A, Salón de Actos.

Hour: 11:00 - 12:00

Chairperson: Joaquín González Rodríguez.

Title: Not too bad? – Forensic Voice Comparison in Theory and Reality.

Summary:

In forensic voice comparison, speech recordings from an unknown voice, usually of an offender, are compared with recordings from a known voice, usually the suspect. The aim, of course, is to help the trier-of-fact decide whether the suspect has said the incriminating speech.

For some time now, the use of a likelihood ratio has been both theoretically recognised as the correct logical framework for the evaluation of forensic evidence, and implemented as a matter of course in some areas, e.g. DNA profiling. It is logically correct, since by Bayes' Theorem a posterior probability – like “it is highly likely the suspect said the incriminating speech” – cannot be estimated absent prior odds, to which the expert is not usually privy. Since a posterior may well impinge on considerations of ultimate issue, the use of a likelihood ratio may also be the legally correct option.

The use of likelihood ratios in forensic voice comparison was an idea whose time came around the beginning of the new century, when its efficacy first began to be demonstrated both with automatic and traditional phonetic features (Morrison 2009a). The results from now well over a decade's extensive, and continuing, research testing have shown that the approach works rather well – same-speaker speech samples can for example be rather well discriminated from different-speaker speech samples on the basis of their LRs (although in real cases discrimination is not the appropriate model) – and it has been shown that the approach can emulate the DNA gold-standard (Gonzalez-Rodriguez et al. 2007).

However, despite these results, and the many explanatory texts written for the legal profession (e.g. Robertson & Vignaux 1995, Aitken et al. n.d.), it is safe to say the idea has not yet caught on. Although forensic voice comparison is experiencing a remarkable paradigm shift in theory, in reality this is only in a few small areas in the world (appropriately for this conference, Spain and Australia); and even then only sporadically and in conditions where it is not at all clear that the interested parties actually understand the concepts. This is partly because of the inherent conservatism of the legal profession, and partly because, it appears, many of its practitioners, if they are aware of the approach at all, find it difficult to understand. For example, in a recent set of Australian draft standards, representatives of organizations including *The Australian Attorney-General's Department, The Australian and New Zealand Forensic Science Society, The University of New South Wales' Expertise, Evidence and Law Program, and The New South Wales Bar Association* maintained that the “Interpretation [of forensic evidence] includes answering the question as to whether or not ... items share a common origin” (Standards Australia 2012). Not just the legal profession finds it difficult, however. There have been many objections to the implementation of the approach – some well-founded and some simply misconceived – among many, mostly UK, practitioners (Morrison 2009b). And perhaps the best-known phonetics text-book (Ladefoged 2006)

defines the likelihood ratio as "... the likelihood that the two voices in question are the same as compared with the likelihood that they are different." thus confusing it with the prior odds.

In my talk, I want to first rehearse the conceptual background to likelihood ratio-based forensic voice comparison, and say what I see as the two immense barriers to the acceptance of the approach where it matters – in the courtroom. I then describe a real case where likelihood ratio-based forensic voice comparison played a role in the successful prosecution of a \$150 million telephone fraud. This will give an idea of what the approach actually involves. The case I describe is important for several reasons unrelated to the size of the sum of money involved. It remains the only one where likelihood ratio-based forensic voice comparison has been received in an Australian court, and is a nice example of where traditional phonetic features, as opposed to automatic features, can come into their own. Finally it is important because it suggests that at least one of my barriers – the impasse at the boundary between likelihood ratio and its matrix concept of Bayes' Theorem – may, in the reality of the courtroom as opposed to the theory of our research laboratories, be imaginary.

Aitken, C.G.G., Roberts, P., & Jackson, G. no date ***Fundamentals of Probability and Statistical Evidence in Criminal Proceedings – Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses***, Royal Statistical Society.

Gonzalez-Rodriguez J., Rose P., Ramos, D., Torre, D. & Ortega-Garcia, J. 2007 "Emulating DNA: Rigorous Quantification of Evidential Weight in Transparent and Testable Forensic Speaker Recognition", ***IEEE Trans. on Audio Speech and Language Processing***, 15(7): 2104 – 2115.

Ladefoged, P. 2006 ***A Course in Phonetics***, 5th ed. Thomson.

Morrison, G.S. 2009a "Forensic voice comparison and the paradigm shift", ***Science & Justice***, 49: 298–308.

Morrison, G.S. 2009b "Comments on Coulthard & Johnson's (2007) portrayal of the likelihood-ratio framework", ***Australian Journal of Forensic Sciences***, 41: 155-161.

Robertson, B., & Vignaux, T. 1995 ***Interpreting Evidence***, Wiley.

Standards Australia 2012 "Forensic Analysis Part 3: Interpretation", Draft for Public Comment, DR AS 5388.3 <http://infostore.saiqglobal.com/store/Details.aspx?ProductID=1530224>.

Speaker Bio:

Phil Rose is adjunct associate professor in the School of Language Studies at the Australian National University, where he taught Phonetics and Chinese Linguistics for thirty years. He studies phonetics and phonology, and is an expert on tone languages and on Chinese and Chinese dialects. He is also an expert on forensic voice comparison and in his 2002 and 2003 books pioneered the application of the Likelihood Ratio of Bayes' Theorem to traditional forensic voice comparison, thus bringing it in line with forensic DNA profiling. He has done forensic voice comparison case-work on Australian English and varieties of Chinese for nearly 20 years. He has been a British Academy visiting professor at the University of Edinburgh's ***Joseph Bell Centre for Forensic Statistics and Legal Reasoning***, is chairman of the Forensic Speech Science Committee of the ***Australasian Speech Science & Technology Association*** and a member of the ***Australian Academy of Forensic Sciences***. His attempt to retire from teaching in 2009 failed and in the first half of 2012 he was visiting professor at the Hong Kong University of Science and Technology, where he taught courses on ***Chinese Phonetics*** and ***Forensic Voice Comparison in Cantonese***.

P2 - Poster Session

Date: Thursday, 22 November

Location: Building C, Hall.

Hour: 12:00 - 13:30

Chairperson: Inma Hernández Rioja.

- P2-1: Rule-Based Pronunciation Modeling for Resource-Scarce Varieties, *Simone Ashby, José Pedro Ferreira, Sílvia Barbosa, and Catarina Silva.*
 - P2-2: A quantitative study of disfluencies in formal, informal and media spontaneous speech in Spanish, *Antonio Moreno Sandoval, Leonardo Campillos Llanos, and Doroteo Torre Toledano.*
 - P2-3: A Computational Morphology for Spanish Sign Language Nominal Inflection, *Jordi Porta, Fernando Jesús López Colino, and José Colás Pasamontes.*
 - P2-4: Modeling a Spoken Dialog System using POMDPs and Agenda-based User Simulation, *David Griol, José Manuel Molina, Araceli Sanchis de Miguel, and Zoraida Callejas.*
 - P2-5: Methodology for developing a Speech into Sign Language Translation System in a New Semantic Domain, *Verónica López-Ludeña, Rubén San-Segundo, Carlos González, Juan López, and José Pardo.*
 - P2-6: Architecture for Text Normalization using Statistical Machine Translation techniques, *Verónica López-Ludeña, Rubén San-Segundo, Juan Manuel Montero, Roberto Barra-Chicote, and Jaime Lorenzo.*
 - P2-7: GenProso: a parametric prosody prediction module for text-to-speech applications, *Juan María Garrido.*
 - P2-8: MAVIR: a corpus of spontaneous formal speech in Spanish and English, *Leonardo Campillos Llanos and Antonio Moreno Sandoval.*
 - P2-9: Velar Movement in European Portuguese Nasal Vowels, *Paula Martins, Catarina Oliveira, Samuel Silva, and António Teixeira.*
 - P2-10: Subjective Evaluation of a High Quality Text-to-Speech System for Argentine Spanish, *Jorge Gurlekian, Christian Cossio Mercado, Humberto Torres, and M. Elena Vaccari.*
 - P2-11: Design of a message verification tool to be implemented in CALL systems, *Igor Odriozola, Inma Hernández, and Eva Navas.*
 - P2-12: Vocalic Filler Characterization using Acoustical Properties Derived from Cepstrum, *Arlindo Veiga, Jorge Proença, Dirce Celorico, Sara Candeias, and Fernando Perdigão.*
-

O4 - Oral Session: Dialogue and Multimodal Systems

Date: Thursday, 22 November

Location: Building C, Sala de Grados.

Hour: 15:00 - 16:40

Chairperson: Javier Ferreiros López.

- O4-1: Mutual Information and Perplexity based Clustering of Dialogue Information for Dynamic Adaptation of Language Models, *Juan Lucas-Cuesta, Fernando Fernández, Tirso Moreno, and Javier Ferreiros López.*
 - O4-2: A multilingual SLU system based on semantic decoding of graphs of words, *Marcos Calvo, Lluís Hurtado, Fernando García Granada, and Emilio Sanchís Arnal.*
 - O4-3: Merging intention and emotion to develop adaptive dialogue systems, *Zoraida Callejas, David Griol, and Ramón López-Cózar Delgado.*
 - O4-4: Language Technology for Handwritten text Recognition, *Alejandro Toselli, Nicolas Serrano, Adrià Giménez-Pastor, Ihab Khoury, Alfons Juan, and Enrique Vidal.*
 - O4-5: Character-based Handwritten Text Recognition of Multilingual Documents, *Miguel del Agua, Nicolás Serrano Martínez-Santos, Jorge Civera, and Alfons Juan.*
-

Projects, Demos and Thesis Special Session

Date: Thursday, 22 November

Location: Building C, Hall.

Hour: 17:00 - 18:30

Chairperson: Alberto Abad.

PROJECTS (Poster Presentation)

- Pr-1: Architecture for Text Sign Localization and Recognition, *Arturo Romero, José Moreno, Fernando Fernández, Juan Manuel Montero, and Rubén San-Segundo.*
- Pr-2: TexAFon: a multilingual text processing tool for text-to-speech applications, *Juan María Garrido, Silvia Rustullet, Yesika Laplaza, Montserrat Marquina, and Claudia Schoenfelder.*
- Pr-3: The First European Portuguese Elderly Speech Corpus, *Annika Hamalainen, Fernando Miguel Pinto, Miguel Dias, Ana Júdice, João Freitas, Carlos Pires, Vitor Duarte Teixeira, António Calado, and Daniela Braga.*
- Pr-4: Let's Speak Quechua: The Implementation of a Text-to-Speech System for the Incas' Language, *Richard Alexander Castro Mamani, John Edgar Vargas Muñoz, and Juan Cruz.*
- Pr-5: Cotovía: an open source TTS for Galician and Spanish, *Eduardo Rodríguez Banga, Carmen García-Mateo, Francisco Méndez Pazó, Manuel González González, and Carmen Magariños Iglesias.*
- Pr-6: The Glissando project: Building and exploitation of a large bilingual annotated corpus for multidisciplinary prosodic analysis and applications, *Juan María Garrido, David Escudero Mancebo, and Lourdes Aguilar.*
- Pr-7: The CASMACAT Project: The Next Generation Translator's Workbench, *Daniel Ortiz-Martínez, Germán Sanchis-Trilles, Francisco Casacuberta Nolla, Vicent Alabau, Enrique Vidal, José-miguel Benedí, Jesús González-Rubio, Alberto Sanchís, and Jorge González.*
- Pr-8: Google's Voice Search in Iberian Languages, *Ignacio Lopez Moreno, Fadi Biadisy, Eugene Weinstein, and Pedro Moreno.*
- Pr-9: transLectures, *Alfons Juan, Joan Albert Silvestre, Miguel del Agua, Gonçal Garcés, Guillem Gascó, Adrià Giménez-Pastor, Adrià Martínez, Alejandro Pérez, Isaías Sánchez, Nicolás Serrano Martínez-Santos, Rachel Spencer, Juan Daniel Valor Miró, Jesús Andrés-Ferrer, Jorge Civera, and Alberto Sanchís.*
- Pr-10: META-NET. Official Languages of Spain in the Digital Age, *Asunción Moreno Bilbao, Núria Bel, Maite Melero, Carmen García-Mateo, Inma Hernaez Rioja, Sergio Oller, aljoscha Burchardt, Kathrin Eichler, Georg Rehm, and Hans Uszkoreit.*

DEMOS (Demo + Optional Poster Presentation)

- Dem-1: Hermes: a mobile distributed speech translation system, *Emilio Granell, Carlos David Martínez Hinarejos, and Vicent Tamarit.*
- Dem-2: COPS: a computer programming tool to cope with functional diversity, *Carlos David Martínez Hinarejos, Santiago Sánchez-Alepuz, and Natividad Prieto-Sáez.*
- Dem-3: GTMTranscriber: a Subtitling Program for Broadcast News in Galician Language, *Antonio Cardenal-Lopez, Carmen Garcia-Mateo, Alberto Espiña, Paula Lopez-Otero, and Laura Docio-Fernandez.*
- Dem-4: An on-line system adding subtitles and sign language to Spanish audio-visual content, *Jordi Porta, Fernando Jesús López Colino, Javier Tejedor, and José Colás Pasamontes.*
- Dem-5: EntrenaVoz. Videogame feedback system for speech training and rehabilitation, *Victor Osma-Ruiz, Nicolás Sáenz-Lechón, Juana Gutiérrez-Arriola, Ruben Fraile, and Juan Ignacio Godino Llorente.*
- Dem-6: Interactive and User-Adapted Voice Portals for the City Council, *David Griol, María García Jiménez, José Manuel Molina, and Araceli Sanchis de Miguel.*
- Dem-7: The MovieScript System: Integration of Statistical Methodologies and the VoiceXML standard to develop spoken dialog systems, *David Griol, José Manuel Molina, and Araceli Sanchis de Miguel.*
- Dem-8: VITHEA: On-line word naming therapy in Portuguese for aphasic patients exploiting automatic speech recognition, *Anna Pompili, Pedro Fialho, and Alberto Abad.*
- Dem-9: Learning Portuguese with Speech Technologies, *Isabel Trancoso, Thomas Pellegrini, André Silva, Rui Correia, Nuno Mamede, and Jorge Baptista.*
- Dem-10: Aromo: Argentine Spanish TTS System, *Humberto Torres, Jorge Gurlekian, and Christian Cossio Mercado.*
- Dem-11: Search Engine for Multilingual Audiovisual Contents, *José David Pérez, Antonio Bonafonte Cávez, Antonio Cardenal, Marta Ruiz Costa-jussa, José Adrián Rodríguez Fonollosa, Asunción Moreno Bilbao, Eva Navas Cordón, and Eduardo Rodríguez Banga.*

Ph.D. THESIS (Poster Presentation)

- Th-1: Application of Speech Technology in the Education of Children's Altered Voice, *William Ricardo Rodríguez Dueñas*.
 - Th-2: Advances in Fully-Automatic and Interactive Phrase-based Statistical Machine Translation, *Daniel Ortiz-Martínez*.
 - Th-3: Session Variability Compensation in Automatic Speaker and Language Recognition, *Javier González Domínguez*.
 - Th-4: Detection and Handling of Overlapping Speech for Speaker Diarization, *Martin Zelenák and Francisco Javier Hernando Pericas*.
 - Th-5: Phase in the harmonic models of the speech signal: strategies for representation, processing and applications, *Ibon Saratxaga Couceiro and Inma Hernáez*.
 - Th-6: Building task-oriented machine translation systems, *Germán Sanchis-Trilles and Francisco Casacuberta Nolla*.
 - Th-7: Robust Diarization for Speaker Characterization, *Carlos Vaquero and Alfonso Ortega Giménez*.
 - Th-8: Perceptual Optimization of Unit-Selection Text-to-Speech Synthesis Systems by means of active interactive Genetic Algorithms, *Lluís Formiga and Francesc Alías*.
-

O5 - Oral Session: Robustness in Automatic Speech Recognition

Date: Friday, 23 November

Location: Building C, Sala de Grados.

Hour: 9:00 - 10:40

Chairperson: Ascensión Gallardo Antolín.

- O5-1: A robust pitch extractor based on DTW lines and CASA with application in noisy speech recognition, *Juan Morales Cordovilla, Pablo Cabañas-Molero, Antonio Miguel Peinado Herreros, and Victoria Eugenia Sánchez Calle.*
 - O5-2: Speech Denoising Using Non-Negative Matrix Factorization with Kullback-Leibler Divergence and Sparseness Constraints, *Jimmy Ludeña Choez and Ascensión Gallardo Antolín.*
 - O5-3: MMSE feature reconstruction based on an occlusion model for robust ASR, *Jose Gonzalez, Antonio Miguel Peinado Herreros, and Angel Gomez.*
 - O5-4: Automatic Speech Recognition based on Ultrasonic Doppler Sensing for European Portuguese, *João Freitas, António Teixeira, Francisco Vaz, and Miguel Dias.*
 - O5-5: Pairwise Likelihood Normalization-Based Channel Selection for Multi-Microphone ASR, *Martin Wolf and Climent Nadeu Camprubi.*
-

Keynote Talk: Pedro Moreno

Date: Friday 23 November

Location: Building A, Salón de Actos.

Hour: 11:00 - 12:00

Chairperson: Javier Ortega García.

Title: Google's speech internationalization project: From 1 to 300 languages and beyond

Abstract:

The speech team at google has built speech recognition systems in more than 30 languages in little more than 2 years. In this talk we will describe the history of this project and what technologies have been developed to achieve this goal. I'll explore a bit some of the acoustic modeling, lexicon, language modeling, infrastructure and even social engineering techniques used to achieve our ultimate goal, to build speech recognition systems in the top 300 languages of the planet as fast as possible.

Speaker Bio:

Dr. Pedro J. Moreno leads the speech internationalization engineering group at the Android division of Google. His team is in charge of the infrastructure, engineering, and research needed to deploy and maintain multilingual speech recognition services worldwide.

He joined google 8 years ago after working as a research scientist at HP Labs. During his work at HP he worked mostly in audio indexing systems.

Dr. Moreno completed his Ph.D. studies at Carnegie Mellon University under the direction of Prof. Richard Stern. His work there was focused on noise robustness in speech recognition systems.

His Ph.D. studies were sponsored by a Fulbright scholarship.

Before that he completed an Electrical Engineering degree at Universidad Politecnica de Madrid, Spain.

O6 - Oral Session: Applications of Speech and Language Technologies

Date: Friday, 23 November

Location: Building C, Sala de Grados.

Hour: 12:00 - 13:40

Chairperson: Asunción Moreno.

- O6-1: Integrating a state-of-the-art ASR system into the Opencast Matterhorn platform, *Juan Daniel Valor Miró, Alejandro Pérez González de Martos, Jorge Civera, and Alfons Juan.*
 - O6-2: Speech Reconstruction by Sparse Linear Prediction, *Jan Koloda, Antonio Miguel Peinado Herreros, and Victoria Eugenia Sánchez Calle.*
 - O6-3: Steganographic pulse-based recovery for robust ACELP transmission over erasure channels, *Domingo López, Angel Gomez, José Luis Pérez Córdoba, Bernd Geiser, and Peter Vary.*
 - O6-4: A proposal for a visual speech animation system for European Portuguese, *José Serra, Manuel Ribeiro, João Freitas, Verónica Orvalho, and Miguel Dias.*
 - O6-5: Online learning of log-linear weights in interactive machine translation, *Francisco Javier López-Salcedo, Germán Sanchis-Trilles, and Francisco Casacuberta Nolla.*
-



**VII Jornadas en Tecnologías del Habla
and III Iberian SLTech Workshop**

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

SOCIAL PROGRAM

Welcome Reception

Date: Wednesday 21 November

Location: Building C, Hall.

Hour: 19:30 -

We would like to invite all conference attendees and accompanying persons to meet us for the Welcome Reception that will take place at 19:30, right after finishing the technical program of the first day of the conference.

The Welcome Reception will take place at the hall of building C.

Optional Visit to Museo Reina Sofía

Date: Thursday 22 November

Location: Museo Reina Sofía.

Hour: 18:30 – 20:45

After the technical activities of the second day, we would like to invite you to join us for a guided visit to Museo Reina Sofía.

Transportation from the conference venue to Museo Reina Sofía will be provided by bus. Buses will leave from the conference venue to the museum at 18:30.

EsMadrid provides us with two buses with a total capacity of 100 seats. For this reason the number of tickets available for this visit is 100. If you want to come to the visit by your own means, please do not take a ticket for this visit and come by 19:00 to the following access of the museum (the conference organization will cover entrance and the guided visit even if you don't have a ticket):

Nouvel Building Access

Address: Ronda de Atocha (esquina plaza Emperador Carlos V), 28012 Madrid

The guided visit will be in small groups (max 25 people). After the visit, a short walk will take us to the restaurant where the Gala Dinner will take place. This makes this visit, apart from an interesting visit, also a convenient way to get to the Gala Dinner.

IMPORTANT NOTICE: This activity will be completely free (including entrance to the museum) for conference attendees and accompanying persons until we reach a total of 100 people. This limit is imposed by the capacity of the buses. The organization of the conference will cover both the guides at the museum and the entrance for the conference attendees not having free entrance even if they don't have a ticket for the bus and visit.

IMPORTANT NOTICE: Entrance to the museum is free for all types of teachers, assistant professors, associate professors and full professors (with a valid teacher or professor id card), students (with a valid university card) and young people (with a valid youth card). If you belong to any of these groups, please remember to bring your valid card for free entrance.

Gala Dinner

Date: Thursday 22 November

Location: Restaurant Samarkanda.

Hour: 21:00 –

Atocha Train Station, right over the tropical garden

The Gala Dinner will take place at the Samarkanda Restaurant. The Restaurant is located inside the Atocha Train Station, on the first floor, just over a tropical garden, which makes it a very special place, in particular for the cold autumns and winters of Madrid.

The gala dinner is included in the registration for conference attendees. A limited number of tickets for the Gala Dinner may be available for accompanying persons on a first-come first-served basis. The price of the dinner is 60 €. Payments will be made in cash at the registration desk.

Although we recommend coming to the visit to the museum, which also provides a convenient way to get to the venue for the Gala Dinner, our second recommendation is to get to the Gala Dinner by public transportation.

You can also get to the Gala Dinner by car. The restaurant has a free private parking, although it is a bit difficult to get to it. If you plan to get to the Gala Dinner by car, please ask our secretariat for directions.



VII Jornadas en Tecnologías del Habla and III Iberian SLTech Workshop

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN

ORGANIZATION

General Chairs

Doroteo Torre Toledano
Alfonso Ortega Giménez
António Teixeira

Universidad Autónoma de Madrid
Universidad de Zaragoza
Universidade de Aveiro

Program Chairs

Joaquín González Rodríguez
Rubén San Segundo Hernández
Luis Hernández Gómez

Universidad Autónoma de Madrid
Universidad Politécnica de Madrid
Universidad Politécnica de Madrid

Publication Chair

Daniel Ramos Castro

Universidad Autónoma de Madrid

Demos Chairs

Daniela Braga
Alberto Abad Gareta

Microsoft (MLDC)
INESC

Awards Chairs

Miguel Sales Dias
Climent Nadeu Camprubi

Microsoft (MLDC)
Universitat Politècnica de Catalunya

Evaluation Chairs

Javier Tejedor Noguerales
Javier González Domínguez

Universidad Autónoma de Madrid
Universidad Autónoma de Madrid



**VII Jornadas en Tecnologías del Habla
and III Iberian SLTech Workshop**

21-23 November 2012
Escuela Politécnica Superior
Universidad Autónoma de Madrid, SPAIN