

Giampiero Salvi

Curriculum Vitae

Lisbon, August 10, 2007

Born Rome, Italy, May 1, 1973

Contact UTL (Universidade Técnica de Lisboa)
IST (Instituto Superior Técnico)
ISR (Institute for Systems and Robotics)
Av. Rovisco Pais, 1
1049-001 Lisboa
Portugal

Phone +351 21841 8197
Fax.: +351 21841 8291
giampi@kth.se
<http://www.speech.kth.se/~giampi/>

Research Interests

Areas: machine learning, variable-length sequence classification, unsupervised learning, automatic speech recognition, speech signal processing.

Methods: recurrent neural networks, Bayesian learning and inference, Gaussian mixture modelling, model-based clustering, hierarchical clustering, hidden Markov models.

Education

Ph.D., Computer Science, October 2006
KTH (Kungl Tekniska Högskolan),
CSC (School of Computer Science and Communication),
TMH (Department of Speech, Music, and Hearing),
Stockholm, Sweden

Thesis: *Mining Speech Sounds, Machine Learning Methods for Automatic Speech Recognition and Analysis*
Adviser: Prof. Björn Granström,¹
Opponent: Prof. Torbjørn Svendsen,²
Committee: Prof. Eva Strangert,³ Prof. Anders Lansner,⁴ Ass. Prof. Thomas Eriksson.⁵

M.Sc. (Laurea), Electronics, November 1998
La Sapienza University
Rome, Italy

Thesis: *Developing Acoustic Models for Speech Recognition*
Adviser: Prof. Maria-Gabriella Di Benedetto,⁶ Prof. Rolf Carlson.⁷

¹KTH, CSC, TMH, Stockholm, Sweden, <http://www.speech.kth.se/~bjorn/>

²Norwegian University of Science and Technology, Trondheim, Norway, <http://www.iet.ntnu.no/~torbjorn/>

³Umeå University, Sweden, <http://www.ling.umu.se/staff/person.html?eva.strangert>

⁴KTH, CSC, Computational Biology, Stockholm, Sweden, <http://www.nada.kth.se/~ala/>

⁵Chalmers, Göteborg, Sweden, <http://www.s2.chalmers.se/~thomase/>

⁶La Sapienza, Rome, Italy, <http://wsfalco.ing.uniroma1.it/>

⁷KTH, CSC, TMH, Stockholm, Sweden, <http://www.speech.kth.se/~rolf/>

Employment

2007–present	Researcher and Postdoc Instituto Superior Técnico, Institute for Systems and Robotics, Lisbon, Portugal
2006–2007	Researcher KTH, TMH, Stockholm, Sweden
2001–2006	PhD Student KTH, TMH, Stockholm, Sweden
1999–2001	Research Assistant KTH, TMH, Stockholm, Sweden
1996–1998	Technical Responsible for the Scientific Calculus Centre at the Department for Mathematical Methods and Models, La Sapienza University, Rome, Italy

Research Projects

2007–present	CONTACT, Learning and Development of Contextual Action http://eris.liralab.it/contact/ funded by the European commission (NEST-5010)
2004–2007	MILLE, Modelling Language Learning, funded by the Bank of Sweden.
2002–2004	Synface, synthesised talking face derived from speech for hard of hearing users of voice channels, http://www.speech.kth.se/synface/ funded by the European commission (IST-2001-33327)
1999–2002	Teleface, http://www.speech.kth.se/teleface/ funded by the Swedish Transport and Communications Research Board
1999–2000	COST Action 249: Continuous Speech Recognition Over the Telephone http://www.elis.ugent.be/cost249/

Teaching

2005–2006	Spectral Transforms (2F1120) KTH, School of Computer Science and Communication
2003	Speech Recognition Graduate School of Language Technology (GSLT)
2000–2002	Pattern Recognition (2E1395) KTH, School of Electrical Engineering
1999–present	Supervisor to several MSc theses.

Awards

2007–2008	National Science Foundation grant (Fundação para a Ciência e a Tecnologia)
-----------	--

Other Achievements

2006	Co-founder of SynFace AB (http://www.synface.com/)
------	--

Research Experience

General

My work has focused on the statistical characterisation of speech sounds for automatic speech recognition and analysis. The focus has been on large sets of speech recordings thus involving machine learning and data mining methods. Both theoretical aspects and practical implementation problems have been considered resulting in empirical and theoretical studies, but also in software solutions that perform real-time speech processing. There are three main areas of investigation explored during my Ph.D. studies:

- The first area aims at deriving the movements of a computer-animated face from the acoustic speech signal. The system is to be used as a lip-reading support for hearing impaired users of the telephone channel. A low-latency phoneme recogniser based on a hybrid of hidden Markov models (HMMs) and recurrent neural networks (RNNs) was developed and implemented for the purpose. Studies that were performed in this area include the evaluation of the system as a hearing aid (perception studies with hearing impaired subjects), and the analysis of the dynamic properties of the machine learning methods. The latter include the analysis of the interaction between the dynamical model imposed by the HMMs and by the RNNs, and the evaluation of an entropy-based method for the detection of phonetic boundaries. A working prototype was built in the form of a stand-alone application that works in conjunction with IP telephony software (e.g. Skype).
- The second area concerns the analysis of pronunciation variation in a language due to the geographical origin of the speaker. The main characteristic of these studies is the use of collections of speech recordings from a large population of speakers. Hidden Markov models, Gaussian mixture fitting, agglomerative hierarchical clustering and discriminant analysis are used in order to analyse ca 10^8 data points in a high dimensional feature space.
- The third area has focused on modelling the language acquisition process in new-born children on the basis of statistical principles. An incremental version of Model-Based Clustering was implemented in order to simulate the unsupervised emergence of acoustic categories in the first stages of language acquisition.

External Reviewer

WWTF, Five Senses, Call 2006, project proposal review
Acta Cybernetica

Technical Skills

Proficient

C/C++, Tcl/Tk, Perl, \LaTeX , Matlab, R (a language for statistical computing),
Specialised Software Packages: HTK, NICO

Familiar

PHP, Java, CGI

Miscellaneous

Citizenship: Italian, Swedish (naturalised)

Languages: Italian, Swedish, English

Publications

Journal Papers

Giampiero Salvi. Segment boundaries in low latency phonetic recognition. *Lecture Notes in Computer Science*, 3817:267–276, 2006.

Giampiero Salvi. Dynamic behaviour of connectionist speech recognition with strong latency constraints. *Speech Communication*, 48:802–818, 2006.

Giampiero Salvi. Segment boundary detection via class entropy measurements in connectionist phoneme recognition. *Speech Communication*, 2006. in press.

- Eva Agelfors, Jonas Beskow, Inger Karlsson, Jo Kewley, Giampiero Salvi, and Neil Thomas. User evaluation of the synface talking head telephone. *Lecture Notes in Computer Science*, 4061:579–586, 2006.
- Catherine Siciliano, Geoff Williams, Andrew Faulkner, and Giampiero Salvi. Intelligibility of an ASR-controlled synthetic talking face (abstract). *Journal of the Acoustical Society of America*, 115(5):2428, 2004.
- Giampiero Salvi. Developing acoustic models for automatic speech recognition in Swedish. *The European Student Journal of Language and Speech*, 1999.
- Tobias Öhman and Giampiero Salvi. Using HMMs and ANNs for mapping acoustic to visual speech. *TMH-QPSR*, 1-2:45–50, 1999.

Refereed Conference Papers

- Giampiero Salvi. Advances in regional accent clustering in Swedish. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, pages 2841–2844, 2005.
- Giampiero Salvi. Ecological language acquisition via incremental model-based clustering. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, pages 1181–1184, 2005.
- Giampiero Salvi. Segment boundaries in low latency phonetic recognition. In *Proceedings of Non Linear Speech Processing (NOLISP)*, 2005.
- Jonas Beskow, Inger Karlsson, Joe Kewley, and Giampiero Salvi. SYNFACE - A Talking Head Telephone for the Hearing-impaired. In K. Miesenberger, J. Klaus, W. Zagler, and D. Burger, editors, *Proceedings of International Conference Computers Helping People with Special Needs*, 2004.
- Karl-Erik Spens, Eva Agelfors, Jonas Beskow, Björn Granström, Inger Karlsson, and Giampiero Salvi. SYNFACE, a talking head telephone for the hearing impaired. In *IFHOH 7th World Congress*, 2004.
- Giampiero Salvi. Accent clustering in Swedish using the Bhattacharyya distance. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, August 2003.
- Giampiero Salvi. Using accent information in ASR models for Swedish. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, pages 2677–2680, September 2003.
- Giampiero Salvi. Truncation error and dynamics in very low latency phonetic recognition. In *Proceedings of Non Linear Speech Processing (NOLISP)*, Le Croisic, France, 2003.
- Inger Karlsson, Andrew Faulkner, and Giampiero Salvi. SYNFACE - a talking face telephone. In *Proceedings of European Conference on Speech Communication and Technology (Eurospeech)*, pages 1297–1300, 2003.
- Borge Lindberg, Finn Tore Johansen, Narada Warakagoda, Gunnar Lehtinen, Zdravko Kačič, Andrei Žgank, Kjell Elenius, and Giampiero Salvi. A noise robust multilingual reference recogniser based on SpeechDat(II). In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 2000.
- Finn Tore Johansen, Narada Warakagoda, Borge Lindberg, Gunnar Lehtinen, Zdravko Kačič, Andrei Žgank, Kjell Elenius, and Giampiero Salvi. The COST 249 SpeechDat multilingual reference recogniser. In *Proceedings of XLDB Workshop on Very Large Telephone Speech Databases*, Athens, Greece, 2000.
- Finn Tore Johansen, Narada Warakagoda, Borge Lindberg, Gunnar Lehtinen, Zdravko Kačič, Andrei Žgank, Kjell Elenius, and Giampiero Salvi. The COST 249 SpeechDat multilingual reference recogniser. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, 2000.

- Eva Agelfors, Jonas Beskow, Björn Granström, Magnus Lundeberg, Giampiero Salvi, Karl-Erik Spens, and Tobias Öhman. Synthetic visual speech driven from auditory speech. In *Proceedings of Audio-Visual Speech Processing (AVSP)*, Santa Cruz, USA, 1999.
- Eva Agelfors, Jonas Beskow, Martin Dahlquist, Björn Granström, Magnus Lundeberg, Giampiero Salvi, Karl-Erik Spens, and Tobias Öhman. Two methods for visual parameter extraction in the Teleface project. In *Proceedings of Fonetik*, Gothenburg, Sweden, 1999.
- Eva Agelfors, Jonas Beskow, Martin Dahlquist, Björn Granström, Magnus Lundeberg, Giampiero Salvi, Karl-Erik Spens, and Tobias Öhman. A synthetic face as a lip-reading support for hearing impaired telephone users - problems and positive results. In *Proceedings of the 4th European Conference on Audiology*, Oulo, Finland, 1999.