

## Dimitra Vergyri

---

SRI International  
333 Ravenswood Ave., Menlo Park, CA 94025-3493  
e-mail:dverg@speech.sri.com , phone:(650) 859-5698

### RESEARCH INTERESTS

Statistical Speech Recognition  
Statistical Machine Translation  
Speech Recognition for sparse-data languages  
Multilingual speech recognition and information extraction  
Language Modeling for highly inflected languages  
Audio-visual speech recognition  
Information Extraction

### EDUCATION

**Ph.D.** in Electrical and Computer Engineering, 2001  
The Johns Hopkins University, Baltimore, MD.  
**Dissertation:** Integration of Multiple Knowledge Sources in Speech Recognition using Minimum Error Training.  
**Advisor:** Frederick Jelinek.

**M.S.** in Electrical and Computer Engineering, 1996  
The Johns Hopkins University, Baltimore, MD.

**Diploma** in Electrical and Computer Engineering, June 1993  
National Technical University of Athens, Greece.  
**Diploma thesis:** Graphical User Interface for Asynchronous Transfer Mode Communication Control Center.

### RESEARCH EXPERIENCE

#### **SRI International Speech Technology and Research Laboratory**

*Research Engineer*

*October 2000 – present*

Worked on the following projects:

#### *Real-time Spoken Language Translation System for Tactical Use*

Funded by DARPA under TRANSTAC program (2004 – present)

Role: Led the speech recognition component development. Key personnel in machine translation development.

Developed speech recognition and machine translation systems for 2-way communication in the following language pairs: English-Iraqi, English-Malay, English-Farsi, English-Pashto, English-Dari.

#### *Rapid development of speech to speech translation systems*

Funded by DARPA as component of the TRANSTAC project (2006 – 2007)

Role: Co-PI

Investigated novel techniques which leverage limited amounts of data, as well as data and models from similar dialects, languages or domains, in order to make system components easily adaptable to the target languages and domains.

#### *Automatic Translation and information extraction from Arabic and Mandarin Broadcast Speech Sources*

Funded by DARPA under the GALE program (2005 – 2009).

Role: Co-PI, leading the ASR effort (2009).

The goal of the GALE DARPA program was to develop and apply computer software technologies to absorb, analyze and interpret huge volumes of speech and text in multiple languages, eliminating the need for linguists and lower-level military analysts and automatically providing relevant, distilled actionable information to military command and personnel in a timely fashion.

*Meeting Recognition and Understanding*

Funded by DARPA under the CALO project (2007 – 2009)

Worked on Speech Recognition of multi-part meetings, and on using adaptive learning techniques across a series of meeting recordings.

*Spoken Term Detection*

Internally funded SRI effort (2006)

Participated in the development of a Spoken term detection system (the indexing part) which was used for the 2006 NIST evaluations and performed the best in one of the audio conditions (English meetings' data).

*Dialectal Arabic Automatic Speech Recognition*

Funded by DARPA under the EARS program (2002 – 2004)

Led the development of Conversational Telephone Speech (CTS) recognition systems for Egyptian Colloquial Arabic (2003) and Levantine Colloquial Arabic (2004), which were used for the NIST evaluations for the corresponding years. Researched the topics of Factored Language Modeling for highly inflected languages, Cross-Dialectal Acoustic Data sharing and Automatic Diacritization of Arabic Script Orthography.

*Pashto (Afghani-language) Speech-to-Speech Translation*

Funded by DARPA under the Babylon project (2003 – 2005)

Developed an Automatic Speech Recognition system for Pashto, which was used as the input to the SRI machine-translation system. The work was particularly challenging due to the limited data resources, and the platform requirements: it runs on a portable device and has been evaluated by military, medical, and security personnel in Afghanistan.

*English Large Vocabulary Automatic Speech Recognition*

Funded by DARPA under the EARS program (2002 – 2005)

Participated in the development of state-of-the-art acoustic and language models for various systems, and the algorithm for optimal combination of system components. Participated with the SRI team in government (NIST) evaluations for Conversational Telephone Speech (CTS) (years 2002, 2003, 2004) and Broadcast News (BN) (2003, 2004) systems. Investigated the effect of high volumes of poor quality training data in ASR and the combination of multiple knowledge sources in an ASR system.

*Prosodic information for ASR*

Funded by DARPA and partially by NSF (2001 – 2002)

Investigated the use of prosodic information for improving word recognition accuracy.

*Speech In Noisy Environments (SPINE)*

Internally funded effort (2001)

Participated with SRI team in the SPINE NIST evaluations, in which the SRI system had the best performance.

**LIMSI/CNRS, Orsay, France**

*Visiting Researcher*

*Feb 2010-May 2010*

Visiting Research Position sponsored by Digiteo.

**University Paris Sud, Orsay, France**

*One-month Visiting Professor*

*March 2009*

Worked on arabic cross-dialectal ASR.

**Summer Workshops at the Center for Language and Speech Processing, The Johns Hopkins University**

*Senior group member*

*July 2002 – August 2002*

Novel Speech Recognition models for Arabic (Group Leader: K. Kirchhoff).

*Senior group member*  
Audio Visual Speech Recognition (Group Leader: C. Neti)

*July 2000 – August 2000*

*Student group member*  
Language Modeling for Conversational Speech Recognition (Group Leader: R. Rosenfeld)

*July 1995 – August 1995*

**Center for Language and Speech Processing, The Johns Hopkins University**

*Research Assistant*

*September 1995 – August 2000*

Developed an approach to dynamically vary the weighting of the language model in a recognition system to compensate for variable acoustic model inadequacies. Generalized the approach to combine dynamically multiple knowledge sources in order to minimize recognition errors. Applied the approach to multi-lingual acoustic model combination, for sparse data language speech recognition, and to audio-visual speech recognition. (Advisor: F. Jelinek)

**PROFESSIONAL  
ACTIVITIES**

Associate Editor, IEEE Transactions for Audio, Speech and Language Processing, 2009 – present  
NSF Panelist, Human Language Technologies Area, 2008

Local Organization Chair for IEEE Spoken Language Technologies Workshop (SLT) 2010

Session Chair for Int. Conference of Acoustic Speech and Signal Processing (ICASSP) 2008

Reviewer, Intl. Conference on Acoustic and Computational Linguistics (ACL) 2001, 2004, 2005

Reviewer, Int. Conference of Acoustic Speech and Signal Processing (ICASSP) 2007, 2009

Reviewer, IEEE Transactions on Audio, Speech and Language Processing (T-ASL) 2001, 2003, 2006, 2007, 2009

Reviewer, Intl. Conference on Neural Information Processing Systems (NIPS) 2004, 2005

Reviewer, Journal of Computer Speech and Language (CSL) 2005, 2008, 2009

Reviewer, IEEE workshop on Automatic Speech Recognition and Understanding (ASRU) 2009

**PUBLICATIONS**

**Journal publications:**

1. G. Tur, A. Stolcke, L. Voss, S. Peters, D. Hakkani-Tur, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, F. Yang, "CALO Meeting Assistance System", IEEE Transactions on Audio, Speech and Language Processing, accepted for publication.
2. K. Kirchhoff, D. Vergyri, J. Bilmes, K. Duh, and A. Stolcke, "Morphology-based language modeling for conversational Arabic speech recognition," Computer Speech and Language, vol. 20, pp. 589–608, October 2006
3. Stolcke, B. Chen, H. Franco, V. R. R. Gadde, M. Graciarena, M.-Y. Hwang, K. Kirchhoff, N. Morgan, X. Lin, T. Ng, M. Ostendorf, K. Snmez, A. Venkataraman, D. Vergyri, W. Wang, J. Zheng, and Q. Zhu, "Recent innovations in speech-to-text transcription at SRI-ICSI-UW," IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, pp. 1729–1744, September 2006
4. K. Kirchhoff, D. Vergyri. "Cross-Dialectal Data Sharing for Acoustic Modeling in Arabic Speech Recognition". Speech Communication journal, Volume 46, Issue 1, pages 37-51, May 2005.

**International Conference Publications:**

1. M. Akbacak, H. Franco, M. Frandsen, S. Hasan, H. Jameel, A. Kathol, S. Khadivi, X. Lei, A. Mandal, S. Mansour, K. Precoda, C. Richey, D. Vergyri, W. Wang, M. Yang, J. Zheng, "Recent Advances in SRI's IraqComm: Iraqi Arabic-English Speech-to-Speech Translation System", in Proc. IEEE ICASSP, (Taipei), April 2009
2. D. Vergyri, A. Stolcke, G. Tur, "Exploiting user feedback for language model adaptation in meeting recognition," in Proc. IEEE ICASSP, (Taipei), April 2009
3. A. Mandal, D. Vergyri, W. Wang, J. Zheng, A. Stoleke, G. Tur, D. Hakkani-Tr, and N. F. Ayan, "Efficient data selection for machine translation," in Proc. IEEE Workshop on Spoken Language Technology, (Goa, India), pp. 261–264, December 2008

4. L. Voss, J. Dowding, B. Favre, R. Fernandez, M. Frampton, M. Frandsen, C. Frederickson, M. Graciarena, D. Hakkani-Tur, D. Kintzing, K. Leveque, S. Mason, J. Niekrasz, S. Peters, M. Purver, K. Riedhammer, E. Shriberg, J. Tien, D. Vergyri, F. Yang, "The CALO meeting speech recognition and understanding system," in Proc. IEEE Workshop on Spoken Language Processing, (Goa, India), pp. 69–72, December 2008
5. D. Vergyri, A. Mandal, W. Wang, A. Stolcke, J. Zheng, M. Graciarena, D. Rybach, C. Gollan, R. Schlter, K. Kirchhoff, A. Faria, and N. Morgan, "Development of the SRI/Nightingale Arabic ASR system," in Proc. Interspeech, (Brisbane, Australia), September 2008
6. B. Peintner, W. Jarrold, D. Vergyri, C. Richey, M. G. Tempini, and J. Ogar, "Learning diagnostic models using speech and language measures," in Proc of the 30th Annual International IEEE EMBS Conference, (Vancouver, British Columbia, Canada), August 2008
7. M. Akbacak, D. Vergyri, A. Stolcke, "Open-Vocabulary Spoken Term Detection Using Graphone-Based Hybrid Recognition Systems", ICASSP 2008, Las Vegas, USA
8. H. Lui, J. Bilmes, D. Vergyri and K. Kirchhoff, "OOV Detection by Joint Word/Phone Lattice Alignment", Proceedings of ASRU 2007, Kyoto, Japan
9. D. Vergyri, I. Shafran, A. Stolcke, R. R. Gadde, M. Akbacak, B. Roark, and W. Wang, "The SRI/OGI 2006 Spoken Term Detection system," in Proc. Interspeech/Eurospeech, (Antwerp), pp. 2393–2396, August 2007.
10. A. Stolcke, F. Grzl, M.-Y. Hwang, X. Lei, N. Morgan, and D. Vergyri, "Cross-domain and cross-language portability of acoustic features estimated by multilayer perceptrons," in Proc. IEEE ICASSP, vol. 1, (Toulouse), pp. 321–324, May 2006.
11. W. Wang and D. Vergyri, "The use of word n-grams and parts of speech for hierarchical cluster language modeling," in Proc. IEEE ICASSP, vol. 1, (Toulouse), pp. 1057–1060, May 2006.
12. W. Wang and D. Vergyri, "Using word n-grams and parts of speech for hierarchical cluster language modeling," in Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (Toulouse, France), May 2006.
13. A. Kathol, K. Precoda, D. Vergyri, W. Wang, and S. Riehemann, "Speech translation for low-resource languages: The case of Pashto," in Proc. Eurospeech, (Lisbon), pp. 2273–2276, September 2005.
14. S. Narayanan, P. G. Georgiou, A. Sethy, D. Wang, M. Bulut, S. Sundaram, E. Ettelaie, S. Ananthakrishnan, H. Franco, K. Precoda, D. Vergyri, J. Zheng, W. Wang, R. R. Gadde, V. Abrash, M. Frandsen, and C. Richey, "Speech recognition engineering issues in speech to speech translation system desing for low resource languages and domains," in Proc. IEEE ICASSP, vol. 5, (Toulouse), pp. 1209–1212, May 2006
15. D. Vergyri, K. Kirchhoff, V.R.R. Gadde, A. Stolcke, J. Zheng, "Development of a Conversational Telephone Speech Recognizer for Levantine Arabic", in Proc of the 9th European Conference on Speech Communication and Technology (Interspeech), Lisbon, Portugal, September 2005
16. A. Kathol, K. Precoda, D. Vergyri, W. Wang, S. Riehemann, "Speech Translation for Low-Resource Languages: The Case of Pashto", in Proc of the 9th European Conference on Speech Communication and Technology (Interspeech), Lisbon, Portugal, September 2005.
17. D. Vergyri, K. Kirchhoff, K. Duh, and A. Stolcke, "Morphology-based language modeling for Arabic speech recognition," in Proc .of the International Conference on Spoken Language Processing (ICSLP), (Jeju, Korea), October 2004.
18. A. Venkataraman, A. Stolcke, W. Wang, D. Vergyri, V. R. R. Gadde, and J. Zheng, "An efficient repair procedure for quick transcriptions," in Proc. of the International Conference on Spoken Language Processing (ICSLP), (Jeju, Korea), October 2004.
19. D. Vergyri and K.Kirchhoff, "Automatic Diacritization of Arabic for Acoustic Modeling in Speech Recognition", Computational Linguistics (COLING) Workshop on Arabic-script Based Languages, Geneva, Switzerland, August 2004.

20. K. Kirchhoff and D. Vergyri, "Cross-dialectal acoustic data sharing for Arabic speech recognition," in Proc. of IEEE International Conference of Acoustic Speech and Signal Processing (ICASSP), vol. 1, (Montreal), pp. 765–768, May 2004.
21. M. Graciarena, H. Franco, J. Zheng, D. Vergyri, and A. Stolcke, "Voicing feature integration in SRI's Decipher LVCSR system," in Proc. of IEEE International Conference of Acoustics Speech and Signal Processing (ICASSP), vol. 1, (Montreal), pp. 921–924, May 2004.
22. K. Precoda, H. Franco, A. Dost, M. Frandsen, J. Fry, A. Kathol, C. Richey, S. Riehemann, D. Vergyri, J. Zheng, and C. Culy, "Limited-domain speech-to-speech translation between English and Pashto," in Proc. HLT/NAACL 2004 Demonstrations, (Boston, MA), pp. 9–12, January 2004.
23. H. Franco, J. Zheng, K. Precoda, F. Cesari, V. Abrash, D. Vergyri, A. Venkataraman, H. Bratt, C. Richey, and A. Sarich, "Development of phrase translation systems for handheld computers: from concept to field," in Proc. of the European Conference on Speech Communication and Technology (Eurospeech), (Geneva, Switzerland), 2003.
24. K. Kirchhoff, J. Bilmes, S. Das, N. Duta, M. Egan, G. Ji, F. He, J. Henderson, D. Liu, M. Noamany, P. Schone, R. Schwartz, and D. Vergyri, "Novel approaches to Arabic speech recognition: Report from the 2002 Johns Hopkins Summer Workshop," in Proc. of the IEEE International Conference of Acoustics Speech and Signal Processing (ICASSP), vol. 1, (Hong Kong), pp. 344–347, April 2003.
25. D. Vergyri, A. Stolcke, V. R. R. Gadde, L. Ferrer, and E. Shriberg, "Prosodic knowledge sources for automatic speech recognition," in Proc. of the IEEE International Conference of Acoustics Speech and Signal Processing (ICASSP), vol. 1, (Hong Kong), pp. 208–211, April 2003.
26. B. Hodjat, H. Franco, H. Bratt, K. Precoda, A. Stolcke, A. Venkataraman, D. Vergyri, and J. Zheng, "Iterative statistical language model generation for use with an agent-oriented natural language interface," in 10th International Conference on Human-Computer Interaction, (Crete), June 2003.
27. V. R. R. Gadde, A. Stolcke, D. Vergyri, J. Zheng, K. Sonmez, and A. Venkataraman, "Building an ASR system for noisy environments: SRI's 2001 SPINE evaluation system," in Proc. of the International Conference on Spoken Language Processing (ICSLP), vol. 3, (Denver, CO), pp. 1577–1580, September 2002.
28. C. Neti, G. Potamianos, J. Luettin, Ian Matthews, H. Glotin, and D. Vergyri. "Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop". in Proc. of the IEEE Workshop on Multimedia Signal Processing, Cannes, 2001.
29. H. Glotin, D. Vergyri, C. Neti, G. Potamianos, and J. Luettin, "Weighting schemes for audio-visual fusion in speech recognition", in Proc. of the IEEE International Conference of Acoustics Speech and Signal Processing (ICASSP), vol. I, (Salt Lake City, Utah), May 2001.
30. D. Vergyri, W. Byrne and S. Tsakalidis. "Minimum Risk Acoustic Clustering for Multilingual Acoustic Model Combination," in Proc. of the International Conference of Spoken Language Processing (ICSLP), Beijing, 2000.
31. Dimitra Vergyri. "Use of Word Level Side Information to improve Speech Recognition". In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Istanbul, Turkey, 2000.
32. P. Beyerlein, W. Byrne, J.M. Huerta, S. Khudanpur, B. Marthi, J. Morgan, N. Peterek, J. Picone, D. Vergyri and W. Wang "Towards Language Independent Acoustic Modeling". In Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Istanbul, Turkey, 2000.
33. R. Rosenfeld, R. Agarwal, W. Byrne, R. Iyer, M. Liberman, L. Shriberg, J. Unverferth, D. Vergyri, and E. Vidal. "Error Analysis and Disfluency Modeling in the Switchboard Domain", In Proc. Addendum of the International Conference of Spoken Language Processing (ICSLP), Philadelphia, PA, p. 15, October 1996.

TEACHING  
EXPERIENCE

**Dept. of Electrical and Computer Engineering, The Johns Hopkins University**

*Teaching Assistant*

*Sept 1993 - May 1995*

Circuits; Signals and Systems;  
Information Extraction from Speech and Text.

PERSONAL  
INFORMATION

Greek citizen.

Permanent US resident.

Languages: Greek, English, French(basic), German(basic).