

## Xavier Anguera Miró, Ph.D.

Ausias Marc 124, 5-1 – 08013 Barcelona  
+34 696954775  
[xanguera@gmail.com](mailto:xanguera@gmail.com)  
<http://www.xavieranguera.com>

---

### Motivation and goals

To help envision, research and develop the technologies of tomorrow within a multidisciplinary team of motivated people.  
To hold a research position where I am able to apply my technical skills while learning new ones.

### Research Interests

I am interested in the areas of multimedia information retrieval using multimodal information and speech/audio processing.

### Research impact metrics

(Updated August 5th, 2015)  
Google Scholar h-index: 22  
Google Scholar i10-index: 41

### Education

- 2003 – 2006:       **Ph.D. Program in Telecommunications**  
Speech and language processing group (TALP)  
Department of theory of signal and communications (TSC).  
Technical University of Catalonia (UPC).  
**Thesis title:** “Robust Speaker Diarization for Meetings”.  
Thesis defended on Dec. 2006 and awarded Excellent with Summa Cum Laude.
- 1999 – 2001 :       **European Masters in Language and Speech (EMLS).**  
Sponsored by ISCA and EACL organizations.  
<http://www.cstr.ed.ac.uk/emasters/>  
**Masters Thesis title:** “Concatenation based Spanish Text-to-Speech Synthesis for embedded applications”. Defended Sept. 2011, awarded Excellent with Summa Cum Laude.  
Technical University of Catalonia (UPC), University of Barcelona (UB) and University Pompeu Fabra (UPF).
- 1996 – 2001 :       **Telecommunications engineering (MS-EE)**  
Technical and superior School of Telecommunications engineering of Barcelona (ETSETB).  
Technical University of Catalonia (UPC).
-

## Selected publications

- “Speaker Diarization: A Review of Recent Research”, **Xavier Anguera**, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, O. Vinyals., to appear in IEEE Transactions on Audio, Speech, and Language Processing (TASLP), 2012
  - “Acoustic beamforming for speaker diarization of meetings”, **Xavier Anguera**, Chuck Wooters and Javier Hernando, IEEE Transactions on Audio, Speech and Language Processing, September 2007, volume 15, number 7, pp.2011-2023.
  - “MASK: Robust Local Features for Audio Fingerprinting”, **Xavier Anguera**, Antonio Garzón and Tomasz Adamek, in Proc. ICME 2012, Melbourne, Australia. **Best paper award**
  - “Information Retrieval-based Dynamic Time Warping“, **Xavier Anguera**, in Proc. Interspeech 2013, Lyon, France.
- 

## Professional experience

February 2015 - ongoing

**Founder and CEO** of Sinkronigo S.L., Barcelona, Spain

Following a long time dream, I created a technology startup in the area of speech technologies and digital publishing. We create epub3 eBooks with embedded natural audio, aligned at word level with the text. We also make bilingual eBooks where one of the languages contains audio narration. The goal of the company is to help people read better and learn new languages.

January 2008 – January 2015

**Research Scientist** at Telefónica I+D, Barcelona, Spain

One of the founding members of the multimedia scientific group. My research has revolved around many areas in multimedia and speech/audio analysis. Research topics:

- mobile multimedia: multimodal tagging on a camera phone (2008-2009)
- Audio indexing: Advertisement and soccer highlights detection (2008)
- multimodal video-copy detection (2009-2011)
- Binary-keys for fast audio representation: applied to speaker verification and diarization (2010-2011)
- Multimodal synchronization: music to music-videos, text to audio (2010-2011)
- Audio Fingerprinting for music and audio ID (2012-2013)
- Emotion recognition using binary-keys (2012-2013)
- Zero-resources speech processing: query-by-example spoken-term detection, lattice-based search (2012- 2015)
- Speaker characterization (2013- 2015)
- Call center data analytics (2013- 2015)
- ASR training on low-resourced languages (2014- 2015)
- Participation in BISON European project

January 2009 – September 2011

**Associate professor** at Universitat Pompeu Fabra (about 9K students and 1k lecturers)

- Courses taught: Operating systems (2009), C programming and concurrency(2009-2010), digital speech processing (2011) to students from Computer Science and Electrical Engineering majors.

January 2007 – December 2007:

**Research engineer** at Telefónica I+D, Barcelona, Spain

- Research on acoustic indexing within Spanish funded project I3Media (multimedia content analysis and generation in broadcasted media).

November 2006 – December 2006:

**Visiting researcher** at LIMSI-CNRS, Paris, France

- Helped on the improvement of the Spanish ASR system within the European TC-Star project.

September 2005 – September 2006:

**Visiting researcher**, recipient of the ICSI fellowship between the International Computer Science Institute and the Spanish Education and Science Ministry (MEC).

- Research on robust Speaker Segmentation and Clustering for Meetings.
- Participation on the RT06s NIST Meetings evaluation.

September 2004 – September 2005:

**Visiting researcher** within the AMI Training Programme (AMI is funded by the 6<sup>th</sup> European Framework) at ICSI.

- Research on Robust Speaker Segmentation and clustering for Meetings.
- Speaker diarization in Broadcast News. Helped in the participation to the RT04F EARS Evaluation on speaker clustering and rich transcription.
- Participation on the RT06s NIST Meetings evaluation.

September 2003 – September 2004:

**Research assistant** in the TALP group (TSC department), UPC Barcelona.

- Speaker verification system and dialog implementation for PETRA project (Unified messaging system) and integration with the rest of the system. Project funded by the Spanish Government. Project finished in December 2003, worked on it from October to December 2003.
- Dialog system in VXML for retrieval of news from an online newspaper. Developed using Verbio dialog server from ATLAS Company.
- Research in speaker segmentation and tracking.

- Involved in the CHIL project (Computers in the Human Interaction Loop). Development of the speaker tracking system and integration of the demo room's audio technologies. Project funded by the European Union.
- Involved in the BIOSEC project (Biometrics and Security) on speaker recognition systems. Project funded by the European Union.

September 2001 – August 2003:

**Research Software Engineer** at Panasonic Speech Technology Laboratory (PSTL), Santa Barbara, California.  
Speech R&D Department.

- Improvement of Spanish Text-to-Speech system for unlimited text.
- Built a second robot prototype to perform research in human-machine interaction with control over the Internet and PDA by voice. Focus on dialog systems oriented to tele-operation and how robots can help humans.
- Assistance in the creation of an American Spanish Speech recognition Database within SPEECON project (creation of speech resources in multiple conditions in more than 20 languages). Project with participation of 12 Speech Companies.
- Conversion of French, Italian and German isolated words Text-To-Speech synthesis systems to unlimited synthesis.
- Development of software tools to automate Multilingual Speech labeling oriented to concatenation based speech synthesis.

February 2001 – September 2001:

**Full time Internship** at Panasonic Speech Technology Laboratory (PSTL), Santa Barbara, California. Speech R&D department.

- Participated in the creation of Spanish Text-To-Speech system for isolated words. Presented as my final degree thesis, with maximum qualification obtained.
- Research in Speaker Recognition based on NIST evaluations.
- Built a Robot prototype for multimodal dialog applications (human machine interaction). Focused towards dialog capabilities and to become a showcase of the company's speech technologies.
- Definition and design of a Speech Recognition Database for American Spanish (SPEECON). Phonetic description and creation of prompts.

May 2000 – December 2000:

**Part time Internship** at Hewlett Packard, Barcelona Site, Spain.  
Large format printers R & D department.

- Research assistant in the writing system and optical sensors of large format printers.

---

## Skills

- Computers:

Engineering level programming in all operating systems. Knowledge of most corporate tools

- Languages:

Catalan: Fluent (Mother tongue).

Spanish: Fluent.

English: Fluent.

First certificate in English by Cambridge University.

Several years living in USA.

Portuguese: Fluent Listening and speaking, beginner writing.

French: Knowledge of.

Level 1 in DELF exam from the official French institute.

**Academic activities**

- PhD thesis coordination:
  - Hector Delgado: "Fast Cross-session Speaker Diarization" thesis on speaker diarization using binary keys. Expected defense end of 2015
  - Ciro Gracia: ongoing PhD on query-by-example spoken term detection. Expected to finish in 2015
- Masters thesis coordination:
  - Antonio Garzón Lorenzo: "Robust Acoustic fingerprinting", in requirement for the UPF Masters in music and sound computing, UPF, September 2011
- Bachelor thesis coordination:
  - Carlos Fernández: "Pattern-based system for helping programmers", in requirement for the Engineering bachelor degree, UPF, December 2011. Obtained Excellent Suma Cum Laude grade.
  - Esperança Movellan: "Reconocimiento de locutor e idioma mediante fingerprints binarios", in requirement for the Engineering bachelor degree, UPF, December 2011. Obtained Excellent Suma Cum Laude grade.
  - Guillem Pérez: "Language Learning with Audio Recognition", in requirement for the Engineering bachelor degree, UPF, September 2012. Obtained Excellent grade.
- PhD Thesis Committees:
  - Gautam Varma Mantena: IIIT Hyderabad, India, May 2015. Topic: Query-by-example spoken-term detection.
  - Binyam Gebrekidan Gebre: Radboud University, Nijmegen, The Netherlands. February 2015. Topic: Machine learning for gesture recognition in videos.
  - Sree Harsha Yella: IDIAP Institute/EPFL, Lausanne, Switzerland. January 2015. Topic: Overlap detection in speaker diarization.
  - Houssemeddine Khemiri: Telecom ParisTech, France, September 2013. Topic: Data-driven audio indexing
  - Simon Bozonnet: Eurecom Institute, Sophia Antipolis, France, May 2012, PhD by Telecom Paris Tech. Topic: Speaker Diarization.
  - Armando Muscariello: University of Rennes, France, January 2011. Topic: Automatic pattern discovery on speech.

- David Fonseca: Universidad La Salle, Barcelona, September 2010.  
Topic: Aesthetic measures for image displays.

---

## Benchmarks and challenges participation

Over the years I have participated and organized several technical challenges and benchmarks to position my technology with that of other research/corporate groups:

- 2004 - NIST Rich Transcription (RT) evaluation. Participated in the speaker Diarization subchallenge for broadcast news data
  - 2005 & 2006 – NIST Rich Transcription (RT) evaluation. Main lead in the speaker diarization subchallenge for Meeting data. Best submitted system for both years. In addition, in 2006 my acoustic beamforming system was also used in the speech recognition subchallenge with great results.
  - 2007 – NIST Rich Transcription (RT) evaluation. Collaborated in a team from UPC University to build a system from scratch.
  - 2007 – NIST Language Recognition evaluation. Participated with an acoustics-only system.
  - 2009 – NIST TRECVID Video-copy detection evaluation. Participated with a novel multimodal change-based system.
  - 2010 & 2011 - NIST TRECVID Video-copy detection evaluation. Participated with multimodal local features. Both years I was invited to present our system (only very few systems were invited per year).
  - 2011 – Mediaeval Query-by-Example Spoken Term Detection (SWS) on Indian languages. Obtained the best results.
  - 2012 – Albayzin Search-on-speech evaluation. Query-by-example search on Spanish data. Obtained the best results in the evaluation.
  - 2012 & 2013 – Mediaeval Query-by-Example Spoken Term Detection (SWS) on multiple languages (African languages in 2012 and 9 languages in 2013). Co-organized the evaluation and participated with my system.
  - 2014 – Mediaeval QUESST (Query by Example Search on Speech) task (~24h database, 6 languages). Coorganized the evaluation
  - 2015 – Mediaeval QUESST (Query by Example Search on Speech) task. Coorganized the evaluation
  - 2015 - Coorganizer of the Zero Resources Speech Challenge, accepted as a special session at Interspeech 2015
- 

## Conference publications

### 2015

- "Query by Example Search on Speech at Mediaeval 2015", Igor Szoke, Luis-Javier Rodriguez-Fuentes, Andi Buzo, Xavier Anguera, Florian Metze, Jorge Proença, Matus Pleva and Xiao Xiong, to appear in proc. Mediaeval 2015 Benchmark evaluation, Wurzen, Germany

- "The Zero Resource Speech Challenge 2015", Maarten Versteegh, Roland Thiolliere, Thomas Schatz, Xuan Nga Cao, **Xavier Anguera**, Aren Jansen and Emmanuel Dupoux, to appear Interspeech 2015, Dresden, Germany
- "Multimodal Read-aloud eBooks for Language Learning", **Xavier Anguera**, Show and Tell session. To appear Interspeech 2015, Dresden, Germany
- "Novel Clustering Selection Criterion for Fast Binary Key Speaker Diarization", Héctor Delgado, **Xavier Anguera**, Corinne Fredouille and Javier Serrano, to appear in proc. Interspeech 2015, Dresde, Germany
- "Effect of gender and call duration on customer satisfaction in call center big data", Quim Llimona, Jordi Luque, **Xavier Anguera**, Zoraida Hidalgo, Souneil Park, Nuria Oliver, to appear in Proc. Interspeech 2015, Dresden, Germany.
- "Improved Binary Key Speaker Diarization System", Héctor Delgado, **Xavier Anguera**, Corinne Freouille and Javier Serrano, to appear in proc. EUSIPCO 2015, Nice, France
- "MASK+:Data-driven Regions Selection for Acoustic Fingerprinting", Lucas Ondel, **Xavier Anguera** and Jordi Luque, in proc. ICASSP 2015, Brisbane, Australia.
- "QUESST2014: Evaluating Query-by-example Speech Search in a Zero-Resource Setting with Real-life Queries", **Xavier Anguera**, Luis-J. Rodriguez-Fuentes, Andi Buzo, Florian Metze, Igor Szöke and Mikel Peñagarikano, in proc. ICASSP 2015, Brisbane, Australia.
- "An Information-Theoretic Metric of Fingerprint Effectiveness", TJ Tsai, Gerald Friedland and **Xavier Anguera**, in proc. ICASSP 2015, Brisbane, Australia.

#### **2014**

- "Query by Example Search on Speech at Mediaeval 2014", **Xavier Anguera**, Luis-Javier Rodriguez-Fuentes, Igor Szöke, Andi Buso and Florian Metze, in proc. Mediaeval 2014
- "Global Speaker Clustering towards Optimal Stopping Criterion in Binary Key Speaker Diarization", Hector Delgado, **Xavier Anguera**, Corinne Fredouille and Javier Serrano, book chapter in Advances in Speech and Language Technologies for Iberian Languages. Lecture Notes in Computer Science, Volume 8854, 2014, pp. 59-68. Presented at Iberspeech 2014, Las Palmas, Spain
- "Phoneme-Lattice to Phoneme-Sequence matching algorithm based on Dynamic Programming", Ciro Gracia, **Xavier Anguera**, Jordi Luque and Ittai Artzi, book chapter in Advances in Speech and Language Technologies for Iberian Languages. Lecture Notes in Computer Science, Volume 8854, 2014, pp. 99-108. Presented at Iberspeech 2014, Las Palmas, Spain
- "Flexible Stand-alone Keyword Recognition Application using Dynamic Time Warping", Miquel Ferrarons, **Xavier Anguera** and Jordi Luque, book chapter in Advances in Speech and Language Technologies for Iberian Languages. Lecture Notes in Computer Science, Volume 8854, 2014, pp. 158-167. Presented at Iberspeech 2014, Las Palmas, Spain
- "Audio-to-text Alignment for speech recognition with very limited resources", **Xavier Anguera**, Jordi Luque and Ciro Gracia, in proc. Interspeech 2014, Singapore
- "Query-by-Example Spoken Term Detection on Multilingual Unconstrained Speech", **Xavier Anguera**, Luis Javier Rodriguez-Fuentes, Igor Szoke, Andi Buzo, Florian Metze and Mikel Penagarikano, in proc. Interspeech 2014, Singapore

- “On the Modeling of Natural Vocal Emotion Expressions Through Binary Key”, Jordi Luque and **Xavier Anguera**, in proc. Eusipco 2014, Lisboa, Portugal
- “Combining Temporal and Spectral Information for Query-By-Example Spoken Term Detection”, Ciro Gracia, **Xavier Anguera** and Xavier Binefa, in proc. Eusipco 2014, Lisboa, Portugal
- “Query-by-Example Spoken Term Detection Evaluation on Low-Resource Languages”, **Xavier Anguera**, Luis J. Rodriguez-Fuentes, Igor Szöke, Andi Buzo, Florian Metze, Mikel Penagarikano, in Proc. SLTU 2014, Saint Petersburg, Russia
- “Sentiment Retrieval on Web Reviews Using Spontaneous Natural Speech”, Jose Costa Pereira, Jordi Luque and **Xavier Anguera**, in Proc. ICASSP 2014, Florence, Italy
- “Inferring Social Relationships in a Phone Call from a Single Party’s Speech”, Sree Harsha Yella, **Xavier Anguera** and Jordi Luque, in Proc. ICASSP 2014, Florence, Italy

### 2013

- “Speed Improvements to Information Retrieval-Based Dynamic Time Warping Using Hierarchical K-means Clustering “, Gautam Mantena and **Xavier Anguera**, in Proc. ICASSP 2013, Vancouver, Canada.
- “Perceptually Inspired Features for Speaker Likability Classification “, Sira Gonzalez and **Xavier Anguera**, in Proc. ICASSP 2013, Vancouver, Canada.
- “The Spoken Web Search Task at Mediaeval 2012 “, Florian Metze, **Xavier Anguera**, Etienne Barnard, Marelle Davel and Guillaume Gravier, in Proc. ICASSP 2013, Vancouver, Canada. pdf
- “Memory Efficient Subsequence DTW for Query-by-Example Spoken Term Detection “, **Xavier Anguera** and Miquel Ferrarons, in Proc. ICME 2013, San Jose, CA, USA.
- “Information Retrieval-based Dynamic Time Warping “, **Xavier Anguera**, in Proc. Interspeech 2013, Lyon, France.
- “Activating the Crowd: Exploiting User-Item Reciprocity for Recommendation”, Martha Larson, Alan Said, Yue Shi, Paolo Cremonesi, Domonkos Tikk, Alexandros Karatzoglou, Linas Baltrunas, Jozef Geurts, **Xavier Anguera** and Frank Hopfgartner, in Proc. CrowdRec: Crowdsourcing and Human Computation for Recommender Systems Workshop, ACM RecSys 2013
- “Two-level clustering towards unsupervised discovery of acoustic classes”, Ciro Gracia, **Xavier Anguera** and Xavier Binefa, in Proc. ICMLA 2013, Florida, USA
- “A Riemannian stopping criterion for unsupervised phonetic segmentation”, Ciro Gracia, **Xavier Anguera** and Xavier Binefa, in Proc. ICMLA 2013, Florida, USA
- “The Spoken Web Search Task”, **Xavier Anguera**, Florian Metze, Andi Buso, Igor Szöke and Luis-Javier Rodriguez Fuentes, in Proc. Mediaeval 2013 evaluation Workshop, Barcelona, Spain.
- “The Telefonica Research Spoken Web Search System for Mediaeval 2013”, **Xavier Anguera**, Miroslav Skázel, Volker Vorwerk and Jordi Luque, in Proc. Mediaeval 2013 evaluation Workshop, Barcelona, Spain.

### 2012



- “Multimodal Video Copy Detection Using Local Features”, **Xavier Anguera** and Tomasz Adamek, in IEEE ComSoc MMTC E-letter Vol.7, No.5, September 2012
- “Emotions recognition using binary fingerprints”, **Xavier Anguera**, Esperança Movellan and Miquel Ferrarons, in Proc. Iberspeech 2012, Madrid, Spain.
- “Telefonica Research system for the Query-by-example task at Albayzin 2012”, **Xavier Anguera**, in Proc. Iberspeech 2012, Madrid, Spain
- “Telefonica Research System for the Spoken Web Search task at Mediaeval 2012”, **Xavier Anguera**, in Proc. Mediaeval 2012 evaluation Workshop, Pisa, Italy.
- “The spoken web search task”, Florian Metze, Etienne Barnard, Marelle Davel, Charl van Heerden, **Xavier Anguera**, Guillaume Gravier and Nitendra Rajput, in Proc. Mediaeval 2012 evaluation Workshop, Pisa, Italy
- “The Spoken Web search task at Mediaeval 2011”, Florian Metze, Nitendra Rajput, **Xavier Anguera**, Marelle Davel, Guillaume Gravier, Charl van Heerden, Gautam V. Mantena, Armando Muscariello, Kishore Prahallad, Igor Szoke, and Javier Tejedor, in Proc. ICASSP 2012, Kyoto, Japan.
- “Speaker Independent Discriminant Feature Extraction for Acoustic Pattern-Matching”, **Xavier Anguera**, in Proc. ICASSP 2012, Kyoto, Japan.
- “MASK: Robust Local Features for Audio Fingerprinting”, **Xavier Anguera**, Antonio Garzón and Tomasz Adamek, in Proc. ICME 2012, Melbourne, Australia.

#### **Best paper award**

#### **2011**

- "Telefonica System for the Spoken Web Search Task at Mediaeval 2011 ", **Xavier Anguera**, in Proc. MediaEval Workshop, November 2011, Pisa, Italy
- "Telefonica Research at TRECVID 2011 Content-Based Copy Detection ", **Xavier Anguera**, Tomasz Adamek, Daru Xu and Juan Manuel Barrios, in Proc. NIST-TRECVID workshop 2011
- "Combining Features at Search Time: PRISMA at Video Copy Detection Task ", Juan Manuel Barrios, Benjamin Bustos and **Xavier Anguera**, in Proc. NIST-TRECVID workshop 2011
- "Multimodal fusion for video copy detection ", **Xavier Anguera**, Juan Manuel Barrios, Tomasz Adamek and Nuria Oliver, in Proc. ACM Multimedia 2011.
- “Spoken Wordcloud: Clustering Recurrent Patterns in Speech”, Remi Flamary, **Xavier Anguera** and Nuria Oliver, to appear in CBMI 2011, Madrid, Spain.
- “Speaker Modeling using local Binary Decisions”, Jean-François Bonastre, **Xavier Anguera**, Gabriel Sierra, Pierre-Michel Bousquet, in Proc. Interspeech 2011, Florence, Italy.
- “Real-Time synchronization of multimedia streams in a mobile device”, Robert Macrae, Joachim Neumann, **Xavier Anguera**, Nuria Oliver and Simon Dixon, in Proc. ADMIRE Workshop within ICME 2011, Barcelona, Spain.
- “Automatic Synchronization of Electronic and Audio Books via TTS Alignment and Silence Filtering”, **Xavier Anguera**, Néstor Pérez, Andreu Urruela and Nuria Oliver, in Proc. Hot Topics in Multimedia within ICME 2011, Barcelona, Spain.
- “Fast Speaker Diarization Based on Binary Keys”, **Xavier Anguera** and Jean-François Bonastre, in Proc. ICASSP 2011, Prague, Check Republic
- “Discriminant Binary Data Representation for Speaker Recognition”, Jean-François Bonastre, **Xavier Anguera**, Pierre-Michel Bousquet, Driss Matrouf, in Proc. ICASSP 2011, Prague, Check Republic.

- “Closed-Form Expressions vs. BIC: a Comparison for Speaker Clustering”, Themis Stafylakis, **Xavier Anguera**, Vassilis Katsouros, George Carayannis, in Proc. ICASSP 2011, Prague, Check Republic.

## 2010

- “Telefonica Research at TRECVID 2010 Content-Based Copy Detection”, Ehsan Younessian, **Xavier Anguera**, Tomasz Adamek, Nuria Oliver and David Marimon, NIST Trecvid Workshop notebook paper
- “Novel binary key representation for biometric speaker recognition”, **Xavier Anguera** and Jean-François Bonastre, in Proc. Interspeech 2010, Makuhari, Japan.
- “System output combination for improved speaker diarization”, Simon Bozonet, Nicholas Evans, **Xavier Anguera**, Oriol Vinyals, Gerald Friedland and Corinne Fredouille, in Proc. Interspeech 2010, Makuhari, Japan.
- “Improvements to the equal-parameter BIC for Speaker Diarization”, Themis Stafylakis, **Xavier Anguera**, in Proc. Interspeech 2010, Makuhari, Japan.
- “MuViSync: Realtime Music Video Alignment”, Robert Macrae, Xavier Anguera and Nuria Oliver, in Proc. ICME 2010
- “Enriching Music Mood Annotation by Semantic Association Reasoning”, Jun Wang, **Xavier Anguera**, Xiaou Chen and Deshun Yang, in Proc. AdMiRe Workshop, in ICME 2010
- “Partial Sequence Matching Using an Unbounded Dynamic Time Warping Algorithm”, **Xavier Anguera**, Robert Macrae and Nuria Oliver, in Proc. ICASSP 2010
- “Unrestricted Voice Annotations and Search of Personal Photographs in a Mobile Phone”, **Xavier Anguera**, Mauro Cherubini and Nuria Oliver, in Proc. Of Spoken Query 2010 Workshop on voice search at ICASSP 2010

## 2009

- “Telefónica Research Content-Based Copy Detection TRECVID Submission”, **Xavier Anguera**, Pere Obrador, Tomasz Adamek, David Marimon and Nuria Oliver, NIST Trecvid Workshop notebook paper
- “Revisiting the use of GMM-SVM for speaker verification”, **Xavier Anguera**, in Proc. Interspeech 2009
- “Multimodal video copy detection of social media”, **Xavier Anguera**, Pere Obrador and Nuria Oliver, in Proc. first SIGMM Workshop on Social Media (WSM2009) at ACM MM09
- “The role of tags and image aesthetics in social image search”, Pere Obrador, **Xavier Anguera**, Rodrigo de Oliveira and Nuria Oliver, in Proc. first SIGMM Workshop on Social Media (WSM2009) at ACM MM09
- “Text versus Speech: A Comparison of Tagging Input Modalities for Camera Phones”, M. Cherubini, **Xavier Anguera**, N. Oliver and R. de Oliveira, Proc. MobileHCI, Bonn, Germany, September 2009, (best paper award nominee)
- “Audio-Based Soccer Game Summarization”, Helenca Duxans, **Xavier Anguera** and David Conejero, in Proc. IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB09)
- “Audio-Based Automatic Management of Audio Commercials”, H. Duxans, D. Conejero and **Xavier Anguera**, in Proc. ICASSP 2009, Taipei, Taiwan. April 2009

## 2008

- "TV advertisements detection and clustering based on acoustic information", D. Conejero and **Xavier Anguera**, in proc. International Conference on Computational Intelligence for Modelling, Control and Automation - CIMCA08, Viena, Austria, December 2008
- "MAMI: Multimodal Annotations on a Camera Phone", **Xavier Anguera** and N. Oliver, in Proc. MobileHCI, Amsterdam, September 2008
- "Sistema de Indexación Automática de Contenidos Multimedia", U. Urdapilleta, D. Conejero, **Xavier Anguera**, D. Cacenas and F.J. Caminero, in Proc. XVIII Jornadas Telecom I+D, Bilbao, Spain
- "Multimodal and Mobile Personal Image Retrieval: A User Study", **Xavier Anguera**, N. Oliver and M. Cherubini, in Proc. Workshop on Mobile Information Retrieval, MOBIR'08, Singapore
- "Multimodal Photo Annotation and Retrieval on a Mobile Phone", **Xavier Anguera**, J. Xu, N. Oliver, in Proc. ACM Intl. Conference on Multimedia Information Retrieval, Vancouver, Canada.

### 2007

- "Model Complexity Selection and Cross-Validation EM Training for Robust Speaker Diarization", **Xavier Anguera**, Takahiro Shinozaki, Chuck Wooters and Javier Hernando, ICASSP, Hawaii, USA, April 2007.
- "Automatic Weighting for the Combination of TDOA and Acoustic Features in Speaker Diarization for Meetings", **Xavier Anguera**, Chuck Wooters, J.M Pardo and Javier Hernando, ICASSP, Hawaii, USA, April 2007.
- "Speaker Diarization for Conference Room: The UPC RT07s Evaluation System", Jordi Luque, **Xavier Anguera**, Andrey Temko, and Javier Hernando, RT07s Rich Transcription evaluation workshop, Washington, May 2007
- "The SRI-ICSI Spring 2007 Meeting and Lecture Recognition System", Andreas Stolcke, **Xavier Anguera**, Kofi Boakye, Ozgur Çetin, Adam Janin, Mathew Magimai-Doss, Chuck Wooters, and Jing Zheng, RT07s Rich Transcription evaluation workshop, Washington, May 2007

### 2006

- "Purity Algorithms for Speaker Diarization of Meetings Data", **Xavier Anguera**, Chuck Wooters and Javier Hernando. ICASSP 2006, Toulouse, France, May 2006.
- "Speaker Diarization for Multi-Microphone Meetings Using only Between-Channel Differences", Jose M. Pardo, **Xavier Anguera**, Chuck Wooters, In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2006), Lecture Notes in Computer Science. Springer
- "Automatic Cluster Complexity and Quantity Selection: Towards Robust Speaker Diarization", **Xavier Anguera**, Chuck Wooters, Javier Hernando, In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2006), Lecture Notes in Computer Science. Springer
- "Robust Speaker Diarization for Meetings: ICSI RT06s Meetings Evaluation System", **Xavier Anguera**, Chuck Wooters and Jose M. Pardo, In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2006), Lecture Notes in Computer Science. Springer
- "Purity Algorithms for Speaker Diarization of Meetings Data", **Xavier Anguera**, Chuck Wooters, Javier Hernando, MMUA 2006, Toulouse, France, May 2006.

- "Hybrid Speech/Non-Speech Detector Applied to Speaker Diarization of Meetings", **Xavier Anguera**, Mateu Aguilo, Chuck Wooters, Climent Nadeu and Javier Hernando, Speaker Odyssey 2006, San Juan de Puerto Rico, USA, June 2006.
- "Friends and Enemies: A Novel Initialization for Speaker Diarization", **Xavier Anguera**, Chuck Wooters and Javier Hernando, ICSLP'06, Pittsburgh, Pensilvania, USA, September 2006.
- "Robust Speaker Diarization for Meetings: ICSI RT06s evaluation system", **Xavier Anguera**, Chuck Wooters and Jose M. Pardo, ICSLP'06, Pittsburgh, Pensilvania, USA, September 2006.
- "The ICSI-SRI Spring 2006 Meeting Recognition System", A. Janin, A. Stolcke, **Xavier Anguera**, K. Boakye, O. Cetin, J. Frankel, and J. Zheng, In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2006), Lecture Notes in Computer Science. Springer
- "Multi-Stream Speaker Diarization Systems for the Meetings Domain", Ascension Gallardo, **Xavier Anguera** and Chuck Wooters, ICSLP'06, Pittsburgh, Pensilvania, USA, September 2006.
- "Speaker Diarization for Multiple Distant Microphone Meetings: Mixing Acoustic Features And Inter-Channel Time Differences", Jose M. Pardo, **Xavier Anguera** and Chuck Wooters, ICSLP'06, Pittsburgh, Pensilvania, USA, September 2006.

#### 2005

- "XBIC: Real-Time Cross Probabilities Measure for Speaker Segmentation", **Xavier Anguera**. International Computer Science Institute Technical Report TR-05-008.
- "Robust Speaker Segmentation for Meetings: The ICSI-SRI Spring 2005 Diarization System", **Xavier Anguera**, Chuck Wooters, Barbara Peskin and Mateu Aguilo. In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2005), Lecture Notes in Computer Science. Springer
- "Further Progress in Meeting Recognition: The ICSI-SRI Spring 2005 Speech-to-Text Evaluation System", Andreas Stolcke, **Xavier Anguera**, Kofy Boakye, Ozgur Cetin, Frantisek Grezl, Adam Janin, Arindam Mandal, Barbara Peskin, Chuck Wooters and Jing Zheng. In S. Renals and S. Bengio, editors, Machine Learning for Multimodal Interaction: Third International Workshop (MLMI 2005), Lecture Notes in Computer Science. Springer
- "Speaker Diarization for Multi-Party Meetings Using Acoustic Fusion", **Xavier Anguera**, Chuck Wooters and Javier Hernando. Automatic Speech Recognition and Understanding (ASRU). Puerto Rico, November 2005.
- "PETRA: Advanced Oral Interfaces for Unified Messaging Applications", David Hernando, Javier Hernando and **Xavier Anguera**. Buran magazine, IEEE Barcelona student branch. Number 22, September 2005.

#### 2004

- "Evolutive Speaker Segmentation using a Repository System", **Xavier Anguera** and Javier Hernando. ICSLP, Korea 2004.
- "Segmentació de locutor per a la indexació automàtica de bases de dades multimèdia en català", **Xavier Anguera**, Mireia Farrús, Javier Hernando and Alberto Abad. II Congrés d'enginyeria en llengua catalana, Andorra 2004.
- "Els sistemes de reconeixement de veu i traducció automàtica en català: present i futur", Mireia Farrús, Jan Anguita, **Xavier Anguera**, Josep M. Crego, Adrià de Gispert, Javier Hernando, Climent Nadeu. II Congrés d'enginyeria en llengua catalana, Andorra 2004.

- “XBIC: Nueva Medida para Segmentación de Locutor hacia el Indexado Automático de la Señal de Voz”, **Xavier Anguera**, Javier Hernando and Jan Anguita. III Jornadas en Tecnología del Habla, Valencia, 17-10 Nov 2004.
  - “Towards Robust Speaker Segmentation: The ICSI-SRI Fall 2004 Diarization System”, Chuck Wooters, James Fung, Barbara Peskin and **Xavier Anguera**. EARS Program RT-04 Workshop, nov 7-10 2004.
- 

### Journal articles

- “Language independent search in MediaEval’s Spoken Web Search task”, Florian Metze, **Xavier Anguera**, Etienne Barnard, Marelie Davel and Guillaume Gravier, Elsevier Journal on Computer, Speech and language, January 2014
  - “Query-by-Example Spoken Term Detection ALBAYZIN 2012 evaluation: overview, systems, results, and discussion”, J. Tejedor, D.T. Toledano, **Xavier Anguera**, A. Varona, L.F. Hurtado, A. Miguel and J. Colás, EURASIP Journal on Audio, Speech, and Music Processing 2013, 2013:23, September 2013
  - “Speaker Diarization: A Review of Recent Research”, **Xavier Anguera**, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, O. Vinyals, in IEEE Transactions on Audio, Speech, and Language Processing (TASLP), special issue on New Frontiers in Rich Transcription, VOL. 20, NO. 2, FEBRUARY 2012
  - "The ICSI RT-09 Speaker Diarization System", Gerald Friedland, Adam Janin, David Imseng, **Xavier Anguera**, Luke Gottlieb, Marijn Huijbregts, Mary Tai Knox and Oriol Vinyals, in IEEE Transactions on Audio, Speech and Language Processing (TASLP), special issue on New Frontiers in Rich Transcription, VOL. 20, NO. 2, FEBRUARY 2012
  - “Speaker Diarization For Multiple-Distant-Microphone Meetings Using Several Sources of Information”, Jose M. Pardo, **Xavier Anguera** and Chuck Wooters, IEEE Transactions on Computers, September 2007, volume 56, number 9, pp. 1189-1224.
  - “Acoustic beamforming for speaker diarization of meetings”, **Xavier Anguera**, Chuck Wooters and Javier Hernando, IEEE Transactions on Audio, Speech and Language Processing, September 2007, volume 15, number 7, pp.2011-2023.
- 

### Other publications

- “Future Media Internet: Reseach challenges and the road ahead”, multiple authors, published by the European Commission, April 2010, ISBN: 978-92-79-15864-3
  - “Concatenation based Spanish Text-to-Speech Synthesis for embedded applications”, Xavier Anguera, masters thesis for Telecommunications engineering, UPC, Barcelona, September 2001. Received maximum grade.
  - “Robust Speaker Diarization for Meetings”, Xavier Anguera, PhD Thesis, submitted in partial fulfillment for the PhD in the TSC department, UPC, Barcelona. Received Excellent qualification with Summa Cum Laude mention.
-

**Conference and invited talks**

- "Speaker Segmentation and Characterization", 2h lecture in the summer school of the RTTH (<http://rtthss2015.talp.cat/>), Barcelona, Spain, July 2015
- "Read-aloud Speech Technology by Sinkronigo", invited speaker at 4th LT-Innovate Summit, Brussels, Belgium, June 25-26, 2015
- "Big data speech applications", invited seminar talk at Technical University of Valencia, Spain, May 14th, 2015
- "Language Technologies for Language Learning", invited speaker at LT-solutions international Workshop to present Sinkronigo activities, San Sebastian, Spain, May 21st, 2015
- "Query-by-Example Spoken Term Detection on Multilingual Unconstrained Speech", paper talk at Interspeech 2014, Singapore, September 2014
- "Query-by-Example Spoken Term Detection Evaluation on Low-Resource Languages", paper talk at SLTU 2014, Saint Petersburg, Russia
- "Multimedia analysis for the poor", invited Stimulus talk at Dagstuhl seminar 13451 (computational audio analysis), Dagstuhl November 2013
- "The spoken web search task", Mediaeval Task overview talk at Mediaeval Workshop, Barcelona, Spain, October 2013.
- "Selected topics in Multimedia analysis", invited talk at Northwestern University, Xi'an, China, September 2013
- "Information Retrieval Dynamic Time Warping", paper talk at Interspeech 2013, Lyon, France, August 2013.
- "Zero-Resources Spoken Audio Search", invited talk at Microsoft Research, Menlo Park, Bay Area, California, July 2013.
- "Selected topics on Multimedia analysis", invited talk at Gracenote, Emeryville, California, July 2013.
- "Dynamic Time Warping's New Youth", Plenary Time Machine Session at ICME 2012, Melbourne, July, 2012
- "MASK: Robust Local Features for Audio Fingerprinting", paper presentation at ICME 2012, Melbourne, Australia. **Best paper award**
- "Speaker and content identification", invited talk at Inesc, Lisboa, April 2012
- "Multimodal video copy detection", invited talk at Eurecom Institute, Sofia Antipolis, France, April 2012
- "Zero Resources audio information retrieval", invited talk at Valencia University, Departamento de Sistemas Informáticos y Computación, December 2011.
- "Identification of similar videos and speakers", invited talk at Columbia University, Electrical Engineering department, December 2011
- "Fast speaker diarization based on binary keys", paper presentation, ICASSP 2011, May 2011.
- "Multimodal speech algorithms and applications", invited talk, research seminars, Universitat Nova de Lisboa, CS department, Lisboa, Portugal, May 2011.
- "Acoustic fingerprinting at Into\_now", invited talk, startups day, event organized by Telefonica Research, May 2011.
- "Telefonica Research at TRECVID 2011 Content-Based Copy Detection", system presentation, TRECVID 2011 Workshop, November 2011.
- "Spoken Wordcloud: Clustering Recurrent Patterns in Speech", paper talk at CBMI 2011, Madrid, Spain.

- “Automatic Synchronization of Electronic and Audio Books via TTS Alignment and Silence Filtering”, paper talk at Hot Topics in Multimedia within ICME 2011, Barcelona, Spain.
  - “Fast Speaker Diarization Based on Binary Keys”, paper presentation at ICASSP 2011, Prague, Check Republic.
  - “Multimodal pattern matching algorithms and applications”, invited talk, Laboratoire d’informatique d’Avignon, France, June 2010
  - “Multimodal pattern matching algorithms and applications”, invited talk, Priberam machine learning lunch seminar, Instituto Superior Tecnico, Technical University of Lisboa, Portugal, May 2010.
  - “Recent research on audio and multimodal processing”, invited talk, Stanford Research Institute (SRI), Stanford, California, USA, November 2010.
  - “Telefonica Research multimedia video copy detection”, System presentation, TRECVID 2010 Workshop, November 2010.
  - “Partial sequence matching using an unbounded dynamic time warping algorithm”, paper presentation, ICASSP 2010, March 2010.
  - “Unrestricted Voice Annotations and Search of Personal Photographs in a Mobile Phone”, paper presentation Spoken Query 2010 Workshop on voice search at ICASSP 2010
  - "MAMI: Multimodal Annotations on a Camera Phone", paper presentation MobileHCI, Amsterdam, September 2008
  - "Multimodal and Mobile Personal Image Retrieval: A User Study", paper presentation Workshop on Mobile Information Retrieval, MOBIR'08, Singapore
- 

### Grants and prizes

- Spanish fellowship “Programa Torres Quevedo” with a 3 years grant to work on multimedia Indexing at Telefonica Research. Granted in 2008.
  - ICSI fellowship for Spanish Scientists promoted by the Spanish Education and Science Ministry (MEC). This award is given once a year to 5 proficient Spanish scientists to pursue research stays at ICSI. Awarded for the period of September 2005 to September 2006.
  - AMI Training program for speech scientists. AMI project, March 2004. Awarded to European students in the speech area to perform research at ICSI from September 2004 to September 2005.
  - “Programa BE de mobilitat de personal investigador”. Generalitat de Catalunya, March 2004. Awarded to pursue research in foreign countries.
- 

### Patents

1. “Multilingual Text-to-Speech system with limited resources”, Granted in 2004
2. “Interactive, Personalized Robot for Home use”, Granted in 2003
3. “Intelligent Nurse Robot”. Patent application 2004
4. “Method and System of Real-time identification of an Audiovisual Advertisement in a Data Stream”, Granted 2012

5. “Method and System of Classification of Audiovisual Information”, Patent application 2008
  6. “Method and System of Classification of Audiovisual Information”, Patent application 2009
  7. “Method for Detecting Audio and Video Copy in Multimedia Streams”, Patent application 2009
  8. “Procede de Classification de donees biometriques”, patent Application 2010
  9. “Real Time Music to Music Video Synchronization Method and System”, Patent application 2010
  10. “Fast Partial Pattern Matching and Method”, Granted 2014
  11. “A Method to Generate Audio Fingerprints”, Patent application 2011
  12. “System and Method to Analyze Voice Communications”, Patent application 2012
  13. “Self Awareness Assistant”, Patent application 2012
- 

### **Technical Program Committees and reviewing assignments**

- Member in technical program committees:
  - Interspeech: 2008 - current
  - ICASSP: 2008 - current
  - ACM MM 2008 -2013
  - ACM-MIR
  - IEEE-ICME
  - IEEE ASRU
  - ICMR
  - SocialCom 2012
  - WOSSPA 2013 (8<sup>th</sup> International workshop on systems, signal processing and their applications)
- Journal reviewer
  - IEEE Transactions on Audio Speech and Language Processing (TASLP)
  - IEEE Transactions on Multimedia
  - ACM TOMCCAP
  - Journal of Computer Mathematics
  - Multimedia Tools and Applications (Springer)
  - IET Image processing
  - EURASIP Journal on Audio, Speech, and Music Processing
  - ACM Trans. on Information Systems.
  - IEEE Transactions on Human-Machine System

### **Scientific leadership**

- Conferences:
  - Publicity chair, ACM-MIR 2010
  - Local chair, ACM-RECSYS 2010
  - Local chair, ICMI 2011
  - Student program co-chair, IEEE-ICME 2011
  - Proceedings co-chair, ACM MM 2011
  - Area chair, IEEE-ICME 2012
  - Area chair, IEEE-ICME 2013



- Sponsorship team member, Interspeech 2013
- Technical demos co-chair, ACM MM 2013
- Local chair, Mediaeval 2013
- Area chair, IEEE-ICME 2014
- Area chair, ACM MM 2014
- Area chair, ICPR 2014
- Journals:
  - Associate editor @ Journal of Multimedia (<http://www.academypublisher.com/jmm/>) since 2013
  - Associate editor @ EURASIP Journal on Audio, Speech, and Music Processing (<http://asmp.eurasipjournals.com/>) since 2013

### **Scientific events organization**

- ICME 2011 – Workshop on Multimodal Audio-based Multimedia Content Analysis (MAMCA). July 2011, Barcelona
- Telefonica Research Multimedia Analysis Workshop (in conjunction with ICME 2011) July 2011, Barcelona

### **Other scientific activities**

- IEEE Computer Society's Technical Committee on Multimedia Computing (TCMC)
- Co-chair of SIGASP (Special Interest Group on Audio and Speech Processing for Multimedia) (<http://www.computer.org/portal/web/tcmc/SIGASP>)

---

### **Other activities**

- Toastmasters International (Public speaking and leadership organization). DTM (Distinguished Toastmasters) certification. Served in several leadership roles.
- Robotics enthusiast. Participated in multiple contests during university and built two robots for dialog interaction.
- Occasional ballroom dancer.

---

### **References**

Available upon request.