

# POTAMIANOS, Gerasimos

## Office Address:

IBM Thomas J. Watson Research Center  
Route 134, Room 18-125A  
Yorktown Heights, NY 10598  
Tel.: +1 914 945 2433 / Mobile: +1 973 723 0959  
E-mail: gpotam@us.ibm.com  
URL: [www.research.ibm.com/people/g/gerasimos.potamianos](http://www.research.ibm.com/people/g/gerasimos.potamianos)

## Home Address:

11 Lake Street, Apt. 7L  
White Plains, NY 10603  
Tel.: +1 914 682 5675

**Research** Multimodal speech processing with applications to human-computer interaction and ambient intelligence. Particular emphasis on audio-visual speech processing, automatic speech recognition, multimedia signal processing and fusion, computer vision for human detection and tracking.

**Education** **Ph.D.**, November 1994, Electrical and Computer Engineering,  
THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland. GPA: 3.97/4.00  
*Advisor*: Prof. John Goutsias. *Dissertation*: "Stochastic Simulation Algorithms for Partition Function Estimation of Markov Random Field Images."

**M.S.E.**, June 1990, Electrical and Computer Engineering,  
THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland. GPA: 3.97/4.00  
*Advisor*: Prof. John Goutsias.

**Diploma**, July 1988, Electrical Engineering and Computer Science,  
NATIONAL TECHNICAL UNIVERSITY OF ATHENS, Greece. GPA: 9.5/10.0  
*Advisor*: Prof. John Diamessis. *Thesis*: "Designing 2-D IIR Digital Filters Using Continued Fractions."

**Experience** Sept. 1999 – Present: **Research Staff Member and Manager** (since Feb. 2007), *Multimodal Conversational Solutions Department*, IBM T. J. WATSON RESEARCH CENTER, Yorktown Heights, New York.

Leading the development and integration of multiple input modalities, natural language processing, and dialog management into conversational platforms and solutions, as manager of a research group within the User Interface / Human Language Technology Department at the IBM T. J. Watson Research Center.

Own research work focuses on audio-visual speech technologies with applications to human-computer interaction and ambient intelligence. Effort has been concentrating on a variety of problems including 2D and 3D person tracking, visual speech representation, fusion algorithms for large-vocabulary continuous audio-visual speech recognition, bimodal speech enhancement, speaker recognition, audio-visual speech activity detection, and real-time algorithmic implementation, with emphasis on applications to the automobile and smart rooms/homes. Recent efforts have been driven by three European research projects (CHIL, NETCARITY, and DICIT).

Aug. 1996 – Aug. 1999: **Senior Technical Staff Member**, *Speech and Image Processing Lab*, AT&T LABS-RESEARCH, Murray Hill / Florham Park, New Jersey.

Conducted research on audio-visual automatic speech recognition and synthesis. Work accomplishments include: Algorithms for speaker independent face localization and segmentation, visual feature extraction, audio-visual information fusion, collection of large audio-visual corpora, development of a real-time audio-visual speech recognition prototype system, and algorithms for visual unit selection in audio-visual speech synthesis.

Nov. 1994 – June 1996: **Postdoctoral Fellow**, *Center for Language and Speech Processing*, THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland.

Conducted research, under the direction of Prof. F. Jelinek, towards the development of statistical language models for continuous speech recognition, using decision trellises, decision trees, as well as traditional n-gram approaches.

July 1994 – Oct. 1994: **Summer Intern / Coop Member of Technical Staff**, *Integrated Systems Laboratory*, TEXAS INSTRUMENTS, Dallas, Texas.

Developed a statistical channel model for digital video transmission over a coaxial cable system. Published two Texas Instruments technical reports.

Jan. 1994 – May 1994: **Teaching Assistant**, *Center for Language and Speech Processing and Department of Electrical and Computer Engineering*, THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland.

Helped with the design of research projects for the “Statistical Models for Language Analysis” graduate-level course, as well as with the grading for the “Image Analysis” graduate-level course.

June 1990 – Dec. 1993: **Research Assistant**, *Department of Electrical and Computer Engineering*, THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland.

Developed, implemented, and analyzed new algorithms for partition function estimation, maximum likelihood parameter estimation, and hypothesis testing of Markov random field images.

Sept. 1989 – May 1990: **Instructor**, *G.W.C. Whiting School of Engineering, Continuing Professional Programs*, THE JOHNS HOPKINS UNIVERSITY, Baltimore, Maryland.

Taught two Electrical Engineering laboratory courses.

July – Aug. 1987: **Summer Intern**, WSZ-ELEKTRONIK GMBH, Wolfratshausen, Germany.

Participated in a program of the International Association for the Exchange of Students for Technical Experience (IAESTE). Took active part in the design, manufacturing and testing of amplifier systems, a project supervised by Siemens, Munich.

## Skills

*Research:* In-depth knowledge of signal and image processing, speech recognition, language modeling, pattern recognition, multimedia signal processing, and computer vision. Seven years academic and eleven years of industrial research experience. Recent work focuses on audio-visual speech processing, speech recognition, and computer vision for robust and natural human-computer interaction and ambient intelligence.

*Supervision (IBM):* Manager, Multimodal Conversational Solutions Department (2007–). Team lead for audio-visual speech technologies research work (2003–). Mentor and supervisor of seven summer students (2000–).

*Projects, Proposals, Academic Collaborations:* Has participated in various proposal writing efforts resulting in funding by DARPA, the European Union (FP6 and FP7), internal IBM divisions, and external IBM customers. Currently in collaboration within multi-partner consortia in the NETCARITY and DICIT EU-funded projects. Successfully completed the CHIL FP6 integrated project, having led the speech-technology work-package. Additional current and past collaborations with faculty at US Universities (Beckmann Institute at the University of Illinois, Urbana-Champaign, ECE Department of Northwestern University, the Robotics Institute at Carnegie Mellon University, and the CSE Department at Ohio State).

*Teaching:* Instructor and teaching assistant at the Johns Hopkins University (1990, 1994); short-course and tutorial speaker (ELSNET 2001; ICIP 2003); has given invited lectures at Columbia University as part of a course on automatic speech recognition (Spring 2004; Fall 2005).

*Coursework:* Has completed numerous courses in signal and image processing, pattern recognition, information theory, control, nonlinear systems, probability, and statistics.

*Computer Skills:* Extensive programming experience in C, C++, PERL, Fortran, Pascal; good working knowledge of Matlab, Splus, Mathematica, HTK, waves+.

*Languages:* Fluent in Greek (mother tongue) and English; some knowledge of German; elementary reading capability of French.

## Publications

- Eleven Articles in Journals and Books (Published / In Press).
- Three Additional Book Chapters Submitted.
- Sixty-Eight Articles in Conference and Workshop Proceedings.
- Seven Technical Reports.
- Two Short Courses / Tutorials.
- Five Granted Patents.
- Numerous Works Citing these Publications (Sample List of 402 Attached).

## Affiliations

Member of the International Institute of Electrical and Electronics Engineers (IEEE).

Member of the National Technical Chamber of Greece.

- Honors**
- Co-author of paper that was awarded the *Best Student Paper Award* at the 2007 Interspeech Conference, in Antwerp, Belgium.
  - Contributed significantly to the 2006 *North American Frost & Sullivan Award for Excellence in Research* in the speech recognition field awarded to the IBM Corporation. Award explicitly emphasizes IBM research work on audio-visual speech recognition.
  - Received the *Best Paper Award* at the 2005 International Conference on Multimedia and Expo. (ICME), in Amsterdam, the Netherlands (among over 800 submitted papers).
  - Received a number of internal IBM Research awards for work on audio-visual speech recognition, including the 2002 *IBM Research Accomplishment Award*.
  - Invited to the summer 2000 research *workshop* (WS2000) at the Center for Language and Speech Processing at the Johns Hopkins University (July – Aug. 2000).
  - Awarded a *post-doctoral* fellowship with the Center for Language and Speech Processing at the Johns Hopkins University (Nov. 1994 – July 1996).
  - Awarded a *post-doctoral* fellowship through the Human Capital and Mobility Program of the Commission of the European Communities for post-doctoral study with IRISA, Rennes, France (1994-1995), that he later declined.
  - Awarded one of the five annual *Abel Wolman Fellowships* of the G.W.C. Whiting School of Engineering for the first academic year as an entering graduate student at the Johns Hopkins University (Sept. 1988 – May 1989).
  - Given *full financial support* during graduate study at the Johns Hopkins University, as an instructor, research, and teaching assistant (Sept. 1989 – June 1994).
  - Awarded *full tuition fellowship* for the entire duration of graduate study at the Johns Hopkins University (Sept. 1988 – Nov. 1994).
  - Awarded prizes from the *Greek National Scholarship Foundation* for academic performance during the years of undergraduate study (1983-1988). Received a financial award from the *National Technical Chamber of Greece* (1987).
- Activities**
- Guest editor of the special issue on Multimodal Speech Processing for the IEEE Transactions on Audio, Speech and Language Processing, September 2008.
  - Special session organizer on “Multimodal Speech Technology” at Acoustics’08 Conference.
  - Guest editor of the special issue on Joint Audio-Visual Speech Processing of the EURASIP Journal of Applied Signal Processing, November 2002.
  - Technical Program Committee Member, IMAP 2007 Workshop.
  - Participant at the CLSP’00 summer workshop, Center for Language and Speech Processing, the Johns Hopkins University, Baltimore, MD (July–Aug. 2000).
  - Reviewer for numerous journals (Speech Communication J.; Computer Speech and Language; J. of Visual Communication and Image Representation; IEEE Trans. on Image Processing; IEEE Trans. on Signal Processing; IEEE Trans. on Systems, Man, and Cybernetics; IEEE Trans. on Audio, Speech and Language Processing; IEEE Trans. on Multimedia).
  - Reviewer for numerous conferences (ASRU, AVSP, ICASSP, ICME, ICMI, MLMI, MMSP).
- Presentations**
- Keynote speaker at the VisHCI 2006 workshop, Canberra, Australia.
  - Panelist at the MMSP 2006 workshop, Victoria, Canada.
  - Tutorial speaker at the ICIP 2003 conference, Barcelona, Spain.
  - Guest lecturer at the “Topics in Signal Processing: Speech Recognition” course, Electrical Eng. Dept., Columbia University, NY, Spring Semester, 2004; Fall Semester, 2005.
  - Plenary speaker at the AVSP 2003 workshop, St. Jorioz, France.
  - Instructor at the ELSNET 2001 summer school, Prague, Czech Republic.
  - Forty conference paper presentations (1989-2007).
  - Numerous invited presentations at Universities and Industrial Research Labs.
- Personal**
- Born in 1965;
  - Greek Citizen; US Permanent Resident;
  - Military Service: Hellenic Navy, Completed, July 2003.

## Journal Articles and Book Chapters

- 1) **G. Potamianos**, C. Neti, J. Luetttin, and I. Matthews, "Audio-Visual Automatic Speech Recognition: An Overview," (To Appear) *Audio-Visual Speech Processing*, E. Vatikiotis-Bateson, G. Bailly, and P. Perrier (Eds.), MIT Press, ISBN: 0-26-222078-4, 2008.
- 2) D. Mostefa, N. Moreau, K. Choukri, **G. Potamianos**, S.M. Chu, A. Tyagi, J.R. Casas, J. Turmo, L. Christoforetti, F. Tobia, A. Pnevmatikakis, V. Mylonakis, F. Talantzis, S. Burger, R. Stiefelhagen, K. Bernardin, and C. Rochet, "The CHIL audiovisual corpus for lecture and meeting analysis inside smart rooms," (To Appear) *Journal of Language Resources and Evaluation*, March 2008.
- 3) Z. Zhang, **G. Potamianos**, A.W. Senior, and T.S. Huang, "Joint face and head tracking inside multi-camera smart rooms," *Signal, Image and Video Processing*, vol. 1, pp. 163–178, 2007.
- 4) **G. Potamianos**, "Audio-Visual Speech Recognition," Short Article, *Encyclopedia of Language and Linguistics, Second Edition, (Speech Technology Section – Computer Understanding of Speech)*, K. Brown (Ed. In Chief), Elsevier, Oxford, United Kingdom, ISBN: 0-08-044299-4, vol. 11, pp. 800–805, 2006.
- 5) P.S. Aleksic, **G. Potamianos**, and A.K. Katsaggelos, "Exploiting Visual Information in Automatic Speech Processing," *Handbook of Image and Video Processing*, Second Edition, Al. Bovik (Ed.), ch. 10.8, pp. 1263–1289, Elsevier Academic Press, Burlington, MA, ISBN: 0-12-119792-1, 2005.
- 6) J. Huang, **G. Potamianos**, J. Connell, and C. Neti. "Audio-visual speech recognition using an infrared headset," *Speech Communication*, vol. 44, no. 4, pp. 83–96, 2004.
- 7) **G. Potamianos**, C. Neti, G. Gravier, A. Garg, and A.W. Senior, "Recent advances in the automatic recognition of audio-visual speech," Invited, *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1306–1326, 2003.
- 8) **G. Potamianos**, C. Neti, G. Iyengar, A.W. Senior, and A. Verma, "A cascade visual front end for speaker independent automatic speechreading," *International Journal of Speech Technology, Special Issue on Multimedia*, vol. 4, pp. 193–208, 2001.
- 9) **G. Potamianos** and F. Jelinek, "A study of n-gram and decision tree letter language modeling methods," *Speech Communication*, vol. 24, no. 3, pp. 171–192, 1998.
- 10) **G. Potamianos** and J. Goutsias, "Stochastic approximation algorithms for partition function estimation of Gibbs random fields," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1948–1965, 1997.
- 11) **G.G. Potamianos** and J. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simulations," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1322–1332, 1993.

## Book Chapters Submitted

- 1) P. Lucey, **G. Potamianos**, and S. Sridharan, "Visual Speech Recognition Across Multiple Views," (submitted to:) *Visual Speech Recognition: Lip Segmentation and Mapping*, A. Wee-Chung Liew and S. Wang (Eds.), Information Science Publishing Press, 2008.
- 2) **G. Potamianos**, L. Lamel, M. Wölfel, J. Huang, E. Marcheret, C. Barras, J. McDonough, J. Hernando, D. Macho, and C. Nadeu, "Automatic Speech Recognition in CHIL," (submitted to:) *Computers in the Human Interaction Loop*, A. Waibel and R. Stiefelhagen (Eds.), Springer, 2008.
- 3) K. Bernardin, R. Stiefelhagen, A. Pnevmatikakis, O. Lanz, A. Brutti, J. Casas, and **G. Potamianos**, "Person Tracking in CHIL," (submitted to:) *Computers in the Human Interaction Loop*, A. Waibel and R. Stiefelhagen (Eds.), Springer, 2008.

## Conference Publications

- 1) A. Tyagi, J.W. Davis, and **G. Potamianos**, "Steepest descent for efficient covariance tracking," (to appear in:) *Proc. IEEE Work. Motion and Video Computing (WMVC)*, Copper Mountain, Colorado, 2008.
- 2) V. Libal, J. Connell, **G. Potamianos**, and E. Marcheret, "An embedded system for in-vehicle visual speech activity detection," *Int. Work. Multimedia Signal Process. (MMSP)*, pp. 255–258, Chania, Greece, 2007.
- 3) P. Lucey, **G. Potamianos**, and S. Sridharan, "A unified approach to multi-pose audio-visual ASR," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 650–653, Antwerp, Belgium, 2007.
- 4) J. Huang, E. Marcheret, K. Visweswariah, V. Libal, and **G. Potamianos**, "Detection, diarization, and transcription of far-field lecture speech," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 2161–2164, Antwerp, Belgium, 2007.
- 5) J. Huang, E. Marcheret, K. Visweswariah, and **G. Potamianos**, "The IBM RT07 evaluation systems for speaker diarization on lecture meetings," (To Appear) *Proc. Rich Transcription Evaluation Work. (RT)*, Baltimore, Maryland, 2007.
- 6) J. Huang, E. Marcheret, K. Visweswariah, V. Libal, and **G. Potamianos**, "The IBM Rich Transcription Spring 2007

- speech-to-text systems for lecture meetings,” (To Appear) *Proc. Rich Transcription Evaluation Work. (RT)*, Baltimore, Maryland, 2007.
- 7) P. Lucey, **G. Potamianos**, and S. Sridharan, “An extended pose-invariant lipreading system,” *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 176–180, Hilvarenbeek, The Netherlands, 2007.
  - 8) A. Tyagi, M. Keck, J.W. Davis, and **G. Potamianos**, “Kernel-based 3D tracking,” *Proc. IEEE Int. Work. Visual Surveillance (VS/CVPR)*, Minneapolis, Minnesota, 2007.
  - 9) A. Tyagi, **G. Potamianos**, J.W. Davis, and S.M. Chu, “Fusion of multiple camera views for kernel-based 3D tracking,” *Proc. IEEE Work. Motion and Video Computing (WMVC)*, Austin, Texas, 2007.
  - 10) E. Marcheret, V. Libal, and **G. Potamianos**, “Dynamic stream weight modeling for audio-visual speech recognition,” *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 4, pp. 945–948, Honolulu, Hawaii, 2007.
  - 11) **G. Potamianos** and P. Lucey, “Audio-visual ASR from multiple views inside smart rooms,” *Proc. Int. Conf. Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp. 35–40, Heidelberg, Germany, 2006.
  - 12) Z. Zhang, **G. Potamianos**, S.M. Chu, J. Tu, and T.S. Huang, “Person tracking in smart rooms using dynamic programming and adaptive subspace learning,” *Proc. Int. Conf. Multimedia Expo. (ICME)*, pp. 2061–2064, Toronto, Canada, 2006.
  - 13) P. Lucey and **G. Potamianos**, “Lipreading using profile versus frontal views,” *Proc. IEEE Work. Multimedia Signal Process. (MMSP)*, pp. 24–28, Victoria, Canada, 2006.
  - 14) A.W. Senior, **G. Potamianos**, S. Chu, Z. Zhang, and A. Hampapur, “A comparison of multicamera person-tracking algorithms,” *Proc. IEEE Int. Work. Visual Surveillance (VS/ECCV)*, Graz, Austria, 2006.
  - 15) **G. Potamianos** and Z. Zhang, “A joint system for single-person 2D-face and 3D-head tracking in CHIL seminars,” *Multimodal Technologies for Perception of Humans: First Int. Eval. Work. on Classification of Events, Activities and Relationships, CLEAR 2006*, R. Stiefelhagen and J. Garofolo (Eds.), LNCS vol. 4122, pp. 105–118, Southampton, United Kingdom, 2006.
  - 16) Z. Zhang, **G. Potamianos**, M. Liu, and T. Huang, “Robust multi-view multi-camera face detection inside smart rooms using spatio-temporal dynamic programming,” *Proc. Int. Conf. Automatic Face and Gesture Recog. (FGR)*, Southampton, United Kingdom, 2006.
  - 17) E. Marcheret, **G. Potamianos**, K. Visweswariah, and J. Huang, “The IBM RT06s evaluation system for speech activity detection in CHIL seminars,” *Proc. RT06s Evaluation Work. – held with Joint Work. on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, S. Renals, S. Bengio, and J.G. Fiscus (Eds.), LNCS vol. 4299, pp. 323–335, Washington DC, 2006.
  - 18) J. Huang, M. Westphal, S. Chen, O. Siohan, D. Povey, V. Libal, A. Soneiro, H. Schulz, T. Ross, and **G. Potamianos**, “The IBM Rich Transcription Spring 2006 speech-to-text system for lecture meetings,” *Proc. RT06s Evaluation Work. – held with Joint Work. on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, S. Renals, S. Bengio, and J.G. Fiscus (Eds.), LNCS vol. 4299, pp. 432–443, Washington DC, 2006. Washington DC, 2006.
  - 19) **G. Potamianos** and P. Scanlon, “Exploiting lower face symmetry in appearance-based automatic speechreading,” *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 79–84, Vancouver Island, Canada, 2005.
  - 20) S.M. Chu, E. Marcheret, and **G. Potamianos**, “Automatic speech recognition and speech activity detection in the CHIL smart room,” *Proc. Joint Work. on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, LNCS vol. 3869, pp. 332–343, Edinburgh, United Kingdom, 2005.
  - 21) Z. Zhang, **G. Potamianos**, A. Senior, S. Chu, and T. Huang, “A joint system for person tracking and face detection,” *Proc. Int. Work. Human-Computer Interaction (ICCV 2005 Work. on HCI)*, pp. 47–59, Beijing, China, 2005.
  - 22) E. Marcheret, K. Visweswariah, and **G. Potamianos**, “Speech activity detection fusing acoustic phonetic and energy features,” *Proc. Europ. Conf. Speech Comm. Technol. (Interspeech)*, pp. 241–244, Lisbon, Portugal, 2005.
  - 23) J. Jiang, **G. Potamianos**, and G. Iyengar, “Improved face finding in visually challenging environments,” *Proc. Int. Conf. Multimedia Expo.*, Amsterdam, The Netherlands, 2005.
  - 24) D. Macho, J. Padrell, A. Abad, C. Nadeu, J. Hernando, J. McDonough, M. Wölfel, U. Klee, M. Omologo, A. Brutti, P. Svaizer, **G. Potamianos**, and S.M. Chu, “Automatic speech activity detection, source localization, and speech recognition on the CHIL seminar corpus,” *Proc. Int. Conf. Multimedia Expo.*, Amsterdam, The Netherlands, 2005.
  - 25) P. Scanlon, **G. Potamianos**, V. Libal, and S.M. Chu, “Mutual information based visual feature selection for lipreading,” *Proc. Int. Conf. Spoken Lang. Process.*, Jeju Island, Korea, 2004.
  - 26) E. Marcheret, S.M. Chu, V. Goel, and **G. Potamianos**, “Efficient likelihood computation in multi-stream HMM based audio-visual speech recognition,” *Proc. Int. Conf. Spoken Lang. Process.*, Jeju Island, Korea, 2004.
  - 27) **G. Potamianos**, C. Neti, J. Huang, J.H. Connell, S. Chu, V. Libal, E. Marcheret, N. Haas, and J. Jiang, “Towards practical deployment of audio-visual speech recognition,” Invited, *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 3, pp. 777–780, Montreal, Canada, 2004.

- 28) J. Jiang, **G. Potamianos**, H. Nock, G. Iyengar, and C. Neti, "Improved face and feature finding for audio-visual speech recognition in visually challenging environments," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 5, pp. 873–876, Montreal, Canada, 2004.
- 29) S.M. Chu, V. Libal, E. Marcheret, C. Neti, and **G. Potamianos**, "Multistage information fusion for audio-visual speech recognition," *Proc. Int. Conf. Multimedia Expo.*, Taipei, Taiwan, 2004.
- 30) **G. Potamianos**, C. Neti, and S. Deligne, "Joint audio-visual speech processing for recognition and enhancement," *Proc. Work. Audio-Visual Speech Process.*, pp. 95–104, St. Jorioz, France, 2003.
- 31) J. Huang, **G. Potamianos**, and C. Neti, "Improving audio-visual speech recognition with an infrared headset," *Proc. Work. Audio-Visual Speech Process.*, pp. 175–178, St. Jorioz, France, 2003.
- 32) **G. Potamianos** and C. Neti, "Audio-visual speech recognition in challenging environments," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1293–1296, Geneva, Switzerland, 2003.
- 33) J.H. Connell, N. Haas, E. Marcheret, C. Neti, **G. Potamianos**, and S. Velipasalar, "A real-time prototype for small-vocabulary audio-visual ASR," *Proc. Int. Conf. Multimedia Expo.*, vol. II, pp. 469–472, Baltimore, MD, 2003.
- 34) U.V. Chaudhari, G.N. Ramaswamy, **G. Potamianos**, and C. Neti, "Information fusion and decision cascading for audio-visual speaker recognition based on time varying stream reliability prediction," *Proc. Int. Conf. Multimedia Expo.*, pp. 9–12, Baltimore, MD, 2003.
- 35) A. Garg, **G. Potamianos**, C. Neti, and T.S. Huang, "Frame-dependent multi-stream reliability indicators for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. I, pp. 24–27, Hong Kong, China, 2003.
- 36) U.V. Chaudhari, G.N. Ramaswamy, **G. Potamianos**, and C. Neti, "Audio-visual speaker recognition using time-varying stream reliability prediction," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. V, pp. 712–715, Hong Kong, China, 2003.
- 37) S. Deligne, **G. Potamianos**, and C. Neti, "Audio-visual speech enhancement with AVCDCN (audio-visual codebook dependent cepstral normalization)," *Proc. Int. Conf. Spoken Lang. Process.*, vol. 3, pp. 1449–1452, Denver, CO, 2002.
- 38) R. Goecke, **G. Potamianos**, and C. Neti, "Noisy audio feature enhancement using audio-visual speech data," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2025–2028, Orlando, FL, 2002.
- 39) G. Gravier, S. Axelrod, **G. Potamianos**, and C. Neti, "Maximum entropy and MCE based HMM stream weight estimation for audio-visual ASR," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 853–856, Orlando, FL, 2002.
- 40) G. Gravier, **G. Potamianos**, and C. Neti, "Asynchrony modeling for audio-visual speech recognition," *Proc. Human Lang. Techn. Conf.*, pp. 1–6, San Diego, CA, 2002.
- 41) **G. Potamianos**, C. Neti, G. Iyengar, and E. Helmuth, "Large-vocabulary audio-visual speech recognition by machines and humans," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1027–1030, Aalborg, Denmark, 2001.
- 42) **G. Potamianos** and C. Neti, "Automatic speechreading of impaired speech," *Proc. Work. Audio-Visual Speech Process.*, pp. 177–182, Aalborg, Denmark, 2001.
- 43) **G. Potamianos** and C. Neti, "Improved ROI and within frame discriminant features for lipreading," *Proc. Int. Conf. Image Process.*, pp. 250–253, Thessaloniki, Greece, 2001.
- 44) C. Neti, **G. Potamianos**, J. Luetin, I. Matthews, H. Glotin, and D. Vergyri, "Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop," *Proc. IEEE Work. Multimedia Signal Process.*, pp. 619–624, Cannes, France, 2001.
- 45) G. Iyengar, **G. Potamianos**, C. Neti, T. Faruquie, and A. Verma, "Robust detection of visual ROI for automatic speechreading," *Proc. IEEE Work. Multimedia Signal Process.*, pp. 79–84, Cannes, France, 2001.
- 46) I. Matthews, **G. Potamianos**, C. Neti, and J. Luetin, "A comparison of model and transform-based visual features for audio-visual LVCSR," *Proc. Intern. Conf. Multimedia Expo.*, Tokyo, Japan, 2001.
- 47) **G. Potamianos**, J. Luetin, and C. Neti, "Hierarchical discriminant features for audio-visual LVCSR," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 165–168, Salt Lake City, UT, 2001.
- 48) J. Luetin, **G. Potamianos**, and C. Neti, "Asynchronous stream modeling for large-vocabulary audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 169–172, Salt Lake City, UT, 2001.
- 49) H. Glotin, D. Vergyri, C. Neti, **G. Potamianos**, and J. Luetin, "Weighting schemes for audio-visual fusion in speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 173–176, Salt Lake City, UT, 2001.
- 50) **G. Potamianos** and C. Neti, "Stream confidence estimation for audio-visual speech recognition," *Proc. Intern. Conf. Spoken Language Process.*, pp. 746–749, Beijing, China, 2000.
- 51) C. Neti, G. Iyengar, **G. Potamianos**, A. Senior, and B. Maison, "Perceptual interfaces for information interaction: Joint processing of audio and visual information for human-computer interaction," *Proc. Intern. Conf. Spoken Language Process.*, vol III, pp. 11–14, Beijing, China, 2000.

- 52) **G. Potamianos**, A. Verma, C. Neti, G. Iyengar, and S. Basu, "A cascade image transform for speaker independent automatic speechreading," *Proc. IEEE Intern. Conf. Multimedia Expo.*, vol. II, pp. 1097–1100, New York, NY, 2000.
- 53) E. Cosatto, **G. Potamianos**, and H.P. Graf, "Audio-visual unit selection for the synthesis of photo-realistic talking-heads," *Proc. IEEE Intern. Conf. Multimedia Expo.*, vol. II, pp. 619–622, New York, NY, 2000.
- 54) **G. Potamianos** and A. Potamianos, "Speaker adaptation for audio-visual automatic speech recognition," *Proc. Europ. Speech Comm. Technol.*, Budapest, Hungary, vol. 3, pp. 1291–1294, 1999.
- 55) **G. Potamianos** and H.P. Graf, "Linear discriminant analysis for speechreading," *Proc. IEEE Work. Multimedia Signal Process.*, Los Angeles, CA, pp. 221–226, 1998.
- 56) **G. Potamianos**, H.P. Graf, and E. Cosatto, "An image transform approach for HMM based automatic lipreading," *Proc. Int. Conf. Image Process.*, Chicago, IL, pp. 173–177, 1998.
- 57) **G. Potamianos** and H.P. Graf, "Discriminative training of HMM stream exponents for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, vol. 6, pp. 3733–3736, 1998.
- 58) H.P. Graf, E. Cosatto, and **G. Potamianos**, "Machine vision of faces and facial features," *Proc. R.I.E.C. Int. Symp. Design Archit. Inform. Process. Systems Based Brain Inform. Princ.*, Sendai, Japan, pp. 48–53, 1998.
- 59) **G. Potamianos**, E. Cosatto, H.P. Graf, and D.B. Roe, "Speaker independent audio-visual database for bimodal ASR," *Proc. Europ. Tutorial Research Work. Audio-Visual Speech Process.*, Rhodes, Greece, pp. 65–68, 1997.
- 60) H.P. Graf, E. Cosatto, and **G. Potamianos**, "Robust recognition of faces and facial features with a multi-modal system," *Proc. Int. Conf. Systems Man Cybern.*, Orlando, FL, pp. 2034–2039, 1997.
- 61) **G. Potamianos**, "Efficient Monte Carlo estimation of partition function ratios of Markov random field images," *Proc. Conf. Inform. Sci. Systems*, Princeton, NJ, vol. II, pp. 1212–1215, 1996.
- 62) **G. Potamianos** and J. Goutsias, "A unified approach to Monte Carlo likelihood estimation of Gibbs random field images," *Proc. Conf. Inform. Sci. Systems*, Princeton, NJ, vol. I, pp. 84–90, 1994.
- 63) **G. Potamianos** and J. Goutsias, "An analysis of Monte Carlo methods for likelihood estimation of Gibbsian images," *Proc. Int. Conf. Acoust. Speech Signal Process.*, Minneapolis, MN, vol. V, pp. 519–522, 1993.
- 64) **G. Potamianos** and J. Goutsias, "On computing the likelihood function of partially observed Markov random field images using Monte Carlo simulations," *Proc. Conf. Inform. Sci. Systems*, Princeton, vol. I, pp. 357–362, 1992.
- 65) **G. Potamianos** and J. Goutsias, "A novel method for computing the partition function of Markov random field images using Monte Carlo simulations," *Proc. Int. Conf. Acoust. Speech Signal Process.*, Toronto, Canada, vol. 4, pp. 2325–2328, 1991.
- 66) **G. Potamianos** and J. Diamessis, "Frequency sampling design of 2-D IIR filters using continued fractions," *Proc. Int. Symp. Circuits Systems*, New Orleans, LA, pp. 2454–2457, 1990.
- 67) J. Diamessis and **G. Potamianos**, "A novel method for designing IIR filters with nonuniform samples," *Proc. Conf. Inform. Sci. Systems*, Princeton, NJ, vol. 1, pp. 192–195, 1990.
- 68) J. Diamessis and **G. Potamianos**, "Modeling unequally spaced 2-D discrete signals by rational functions," *Proc. Int. Symp. Circuits Systems*, Portland, OR, pp. 1508–1511, 1989.

## Thesis

**G. Potamianos**, *Stochastic Simulation Algorithms for Partition Function Estimation of Markov Random Field Images*, The Johns Hopkins University, 1994.

## Technical Reports

- 1) C. Neti, **G. Potamianos**, J. Luettin, I. Matthews, H. Glotin, D. Vergyri, J. Sison, A. Mashari, and J. Zhou, "Audio-Visual Speech Recognition: Final Workshop 2000 Report", *Technical Report*, Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD, 2000.
- 2) **G. Potamianos** and F. Jelinek, "A study of n-gram and decision tree letter language modeling methods," *Research Notes, Center for Language and Speech Processing Technical Report*, no. 13, The Johns Hopkins University, Baltimore, MD, 1997.
- 3) **G. Potamianos** and J. Goutsias, "Stochastic Approximation Algorithms for Partition Function Estimation of Gibbs Random Fields," *Technical Report JHU/ECE*, no. 95-27, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, 1995.
- 4) **G. Potamianos** and A. Gatherer, "Coaxial cable system channel impairments and their effects to the transmission of QAM modulated signals," *Technical Activity Report*, Integrated Systems Laboratory, Texas Instruments, Dallas, 1994.
- 5) **G. Potamianos**, A. Gatherer, and G. Wyman, "Coaxial cable channel impulse response characterization," *Technical Activity Report*, Integrated Systems Laboratory, Texas Instruments, Dallas, 1994.
- 6) **G. Potamianos** and J. Goutsias, "Analysis of stochastic simulation algorithms for partition function estimation of Gibbs images," *Technical Report JHU/ECE*, no. 93-05, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, 1993.
- 7) **G. Potamianos** and J. Goutsias, "Stochastic simulation techniques for partition function approximation of Gibbs random field images," *Technical Report JHU/ECE*, no. 91-02, Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, 1991.

## Tutorials / Short Courses

- 1) A. Katsaggelos and **G. Potamianos**, *Joint Audio-Visual Signal Processing*, Tutorial, Int. Conf. Image Process., Barcelona, Spain, 2003.
- 2) C. Neti, **G. Potamianos**, and G. Iyengar, *Joint Audio-Visual Processing and Resources*, Elsnet Summer School, Prague, Czech Republic, 2001.

## U.S. Patents (Granted)

- 1) J.H. Connell, N. Haas, E. Marcheret, C.V. Neti, and **G. Potamianos**, *Audio-Only Backoff in Audio-Visual Speech Recognition System*, Patent No.: US007251603B2, July 31, 2007.
- 2) U.V. Chaudhari, C. Neti, **G. Potamianos**, and G.N. Ramaswamy, *Automated Decision Making Using Time-Varying Stream Reliability Prediction*, Patent No.: US007228279B2, June 5, 2007.
- 3) P. de Cuetos, G.R. Iyengar, C.V. Neti, and **G. Potamianos**, *System and Method for Microphone Activation Using Visual Speech Cues*, Patent No.: US006754373B1, June 22, 2004.
- 4) E. Cosatto, H.P. Graf, **G. Potamianos**, and J. Schroeter, *Audio-Visual Selection Process for the Synthesis of Photo-Realistic Talking-Head Animations*, Patent No.: US006654018B1, Nov. 25, 2003.
- 5) E. Cosatto, H.P. Graf, and **G. Potamianos**, *Robust multi-modal method for recognizing objects*, Patent No.: US006118887A, Sep. 12, 2000.

## Sample Citations of Published Work (402 listed; list is slightly outdated – Jan. 2007)

NOTE: Most commonly cited papers (with five or more citations) are listed (excluding self author citations).

- **51 citations of:** G. Potamianos, C. Neti, G. Gravier, A. Garg, and A.W. Senior, “Recent advances in the automatic recognition of audio-visual speech,” Invited, *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1306–1326, 2003.
  - 1) S. Oviatt, “User-centered modeling and evaluation of multimodal interfaces,” *Proceedings of the IEEE*, 91(9): 1457–1468, 2003.
  - 2) P.S. Aleksic and A.K. Katsaggelos, “Speech-to-video synthesis using MPEG-4 compliant visual features,” *IEEE Trans. Circuits Syst. Video Technol.*, 14(5): 682–692, 2004.
  - 3) S. Oviatt, T. Darrell, and M. Flickner, “Introduction. Special Issue on Multimodal Interfaces that Flex Adapt and Persist,” *Communications of the ACM*, 47(1): 30–75, 2004.
  - 4) P. Aarabi and B.V. Dasarathy, “Robust speech processing using multi-sensor multi-source information fusion – an overview of the state of the art,” *Information Fusion*, 5(2): 77–80, 2004.
  - 5) M.J. Sánchez Martínez and J.P. de la Cruz Gutiérrez, “Audio-visual speech recognition using motion based lipreading,” *Proc. Int. Conf. Spoken Language Process.*, 2004.
  - 6) P. Lucey, T. Martin, and S. Sridharan, “Confusability of phonemes grouped according to their viseme classes in noisy environments,” *Proc. Australian Int. Conf. on Speech Science and Technology*, pp. 265–270, 2004.
  - 7) H.E. Cetingul, Y. Yemez, E. Erzlin, and A.M. Tekalp, “Discriminative lip-motion features for biometric speaker identification,” *Proc. Int. Conf. Image Process.*, vol. 3, pp. 2023–2026, 2004.
  - 8) M. Liu, Z. Xiong, S.M. Chu, Z. Zhang, and T.S. Huang, “Audio-visual word spotting,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 3, pp. 785–788, 2004.
  - 9) P.S. Aleksic and A.K. Katsaggelos, “Comparison of low- and high-level visual features for audio-visual continuous automatic speech recognition,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 5, pp. 917–920, 2004.
  - 10) Z. Wu and P.S. Aleksic, “Inner lip feature extraction for MPEG-4 facial animation,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 3, pp. 633–636, 2004.
  - 11) A. Ortega, F. Sukno, E. Lleida, A. Frangi, A. Miguel, L. Buera, and E. Zacur, “AV@CAR: A Spanish multichannel multimodal corpus for in-vehicle automatic audio-visual speech recognition,” *Proc. Int. Conf. on Language Resources and Evaluation*, 2004.
  - 12) K. Saenko, T. Darrell, and J. Glass, “Articulatory features for robust visual speech recognition,” *Proc. Int. Conf. Multimodal Interfaces*, pp. 152–158, 2004.
  - 13) D.W. Massaro, “A framework for evaluating multimodal integration by humans and a role for embodied conversational agents,” *Proc. 6th Int. Conf. on Multimodal Interfaces*, pp. 24–31, 2004.
  - 14) P. Yin, I. Essa, and J.M. Rehg, “Asymmetrically boosted HMM for speech reading,” *Proc. Conf. Computer Vision and Pattern Recog.*, vol. 2, pp. 755–761, 2004.
  - 15) M. Turk, “Multimodal human computer interaction,” In *Real-Time Vision for Human-Computer Interaction*, B. Kisanin, V. Pavlovic, and T. Huang (eds.), Springer, Aug. 2005.
  - 16) K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, and T. Darrell, “Visual speech recognition with loosely synchronized feature streams,” *Proc. Int. Conf. Comp. Vision*, vol. 2, pp. 1424–1431, 2005.
  - 17) R. Goecke, “3D lip tracking and co-inertia analysis for improved robustness of audio-video automatic speech recognition,” *Proc. Int. Conf. Auditory-Visual Speech Process.*, pp. 109–114, 2005.
  - 18) R. Goecke, “Current trends in joint audio-visual signal processing: a review,” *Proc. Int. Symp. Signal Process. and its Applications*, pp. 70–73, 2005.
  - 19) R. Goecke, “Audio-video automatic speech recognition: An example of improved performance through multimodal sensor input,” *Proc. 2005 NICTA-HCSNet Multimodal User Interaction Works.*, pp. 25–32, 2005.
  - 20) H.E. Cetingul, Y. Yemez, E. Erzlin, and A.M. Tekalp, “Robust lip-motion features for speaker identification,” *Proc. Int. Conf. Image Process.*, vol. 1, pp. 509–512, 2005.
  - 21) J.-S. Lee and C.H. Park, “Discriminative training of hidden Markov models by multiobjective optimization for visual speech recognition,” *Proc. Int. Conf. Neural Networks*, vol. 4, pp. 2053–2058, 2005.
  - 22) I. Arsic, N. Marina, and J.-P. Thiran, “Impact of sample sizes on information theoretic measures for audio-visual signal processing,” *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2005.
  - 23) J.F.G. Pérez, A.F. Frangi, E.L. Solano, and K. Lukas, “Lip reading for robust speech recognition on embedded devices,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 473–476, 2005.
  - 24) T.W. Lewis and D.M.W. Powers, “Distinctive feature fusion for improved audio-visual phoneme recognition,” *Proc. Int. Symp. Signal Process. and its Applications*, pp. 62–65, 2005.
  - 25) P.S. Aleksic and A.K. Katsaggelos, “Comparison of MPEG-4 facial animation parameter groups with respect to audio-visual speech recognition performance,” *Proc. Int. Conf. Image Process.*, 2005.
  - 26) N.A. Fox, B.A. O’Mullane, and R.B. Reilly, “VALID: A new practical audio-visual database, and comparative results,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 3546, pp. 777–786, 2005.
  - 27) N.A. Fox, B.A. O’Mullane, and R.B. Reilly, “Audio-visual speaker identification via adaptive fusion using reliability estimates of both modalities,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 3546, pp. 787–796, 2005.
  - 28) L.J.M. Rothkrantz, J.C. Wojdel, and P. Wiggers, “Fusing data streams in continuous audio-visual speech recognition,” *Proc. Int. Conf. Text, Speech and Dialogue*, J.G. Carbonell and J. Siekmann (eds.), Springer LNAI 3658, pp. 33–44, 2005.
  - 29) X. Li and C. Kwan, “Geometrical feature extraction for robust speech recognition,” *Proc. Asilomar Conf. Signals, Systems, and Computers*, pp. 558–562, 2005.
  - 30) J. Huang and K. Visweswariah, “Improving lip-reading with feature space transforms for multi-stream audio-visual speech recognition,” *Proc. Interspeech*, pp. 1221–1224, 2005.
  - 31) J. Huang and D. Povey, “Discriminatively trained features using fMPE for multi-stream audio-visual speech recognition,” *Proc. Interspeech*, pp. 777–780, 2005.
  - 32) J. Huang, E. Marcheret, and K. Visweswariah, “Rapid feature space speaker adaptation for multi-stream HMM-based audio-visual speech recognition,” *Proc. Int. Conf. Multimedia Expo*, pp. 338–341, 2005.
  - 33) M.E. Sargin, E. Erzlin, Y. Yemez, and A.M. Tekalp, “Lip feature extraction based on audio-visual correlation,” *Proc. Eusipco*, 2005.
  - 34) D. Mou, *Autonomous Face Recognition*, Ph.D. Thesis, University of Ulm, Germany, 2005.

- 35) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 36) N.A. Fox, *Audio and Video Based Person Identification*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 37) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
  - 38) M.-L. Bourguet, "Towards a taxonomy of error-handling strategies in recognition-based multimodal human-computer interfaces," *Signal Process. J.*, 86(12): 3625–3643, 2006.
  - 39) D. Nguyen, D. Halupka, P. Aarabi, and A. Sheikholeslami, "Real-time face detection and lip feature extraction using field-programmable gate arrays," *IEEE Trans. Systems, Man and Cybernetics*, 36(4): 902–912, 2006.
  - 40) H. Bredin, A. Miguel, I.H. Witten, and G. Chollet, "Detecting replay attacks in audiovisual identity verification," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 621–624, 2006.
  - 41) A. Potamianos, G. Bouselmi, D. Dimitriadis, D. Fohr, R. Gemello, I. Illina, F. Maha, P. Maragos, M. Matassoni, V. Pitsikalis, J. Ramirez, E. Sanchez-Soto, J. Segura, and P. Svaizer, "Towards speaker and environmental robustness in ASR: The HIWIRE project," *Proc. Work. Speech Recognition and Intrinsic Variation (SRIV)*, pp. 135–142, 2006.
  - 42) D. Sodoyer, B. Rivet, L. Girin, J.-L. Schwartz, and C. Jutten, "An analysis of visual speech information applied to voice activity detection," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 601–604, 2006.
  - 43) W.C. Yau, D.K. Kumar, and S.P. Arjunan, "Visual speech recognition method using translation, scale, and rotation invariant features," *Proc. Int. Conf. Video Signal Based Surveillance*, p. 63, 2006.
  - 44) I. Lee Hetherington, H. Shu, and J.R. Glass, "Flexible multi-stream framework for speech recognition using multi-tape finite state transducers," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 417–420, 2006.
  - 45) X. Hong, H. Yao, Y. Wan, and R. Chen, "A PCA based visual DCT feature extraction method for lip-reading," *Proc. Int. Conf. Intelligent Information Hiding and Multimedia*, pp. 321–326, 2006.
  - 46) M.I. Faraj and J. Bigun, "Person verification by lip-motion," *Proc. Conf. Computer Vision Pattern Recog. Work.*, pp. 37, 2006.
  - 47) M.E. Sargin, E. Erzincan, Y. Yemez, and A.M. Tekalp, "Multimodal speaker identification using canonical correlation analysis," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 613–616, 2006.
  - 48) W.C. Yau, D.K. Kumar, S.P. Arjunan, and S. Kumar, "Visual speech recognition method using image moments and multiresolution wavelet images," *Proc. Int. Conf. Computer Graphics, Imaging and Visualisation*, pp. 194–199, 2006.
  - 49) A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, "Multimodal fusion by adaptive compensation for feature uncertainty with application to audiovisual speech recognition," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
  - 50) G. Monaci, P. Jost, P. Vanderghyest, B. Mailhe, S. Lesage, and R. Gribonval, "Learning multi-modal dictionaries," Submitted To: *IEEE Trans. Image Process.* (2006).
  - 51) N.A. Fox, R. Gross, J.F. Cohn, and R.B. Reilly, "Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts," To Appear: *IEEE Trans. Multimedia*, 2007.
- **49 citations of:** C. Neti, Potamianos, J. Luetttin, I. Matthews, H. Glotin, D. Vergyri, J. Sison, A. Mashari, and J. Zhou, "Audio-Visual Speech Recognition: Final Workshop 2000 Report", *Technical Report*, Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD, 2000.
    - 1) P. Daubias and P. Deléglise, "Evaluation of an automatically obtained shape and appearance model for automatic audio visual speech recognition," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1031–1034, 2001.
    - 2) M. Slaney, D. Ponceleon, and J. Kaufman, "Temporal events in all dimensions and scales," *Proc. IEEE Work. on Detection and Recognition of Events in Video*, pp. 83–91, 2001.
    - 3) M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1260–1273, 2002.
    - 4) X. Zhang, C.C. Broun, R.M. Mersereau, and M. Clements, "Automatic speechreading with applications to human-computer interfaces," *EURASIP J. Appl. Signal Processing*, 2002(11): 1228–1247, 2002.
    - 5) P. Daubias and P. Deléglise, "Statistical lip-appearance models trained automatically using audio information," *EURASIP J. Appl. Signal Processing*, 2002(11), 1202–1212, 2002.
    - 6) A.V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, "A coupled HMM for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2013–2016, 2002.
    - 7) C.C. Chibelushi, F. Deravi, and J.S.D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. Multimedia*, 4(1): 23–37, 2002.
    - 8) P.S. Aleksic, J.J. Williams, Z. Wu, and A.K. Katsaggelos, "Audio-visual speech recognition using MPEG-4 compliant visual features," *EURASIP J. Appl. Signal Process.*, 2002(11): 1213–1227, 2002.
    - 9) S.M. Chu and T.S. Huang, "Audio-visual speech modeling using coupled hidden Markov models," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2009–2012, 2002.
    - 10) L. Liang, X. Liu, Y. Zhao, X. Pi, and A.V. Nefian, "Speaker independent audio-visual continuous speech recognition," *Proc. Int. Conf. Multimedia and Expo*, 2002.
    - 11) X. Liu, Y. Zhao, X. Pi, L. Liang, and A.V. Nefian, "Audio-visual continuous speech recognition using a coupled hidden Markov model," *Proc. Int. Conf. Spoken Lang. Process.*, pp. 213–216, 2002.
    - 12) X. Zhang, R.M. Mersereau, and M. Clements, "Bimodal fusion in audio-visual speech recognition," *Proc. Int. Conf. Image Process.*, vol. 1, pp. 964–967, 2002.
    - 13) P.K. Kakumanu, *Audio-visual processing for speech-driven facial animation*, M.Sc. Thesis, Wright State University, 2002.
    - 14) G. Pingali, C. Pinhanez, A. Levas, R. Kjeldsen, M. Podlaseck, H. Chen, and N. Sukaviriya, "Steerable interfaces for pervasive computing spaces," *Proc. IEEE Conf. Pervasive Computing and Communications*, 2003.
    - 15) P.S. Aleksic and A.K. Katsaggelos, "Product HMMs for audio-visual continuous speech recognition using facial animation parameters," *Proc. Int. Conf. Multimedia and Expo*, vol. 1, pp. 481–484, 2003.
    - 16) V. Pera, F. Sá, P. Afonso, and R. Ferreira, "Audio-visual speech recognition in a Portuguese language based application," *Proc. Int. Conf. Industrial Technology*, vol. 2, pp. 688–692, 2003.
    - 17) Y. Zhang, Q. Diao, S. Huang, W. Hu, C. Bartels, and J. Bilmes, "DBN-based multi-stream models for speech," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 884–887, 2003.
    - 18) T. Yoshinaga, S. Tamura, K. Iwano, and S. Furui, "Audio-visual speech recognition using lip motion movement extracted from side-face images," *Proc. Work. Audio-Visual Speech Process.*, pp. 117–120, 2003.

- 19) M.-W. Mak, M.-C. Cheung, and S.-Y. Kung, "Robust speaker verification from GSM-transcoded speech based on decision fusion and feature transformation," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 2, pp. 745–748, 2003.
- 20) P. Daubias and P. Deléglise, "The LIUM-AVS database: a corpus to test lip segmentation and speechreading systems in natural conditions," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1569–1571, 2003.
- 21) A.V. Nefian, L.H. Liang, T. Fu, and X.X. Liu, "A Bayesian approach to audio-visual speaker identification," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 761–769, 2003.
- 22) X.X. Liu, L.H. Liang, X. Pi, and A.V. Nefian, "Audio-visual speaker identification using coupled hidden Markov models," *Proc. Int. Conf. Image Process.*, 2003.
- 23) Y.K. Tan, N. Sherkat, and T. Allen, "Eye gaze and speech for data entry: a comparison of different data entry methods," *Proc. Int. Conf. Multimedia Expo*, vol. 1, pp. 41–44, 2003.
- 24) E.K. Patterson and J.N. Gowdy, "An audio-visual approach to simultaneous-speaker speech recognition," *Proc. Int. Conf. Acoustics Speech Signal Process.*, vol. 5, pp. 780–783, 2003.
- 25) P.S. Aleksic and A.K. Katsaggelos, "Speech-to-video synthesis using MPEG-4 compliant visual features," *IEEE Trans. Circuits Syst. Video Technol.*, 14(5): 682–692, 2004.
- 26) G.F. Meyer, J.B. Mulligan, and S.M. Wuerger, "Continuous audio-visual digit recognition using N-best decision fusion," *Information Fusion*, 5(2): 91–101, 2004.
- 27) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, The Australian National University, Jan. 2004.
- 28) R. Goecke and J.B. Millar, "The audio-video Australian English speech data corpus (AVOZES)," *Proc. Int. Conf. Spoken Lang. Process.*, 2004.
- 29) J.B. Millar, M. Wagner, and R. Goecke, "Aspects of speaking-face data corpus design methodology," *Proc. Int. Conf. Spoken Lang. Process.*, 2004.
- 30) R. Goecke and J.B. Millar, "A detailed description of the AVOZES data corpus," *Proc. Australian Int. Conf. on Speech Science and Technology*, pp. 486–489, 2004.
- 31) T. Yoshinaga, S. Tamura, K. Iwano, and S. Furui, "Audio-visual speech recognition using new lip features extracted from side-face images," *Proc. ISCA Tutorial and Works. on Robustness Issues in Conversational Interaction (ROBUST)*, 2004.
- 32) J.N. Gowdy, A. Subramanya, C. Bartels, and J. Bilmes, "DBN based multi-stream models for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 993–996, 2004.
- 33) T.J. Hazen, K. Saenko, C.-H. La, and J.R. Glass, "A segment-based audio-visual speech recognizer: Data collection, development, and initial experiments," *Proc. Int. Conf. Multimodal Interfaces*, pp. 235–242, 2004.
- 34) X. Zhang, K. Takeda, J.H.L. Hansen, and T. Maeno, "Audio-visual integration for hands-free voice interaction in automobile route navigation," *Proc. Int. Congress on Acoustics*, pp. 2821–2824, 2004.
- 35) M.J. Sánchez Martínez and J.P. de la Cruz Gutiérrez, "Audio-visual speech recognition using motion based lipreading," *Proc. Int. Conf. Spoken Language Process.*, 2004.
- 36) P.S. Aleksic and A.K. Katsaggelos, "Comparison of low- and high-level visual features for audio-visual continuous automatic speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 5, pp. 917–920, 2004.
- 37) M. Liu, Z. Xiong, S.M. Chu, Z. Zhang, and T.S. Huang, "Audio-visual word spotting," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 3, pp. 785–788, 2004.
- 38) P. Liu and Z. Whang, "Stream weight training based on MCE for audio-visual LVCSR," *Tsinghua Science and Technology*, 10(2): 141–144, 2005.
- 39) O. Balter, O. Engwall, A.-M. Oster, and H. Kjellstrom, "Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction," *Proc. Int. ACM Conf. on Computers and Accessibility*, pp. 36–43, 2005.
- 40) R. Goecke, "Current trends in joint audio-visual signal processing: a review," *Proc. Int. Symp. Signal Process. and its Applications*, pp. 70–73, 2005.
- 41) P.S. Aleksic and A.K. Katsaggelos, "Comparison of MPEG-4 facial animation parameter groups with respect to audio-visual speech recognition performance," *Proc. Int. Conf. Image Process.*, 2005.
- 42) K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, and T. Darrell, "Visual speech recognition with loosely synchronized feature streams," *Proc. Int. Conf. Comp. Vision*, vol. 2, pp. 1424–1431, 2005.
- 43) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
- 44) P.S. Aleksic and A.K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Trans. on Information Forensics and Security*, 1(1): 3–11, 2006.
- 45) O. Engwall, O. Balter, A.-M. Oster, and H. Kjellstrom, "Designing the user interface of the computer-based speech training system ARTUR based on early user tests," *Behaviour and Information Technology*, 25(4): 353–365, 2006.
- 46) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
- 47) J. Kaukenas, G. Navickas, and L. Telksnys, "Human-computer audiovisual interface," *Information Technology and Control*, 35(2): 87–92, 2006.
- 48) X. Hong, H. Yao, Y. Wan, and R. Chen, "A PCA based visual DCT feature extraction method for lip-reading," *Proc. Int. Conf. Intelligent Information Hiding and Multimedia*, pp. 321–326, 2006.
- 49) S. Sakti, K. Markov, and S. Nakamura, "Incorporation of pentaphone-context dependency based on hybrid HMM/BN acoustic modeling framework," *Proc. Int. Conf. Acoustics Speech Signal Process.*, vol. 1, pp. 1177–1180, 2006.
- **35 citations of:** G. Potamianos, H.P. Graf, and E. Cosatto, "An image transform approach for HMM based automatic lipreading," *Proc. Int. Conf. Image Process.*, Chicago, IL, pp. 173–177, 1998.
  - 1) A. Rosenfeld, "Image analysis and computer vision: 1998," *Comp. Vision and Image Understanding*, 74(1): pp. 36–95, 1999.
  - 2) S. Dupont and J. Luetttin, "Audio-visual speech modeling for continuous speech recognition," *IEEE Trans. Multimedia*, 2(3): 141–151, 2000.
  - 3) S. Gurbuz, Z. Tufekci, E. Patterson, and J.N. Gowdy, "Application of affine-invariant Fourier descriptors to lipreading for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, 2001.
  - 4) S. Gurbuz, E.K. Patterson, Z. Tufekci, and J.N. Gowdy, "Lip-reading from parametric lip contours for audio-visual speech recognition," *Proc. Eurospeech*, pp. 1181–1184, 2001.
  - 5) P. Daubias and P. Deléglise, "Statistical lip-appearance models trained automatically using audio information," *EURASIP J. Appl.*

*Signal Processing*, 2002(11), 1202–1212, 2002.

- 6) P. Daubias and P. Deléglise, “Lip-reading based on a fully automatic statistical model,” *Proc. Int. Conf. Spoken Lang. Process.*, 2002.
  - 7) E.K. Patterson, S. Gurbuz, Z. Tufekci, and J.N. Gowdy, “Moving talker, speaker-independent feature study, and baseline results using the CUAVE multimodal speech corpus,” *EURASIP J. Appl. Signal Process.*, 2002(11): 1189–1201, 2002.
  - 8) M. Heckmann, K. Kroschel, C. Savariaux, and F. Berthommier, “DCT-based video features for audio-visual speech recognition,” *Proc. Int. Conf. Spoken Lang. Process.*, pp. 1925–1928, 2002.
  - 9) G. Meyer and J. Mulligan, “Continuous audio-visual digit recognition using decision fusion,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 305–308, 2002.
  - 10) E. Patterson, S. Gurbuz, Z. Tufekci, and J.N. Gowdy, “CUAVE: A new audio-visual database for multimodal human-computer interface research,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2017–2020, 2002.
  - 11) S. Lucey, S. Sridharan, and C. Chandran, “Improved facial-feature detection for AVSP via unsupervised clustering and discriminant analysis,” *EURASIP J. Applied Signal Process.*, 2003(3): 264–275, 2003.
  - 12) S. Lucey, “An evaluation of visual speech features for the tasks of speech and speaker recognition,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 2688, pp. 260–267, 2003.
  - 13) P. Daubias and P. Deléglise, “The LIUM-AVS database: a corpus to test lip segmentation and speechreading systems in natural conditions,” *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1569–1571, 2003.
  - 14) P. Scanlon, R. Reilly, and P. de Chazal, “Visual feature analysis for automatic speechreading,” *Proc. Work. Audio-Visual Speech Process.*, pp. 127–132, 2003.
  - 15) N. Fox and R.B. Reilly, “Audio-visual speaker identification based on the use of dynamic audio and visual features,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 743–751, 2003.
  - 16) M. Heckmann, F. Berthommier, C. Savariaux, and K. Kroschel, “Effects of image distortions on audio-visual speech recognition,” *Proc. Work. Audio-Visual Speech Process.*, pp. 163–168, 2003.
  - 17) V. Pera, F. Sá, P. Afonso, and R. Ferreira, “Audio-visual speech recognition in a Portuguese language based application,” *Proc. Int. Conf. Industrial Technology*, vol. 2, pp. 688–692, 2003.
  - 18) W. Yu, *Bimodal Voice Recognition Based Computer Input*, Master Thesis, Dept. of Human Work Sciences, Division of Industrial Ergonomics, Luleå University of Technology, Sweden, 2003
  - 19) G.F. Meyer, J.B. Mulligan, and S.M. Wuerger, “Continuous audio-visual digit recognition using N-best decision fusion,” *Information Fusion*, 5: 91–101, 2004.
  - 20) P.S. Aleksic and A.K. Katsaggelos, “Comparison of low- and high-level visual features for audio-visual continuous automatic speech recognition,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 5, pp. 917–920, 2004.
  - 21) N.A. Fox, R. Gross, J.F. Cohn, and R.B. Reilly, “Robust automatic human identification using face, mouth, and acoustic information,” *Proc. Works. Analysis and Modelling of Faces and Gestures (AMFG)*, Springer LNCS 3723, pp. 264–278, 2005.
  - 22) N.A. Fox, B.A. O’Mullane, and R.B. Reilly, “VALID: A new practical audio-visual database, and comparative results,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 3546, pp. 777–786, 2005.
  - 23) N.A. Fox, B.A. O’Mullane, and R.B. Reilly, “Audio-visual speaker identification via adaptive fusion using reliability estimates of both modalities,” *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 3546, pp. 787–796, 2005.
  - 24) N.A. Fox, *Audio and Video Based Person Identification*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 25) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 26) P. Lucey, S. Lucey, and S. Sridharan, “Using a free-parts representation for visual speech recognition,” *Proc. Digital Image Computing: Techniques and Applications (DICTA)*, p. 55, 2005.
  - 27) P. Lucey, D. Dean, and S. Sridharan, “Problems associated with current area-based visual speech feature extraction techniques,” *Proc. Int. Conf. Auditory-Visual Speech Process.*, pp. 73–78, 2005.
  - 28) S. Lucey and P. Lucey, “Improved speech reading through a free-parts representation,” *Proc. Int. Conf. Auditory-Visual Speech Process.*, pp. 85–86, 2005.
  - 29) A. Sagheer, N. Tsuruta, R.-I. Taniguchi, and S. Maeda, “Visual speech features representation for automatic lip-reading,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 2, pp. 781–784, 2005.
  - 30) P.S. Aleksic and A.K. Katsaggelos, “Comparison of MPEG-4 facial animation parameter groups with respect to audio-visual speech recognition performance,” *Proc. Int. Conf. Image Process.*, 2005.
  - 31) D. Nguyen, D. Halupka, P. Aarabi, and A. Sheikholeslami, “Real-time face detection and lip feature extraction using field-programmable gate arrays,” *IEEE Trans. Systems, Man and Cybernetics*, 36(4): 902–912, 2006.
  - 32) X. Hong, H. Yao, Y. Wan, and R. Chen, “A PCA based visual DCT feature extraction method for lip-reading,” *Proc. Int. Conf. Intelligent Information Hiding and Multimedia*, pp. 321–326, 2006.
  - 33) I. Arsic and J.-P. Thiran, “Mutual information eigenlips for audio-visual speech recognition,” *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
  - 34) F. Shafait, R. Kricke, I. Shdaifat, and R.-R. Grigat, “Real time lip motion analysis for a person authentication system using near infrared illumination,” *Proc. Int. Conf. Image Process.*, pp. 1957–1960, 2006.
  - 35) N.A. Fox, R. Gross, J.F. Cohn, and R.B. Reilly, “Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts,” To Appear: *IEEE Trans. Multimedia*, 2007.
- **34 citations of:** G. Potamianos and H.P. Graf, “Discriminative training of HMM stream exponents for audio-visual speech recognition,” *Proc. Int. Conf. Acoust. Speech Signal Process.*, Seattle, WA, vol. 6, pp. 3733–3736, 1998.
- 1) M.T. Chan, Y. Zhang, and T.S. Huang, “Real-time lip tracking and bimodal continuous speech recognition,” *Proc. Works. Multimedia Signal Process.*, pp. 65–70, 1998.
  - 2) J. Huang, Z. Liu, Y. Wang, Y. Chen, and E.K. Wong, “Integration of multimodal features for video scene classification based on HMM,” *Proc. IEEE Work. Multimedia Signal Process.*, pp. 53–58, 1999.
  - 3) S. Basu, C. Neti, N. Rajput, A. Senior, L. Subramaniam, and A. Verma, “Audio-visual large vocabulary continuous speech recognition in the broadcast domain,” *Proc. IEEE Work. Multimedia Signal Process.*, pp. 475–481, 1999.
  - 4) A. Verma, T. Faruque, C. Neti, S. Basu, A. Senior, “Late integration in audio-visual speech recognition,” *Proc. Work. Automatic Speech Recognition Understanding*, 1999.
  - 5) P.M. McCourt, S.V. Vaseghi, and B. Doherty, “Multi-resolution sub-band features and models for HMM-based phonetic modelling,”

*Technical Report CSL023/99*, School of Electrical and Electronic Engineering, Queens Univ. Belfast, 1999.

- 6) Y. Wang, Z. Liu, and J.-C. Huang, "Multimedia content analysis using both audio and visual cues," *Signal Process. Mag.*, 17(6): 12–36, 2000.
  - 7) S. Dupont and J. Luetttin, "Audio-visual speech modeling for continuous speech recognition," *IEEE Trans. Multimedia*, 2(3): 141–151, 2000.
  - 8) S. Nakamura, H. Ito, and K. Shikano, "Stream weight optimization of speech and lip image sequence for audio-visual speech recognition," *Proc. Int. Conf. Spoken Lang. Process.*, 2000.
  - 9) C. Miyajima, K. Tokuda, and T. Kitamura, "Audiovisual speech recognition using MCE-based HMMs and model-dependent stream weights," *Proc. Int. Conf. Spoken Lang. Process.*, 2000.
  - 10) S.M. Chu and T.S. Huang, "Bimodal speech recognition using coupled hidden Markov models," *Proc. Int. Conf. Spoken Lang. Process.*, 2000.
  - 11) T.A. Faruque, A. Majumdar, N. Rajput, and L.V. Subramaniam, "Large vocabulary audio-visual speech recognition using active shape models," *Proc. Int. Conf. Pattern Recog.*, vol. 3, pp. 106–109, 2000.
  - 12) K. Kirchhoff and J. Bilmes, "Combination and joint training of acoustic classifiers for speech recognition," *Proc. ISCA ASR 2000 Tutorial and Research Works.*, 2000.
  - 13) T. Wark and S. Sridharan, "Adaptive fusion of speech and lip information for robust speaker identification," *Digital Signal Process.*, 11: 169–186, 2001.
  - 14) K. Kumatani, S. Nakamura, and K. Shikano, "An adaptive integration based on product HMM for audio-visual speech recognition," *Proc. Int. Conf. Multimedia and Expo*, 2001.
  - 15) S. Nakamura, "Fusion of audio-visual information for integrated speech processing," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 127–143, 2001.
  - 16) S. Basu, H.S.M. Beigi, S.H. Maes, B.E.G. Maison, C.V. Neti, and A.W. Senior, *Methods and Apparatus for Audio-Visual Speaker Recognition and Utterance Verification*, U.S. Patent 6,219,640 B1, Apr. 2001.
  - 17) C.C. Chibelushi, F. Deravi, and J.S.D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. Multimedia*, 4(1): 23–37, 2002.
  - 18) K. Kirchhoff, G.A. Fink, and G. Sagerer, "Combining acoustic and articulatory information for robust speech recognition," *Speech Communication*, 37(3–4): 303–319, 2002.
  - 19) P.S. Aleksic, J.J. Williams, Z. Wu, and A.K. Katsaggelos, "Audio-visual speech recognition using MPEG-4 compliant visual features," *EURASIP J. Appl. Signal Process.*, 2002(11): 1213–1227, 2002.
  - 20) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
  - 21) S. Nakamura, K. Kumatani, and S. Tamura, "Robust bi-modal speech recognition based on state synchronous modeling and stream weight optimization," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 309–312, 2002.
  - 22) S.M. Chu and T.S. Huang, "Audio-visual speech modeling using coupled hidden Markov models," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2009–2012, 2002.
  - 23) S. Lucey, *Audio-Visual Speech Processing*, Ph.D. Thesis, Queensland Univ. of Technology, 2002.
  - 24) A. Verma, S. Basu, and C. Neti, *Late Integration in Audio-Visual Continuous Speech Recognition*, U.S. Patent 6,633,844 B1, Oct. 2003.
  - 25) A. Xafopoulos, C. Kotropoulos, G. Almpanidis, and I. Pitas, "Language identification in web documents using discrete HMMs," *Pattern Recog.*, 37(3): 583–594, 2004.
  - 26) J.C. Nascimento and J.S. Marques, "Robust shape tracking in the presence of cluttered background," *IEEE Trans. Multimedia*, 6(6): 852–861, 2004.
  - 27) M.N. Kanyak, Q. Zhi, A.D. Cheok, S. Kuntal, Z. Jian, and K. Chung, "Lip geometric features for human-computer interaction using bimodal speech recognition: comparison and analysis," *Speech Communication*, 43(1–2): 1–16, 2004.
  - 28) C. Sanderson and K.K. Paliwal, "Identity verification using speech and face information," *Digital Signal Processing*, 14: 449–480, 2004.
  - 29) S. Basu, P.C. de Cuetos, S.H. Maes, C.V. Neti, and A.W. Senior, *Method and Apparatus for Audio-Visual Speech Detection and Recognition*, U.S. Patent 6,816,836 B2, Nov. 2004.
  - 30) S. Lucey, T. Chen, S. Sridharan, and C. Chandran, "Integration strategies for audio-visual speech processing: Applied to text dependent speaker identification/verification," *IEEE Trans. Multimedia*, 7(3): 496–506, 2005.
  - 31) L. Xie, R.-C. Zhao, and Z.-Q. Liu, "Adaptive stream reliability modeling based on local dispersion measures for audio visual speech recognition," *Proc. Int. Conf. Machine Learning Cybernetics*, pp. 4852–4857, 2005.
  - 32) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
  - 33) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 34) A. Potamianos, G. Bouselmi, D. Dimitriadis, D. Fohr, R. Gemello, I. Illina, F. Maha, P. Maragos, M. Matassoni, V. Pitsikalis, J. Ramirez, E. Sanchez-Soto, J. Segura, and P. Svaizer, "Towards speaker and environmental robustness in ASR: The HIWIRE project," *Proc. Work. Speech Recognition and Intrinsic Variation (SRIV)*, pp. 135–142, 2006.
- **25 citations of:** G. Potamianos, E. Cosatto, H.P. Graf, and D.B. Roe, "Speaker independent audio-visual database for bimodal ASR," *Proc. Europ. Tutorial Research Work. Audio-Visual Speech Process.*, Rhodes, Greece, pp. 65–68, 1997.
- 1) T. Chen and R.R. Rao, "Audio-visual integration in multimodal communication," *Proceedings of the IEEE*, 86(5): 837–851, 1998.
  - 2) J. Ostermann, L.S. Chen, and T.S. Huang, "Animated talking head with personalized 3D head model," *J. VLSI Signal Process.*, 20(1–2): 97–105, 1998.
  - 3) K. Toyama, "Look, Ma – No hands! Hands-free cursor control with real-time 3D face tracking," *Proc. Works. Perceptual User Interface*, 1998.
  - 4) P. Teissier, J. Robert-Ribes, and J.L. Schwartz, "Comparing models for audiovisual fusion in a noisy-vowel recognition task," *IEEE Trans. Speech Audio Processing*, 7(6): 629–642, 1999.
  - 5) T. Chen, "Technologies for building networked collaborative environments," *Proc. Int. Conf. Image Process.*, vol. 3, pp. 16–20, 1999.
  - 6) C. Benoit, J. Martin, C. Pelachaud, L. Schomaker, and B. Suhm, "Audiovisual and multimodal speech systems", In *Handbook of Standards and Resources in Spoken Language Systems*, D. Gibbon (Ed.), Kluwer Academic Publishers, 2000.
  - 7) T. Chen, "Audiovisual speech processing," *Signal Process. Magazine*, 18(1): 9–21, 2001.

- 8) K. Iwano, S. Tamura, and S. Furui, "Bimodal speech recognition using lip movement measured by optical-flow analysis," *Proc. Work. on Hands-Free Speech Communication (HSC)*, pp. 187–190, 2001.
  - 9) P. Scanlon and R. Reilly, "Lessons from speechreading," *Proc. Int. Conf. Multimedia Expo*, pp. 555–558, 2001.
  - 10) I. Matthews, T.F. Cootes, J.A. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Analysis Machine Intell.*, 24(2): 198–213, 2002.
  - 11) E.K. Patterson, S. Gurbuz, Z. Tufekci, and J.N. Gowdy, "Moving talker, speaker-independent feature study, and baseline results using the CUAVE multimodal speech corpus," *EURASIP J. Appl. Signal Process.*, 2002(11): 1189–1201, 2002.
  - 12) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
  - 13) E. Patterson, S. Gurbuz, Z. Tufekci, and J.N. Gowdy, "CUAVE: A new audio-visual database for multimodal human-computer interface research," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2017–2020, 2002.
  - 14) M. Zelezny and Petr Cisar, "Czech audio-visual speech corpus of a car driver for in-vehicles audio-visual speech recognition," *Proc. Work. Audio-Visual Speech Process.*, pp. 169–173, 2003.
  - 15) P. Daubias and P. Deléglise, "The LIUM-AVS database: a corpus to test lip segmentation and speechreading systems in natural conditions," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1569–1571, 2003.
  - 16) T. Yoshinaga, S. Tamura, K. Iwano, and S. Furui, "Audio-visual speech recognition using lip motion movement extracted from side-face images," *Proc. Work. Audio-Visual Speech Process.*, pp. 117–120, 2003.
  - 17) P. Scanlon, R. Reilly, and P. de Chazal, "Visual feature analysis for automatic speechreading," *Proc. Work. Audio-Visual Speech Process.*, pp. 127–132, 2003.
  - 18) S.W. Foo, Y. Lian, and L. Dong, "Recognition of visual speech elements using adaptively boosted hidden Markov models," *IEEE Trans. Circuits Systems Video Techn.*, 14(5): 693–705, 2004.
  - 19) S. Tamura, K. Iwano, and S. Furui, "Multi-modal speech recognition using optical-flow analysis for lip-images," *J. VLSI Signal Process.*, 36(2–3): 117–124, 2004.
  - 20) S. Tamura, K. Iwano, and S. Furui, "A stream-weight optimization method for audio-visual speech recognition using multi-stream HMMs," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 857–860, 2004.
  - 21) S. Tamura, K. Iwano, and S. Furui, "Improvement of audio-visual speech recognition in cars," *Proc. Int. Congress on Acoustics (ICA)*, vol. 4, pp. 2595–2598, 2004.
  - 22) T. Yoshinaga, S. Tamura, K. Iwano, and S. Furui, "Audio-visual speech recognition using new lip features extracted from side-face images," *Proc. ISCA Tutorial and Works. on Robustness Issues in Conversational Interaction (ROBUST)*, 2004.
  - 23) A. Ortega, F. Sukno, E. Lleida, A. Frangi, A. Miguel, L. Buera, and E. Zacur, "AV@CAR: A Spanish multichannel multimodal corpus for in-vehicle automatic audio-visual speech recognition," *Proc. Int. Conf. on Language Resources and Evaluation*, 2004.
  - 24) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, Research School of Information Sciences and Engineering, The Australian National University, Canberra, Australia, 2004.
  - 25) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
- **21 citations of:** G. Potamianos, J. Luetttin, and C. Neti, "Hierarchical discriminant features for audio-visual LVCSR," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 165–168, Salt Lake City, UT, 2001.
    - 1) A.V. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, "Dynamic Bayesian networks for audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1274–1288, 2002.
    - 2) J. Jiang, A. Alwan, P.A. Keating, B. Chaney, E.T. Auer, Jr., and L.E. Bernstein, "On the relationship between face movements, tongue movements, and speech acoustics," *EURASIP J. Appl. Signal Processing*, 2002(11): 1174–1188, 2002.
    - 3) P.S. Aleksic, J.J. Williams, Z. Wu, and A.K. Katsaggelos, "Audio-visual speech recognition using MPEG-4 compliant visual features," *EURASIP J. Appl. Signal Process.*, 2002(11): 1213–1227, 2002.
    - 4) X. Zhang, C.C. Broun, R.M. Mersereau, and M. Clements, "Automatic speechreading with applications to human-computer interfaces," *EURASIP J. Appl. Signal Processing*, 2002(11): 1228–1247, 2002.
    - 5) X. Zhang, R.M. Mersereau, M. Clements, and C.C. Broun, "Visual speech feature extraction for improved speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 1993–1996, 2002.
    - 6) X. Zhang, R.M. Mersereau, and M.A. Clements, "Audio-visual speech recognition by speechreading," *Proc. Int. Conf. Digital Signal Process.*, vol. 2, pp. 1069–1072, 2002.
    - 7) S. Gurbuz, Z. Tufekci, E. Patterson, and J.N. Gowdy, "Multi-stream product modal audio-visual integration strategy for robust adaptive speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2021–2024, 2002.
    - 8) H.-X. Yao, W. Gao, W. Shan, and M.-H. Xu, "Visual features extracting and selecting for lipreading," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 251–259, 2003.
    - 9) R. Goecke and B. Millar, "Statistical analysis of the relationship between audio and video speech parameters for Australian English," *Proc. Work. Audio-Visual Speech Process.*, pp. 133–138, 2003.
    - 10) S. Lucey, "An evaluation of visual speech features for the tasks of speech and speaker recognition," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 260–267, 2003.
    - 11) R. Séguier and N. Cladel, "Genetic snakes: Application on lipreading," *Proc. Int. Conf. on Artificial Neural Networks and Genetic Alg. (ICANNGA)*, 2003.
    - 12) P.S. Aleksic and A.K. Katsaggelos, "Speech-to-video synthesis using MPEG-4 compliant visual features," *IEEE Trans. Circuits Syst. Video Technol.*, 14(5): 682–692, 2004.
    - 13) P. Motlicek, L. Burget, J. Cernocky, and I. Potucek, "Phoneme recognition of meetings using audio-visual data," *Proc. Joint AMI/PASCAL/IM2/M4 Works.*, pp. 6–11, 2004.
    - 14) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, Research School of Information Sciences and Engineering, The Australian National University, Canberra, Australia, 2004.
    - 15) R. Goecke, "Audio-video automatic speech recognition: An example of improved performance through multimodal sensor input," *Proc. 2005 NICTA-HCSNet Multimodal User Interaction Works.*, pp. 25–32, 2005.
    - 16) R. Goecke, "Current trends in joint audio-visual signal processing: a review," *Proc. Int. Symp. Signal Process. and its Applications*, pp. 70–73, 2005.
    - 17) H. Glotin, S. Tollari, and P. Giraudet, "Approximation of linear discriminant analysis for word dependent visual features selection," *Proc. Conf. Advanced Concepts for Intell. Vision Systems (ACIV)*, Springer LNCS 3708, pp. 170–177, 2005.

- 18) R. Seymour, J. Ming, and D. Stewart, "A new posterior based audio-visual integration method for robust speech recognition," *Proc. Interspeech*, pp. 1229–1232, 2005.
  - 19) P. Scanlon, *Audio and Visual Feature Analysis for Speech Recognition*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 20) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
  - 21) J. Kaukenas, G. Navickas, and L. Telksnys, "Human-computer audiovisual interface," *Information Technology and Control*, 35(2): 87–92, 2006.
- **20 citations of:** G. Potamianos, C. Neti, J. Luetin, and I. Matthews, "Audio-Visual Automatic Speech Recognition: An Overview," *Audio-Visual Speech Processing*, E. Vatikiotis-Bateson, G. Bailly, and P. Perrier (Eds.), MIT Press, 2006.
    - 1) I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard, and D. Zhang, "Automatic analysis of multimodal group actions in meetings," *IEEE Trans. Pattern Anal. Machine Intell.*, 27(3): 305–317, 2005.
    - 2) S. Oviatt and R. Lunsford, "Multimodal interfaces for cell phones and mobile technology," *J. of Sol-Gel Science and Technology*, 8(2): 127–132, 2005.
    - 3) A. Nurnberger and M. Detyniecki, "Adaptive multimedia retrieval: From data to user interaction," In *Do Smart Adaptive Systems Exist? – Best Practice for Selection and Combination of Intelligent Methods*, J. Strackeljan, K. Leivisk, and B. Gabrys (Eds.), Springer-Verlag, Berlin, 2005.
    - 4) M.H. Coen, "Cross-modal clustering," *Proc. National Conf. Artificial Intelligence (AAAI)*, pp. 932–937, 2005.
    - 5) S. Bengio and H. Bourlard, "Multi-channel sequence processing," *IDIAP Research Report RR 05-04*, Martigny, Switzerland, 2005.
    - 6) S. Lucey and P. Lucey, "Improved speech reading through a free-parts representation," *Proc. Int. Conf. Auditory-Visual Speech Process.*, pp. 85–86, 2005.
    - 7) P. Lucey, D. Dean, and S. Sridharan, "Problems associated with current area-based visual speech feature extraction techniques," *Proc. Int. Conf. Auditory-Visual Speech Process.*, pp. 73–78, 2005.
    - 8) P. Lucey, S. Lucey, and S. Sridharan, "Using a free-parts representation for visual speech recognition," *Proc. Digital Image Computing: Techniques and Applications (DICTA)*, p. 55, 2005.
    - 9) D. Dimitriadis, N. Katsamanis, P. Maragos, G. Papandreou, and V. Pitsikalis, "Towards automatic speech recognition in adverse environments," *Proc. HERCMA – Hellenic European Conf. on Research on Computer Mathematics and its Applications*, 2005.
    - 10) I. Arsic, N. Marina, and J.-P. Thiran, "Impact of sample sizes on information theoretic measures for audio-visual signal processing," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2005.
    - 11) M. Gurban and J.-P. Thiran, "Audio-visual speech recognition with a hybrid SVM-HMM system," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2005.
    - 12) M. Gurban and J.-P. Thiran, "An information theoretic perspective on multimodal signal processing," *ITS Technical Report, TR\_ITS.2005.38*, EPFL, Lausanne, Switzerland, 2005.
    - 13) J. Zibert, N. Pavesic, and F. Mihelic, "Speech/non-speech segmentation based on phoneme recognition features," *EURASIP J. Appl. Signal Processing*, 2006: 1–13, 2006.
    - 14) A. Jaimes, N. Sebe, and D. Gatica-Perez, "Human-centered computing: A multimedia perspective," *Proc. ACM Int. Conf. Multimedia*, pp. 855–864, 2006.
    - 15) E.C. Kaiser, "Using redundant speech and handwriting for learning new vocabulary and understanding abbreviations," *Proc. Int. Conf. Multimodal Interfaces (ICMI)*, pp. 347–356, 2006.
    - 16) M. Gurban and J.-P. Thiran, "Multimodal speaker localization in a probabilistic framework," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
    - 17) I. Arsic, R. Vilagut, and J.-P. Thiran, "Automatic extraction of geometric lip features with application to multi-modal speaker identification," *Proc. Int. Conf. Multimedia Expo*, pp. 161–164, 2006.
    - 18) I. Arsic and J.-P. Thiran, "Mutual information eigenlips for audio-visual speech recognition," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
    - 19) F. Shafait, R. Kricke, I. Shdaifat, and R.-R. Grigat, "Real time lip motion analysis for a person authentication system using near infrared illumination," *Proc. Int. Conf. Image Process.*, pp. 1957–1960, 2006.
    - 20) U. Saeed, F. Matta, and J.-L. Dugelay, "Person recognition based on head and mouth dynamics," *Int. Works. Multimedia Signal Process.*, 2006.
  - **18 citations of:** H. Glotin, D. Vergyri, C. Neti, G. Potamianos, and J. Luetin, "Weighting schemes for audio-visual fusion in speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 173–176, Salt Lake City, UT, 2001.
    - 1) M. Heckmann, T. Wild, F. Berthommier, and K. Kroschel, "Comparing audio- and a-posteriori-probability-based stream confidence measures for audio-visual speech recognition," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1023–1026, 2001.
    - 2) M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1260–1273, 2002.
    - 3) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
    - 4) P. Daubias and P. Deléglise, "Lip-reading based on a fully automatic statistical model," *Proc. Int. Conf. Spoken Lang. Process.*, 2002.
    - 5) S. Gurbuz, Z. Tufekci, E. Patterson, and J.N. Gowdy, "Multi-stream product modal audio-visual integration strategy for robust adaptive speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2021–2024, 2002.
    - 6) T.T. Kristjansson, *Speech Recognition in Adverse Environments: A Probabilistic Approach*, Dept. of Computer Science, Univ. of Waterloo, Waterloo, Ontario, Canada, 2002.
    - 7) T.W. Lewis and D.M.W. Powers, "Audio-visual speech recognition using red exclusion and neural networks," *J. of Research and Practice in Information Theory*, 35(1): 41–64, 2003.
    - 8) P.S. Aleksic and A.K. Katsaggelos, "Product HMMs for audio-visual continuous speech recognition using facial animation parameters," *Proc. Int. Conf. Multimedia and Expo*, vol. 1, pp. 481–484, 2003.
    - 9) P.S. Aleksic and A.K. Katsaggelos, "Speech-to-video synthesis using MPEG-4 compliant visual features," *IEEE Trans. Circuits Syst. Video Technol.*, 14(5): 682–692, 2004.
    - 10) P. Aarabi and B.D. Dasarathy, "Robust speech processing using multi-sensor multi-source information fusion – an overview of the state of the art," *Information Fusion*, 5(2): 77–80, 2004.

- 11) G.F. Meyer, J.B. Mulligan, and S.M. Wuerger, "Continuous audio-visual digit recognition using N-best decision fusion," *Information Fusion*, 5: 91–101, 2004.
  - 12) T.W. Lewis and D.M.W. Powers, "Sensor fusion weighting measures in audio-visual speech recognition," *Proc. 27th conference on Australasian computer science*, pp. 305–314, 2004.
  - 13) P. Lucey, T. Martin, and S. Sridharan, "Confusability of phonemes grouped according to their viseme classes in noisy environments," *Proc. Australian Int. Conf. on Speech Science and Technology*, pp. 265–270, 2004.
  - 14) X. Lei, G. Ji, T. Ng, J. Bilmes, and M. Ostendorf, "DBN-based multi-stream models for Mandarin toneme recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 345–348, 2005.
  - 15) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
  - 16) N.A. Fox, *Audio and Video Based Person Identification*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 17) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
  - 18) A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, "Multimodal fusion by adaptive compensation for feature uncertainty with application to audiovisual speech recognition," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
- **17 citations of:** C. Neti, G. Potamianos, J. Luetttin, I. Matthews, H. Glotin, and D. Vergyri, "Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop," *Proc. IEEE Work. Multimedia Signal Process.*, pp. 619–624, Cannes, France, 2001.
    - 1) M. Gordan, C. Kotropoulos, and I. Pitas, "A support vector machine-based dynamic network for visual speech recognition applications," *EURASIP J. Appl. Signal Processing*, 2002(11): 1248–1259, 2002.
    - 2) J. Healey, R. Hosn, and S.H. Maes, "Adaptive content for device independent multi-modal browser applications," *Proc. Second Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems*, pp. 401–405, 2002.
    - 3) X. Li and R.M. Stern, "Training of stream weights for the decoding of speech using parallel feature streams," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 880–883, 2003.
    - 4) I. Shdaifat, R. Grigat, and D. Langmann, "A system for automatic lip reading," *Proc. Work. Auditory-Visual Speech Process.*, pp. 121–125, 2003.
    - 5) T.W. Lewis and D.M.W. Powers, "Audio-visual speech recognition using red exclusion and neural networks," *J. of Research and Practice in Information Theory*, 35(1): 41–64, 2003.
    - 6) H.-X. Yao, W. Gao, W. Shan, and M.-H. Xu, "Visual features extracting and selecting for lipreading," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 251–259, 2003.
    - 7) B.A. Kim, *Multi-Source Human Identification*, M.S.E. Thesis, MIT, May 2003.
    - 8) X. Li and R.M. Stern, "Parallel feature generation based on maximizing normalized acoustic likelihood," *Proc. Int. Conf. Spoken Lang. Process.*, 2004.
    - 9) K. Maghooli and M.S. Moin, "A new approach on multimodal biometrics based on combining neural networks using Adaboost," *Proc. Biometric Authentication ECCV Workshop*, pp. 332–341, 2004.
    - 10) P. Lucey, T. Martin, and S. Sridharan, "Confusability of phonemes grouped according to their viseme classes in noisy environments," *Proc. Australian Int. Conf. on Speech Science and Technology*, pp. 265–270, 2004.
    - 11) M.J. Sánchez Martínez and J.P. de la Cruz Gutiérrez, "Audio-visual speech recognition using motion based lipreading," *Proc. Int. Conf. Spoken Language Process.*, 2004.
    - 12) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, The Australian National University, Jan. 2004.
    - 13) E. Erzin, Y. Yemez, and A.M. Tekalp, "Multimodal speaker identification using an adaptive classifier cascade based on modality reliability," *IEEE Trans. Multimedia*, 7(5): 840–852, 2005.
    - 14) E.C. Kaiser, "Multimodal new vocabulary recognition through speech and handwriting in a whiteboard scheduling application," *Proc. Int. Conf. Intelligent User Interfaces*, pp. 51–58, 2005.
    - 15) L.J.M. Rothkrantz, J.C. Wojdel, and P. Wiggers, "Fusing data streams in continuous audio-visual speech recognition," *Proc. Int. Conf. Text, Speech and Dialogue*, J.G. Carbonell and J. Siekmann (eds.), Springer LNAI 3658, pp. 33–44, 2005.
    - 16) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
    - 17) J. Saragih and R. Goecke, "Learning active appearance models from image sequences," *Proc. HCSNet Works. Use of Vision in HCI (VisHCI)*, 2006.
  - **15 citations of:** J. Luetttin, G. Potamianos, and C. Neti, "Asynchronous stream modeling for large-vocabulary audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 169–172, Salt Lake City, UT, 2001.
    - 1) J.A. Bilmes, G. Zweig, T. Richardson, et al., "Discriminatively structured graphical models for speech recognition," Final Workshop 2001 Report, *Technical Report*, Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD, 2001.
    - 2) H. Glotin, "Dominant speaker detection based on voicing for adaptive audio-visual ASR robust to speech noise," *Proc. ITRW on Adaptation Methods for Speech Recognition (Adaptation)*, pp. 89–92, 2001.
    - 3) X. Zhang, C.C. Broun, R.M. Mersereau, and M. Clements, "Automatic speechreading with applications to human-computer interfaces," *EURASIP J. Appl. Signal Processing*, 2002(11): 1228–1247, 2002.
    - 4) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
    - 5) A.V. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, "Dynamic Bayesian networks for audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1274–1288, 2002.
    - 6) A.V. Nefian, L. Liang, X. Pi, L. Xiaoxiang, C. Mao, and K. Murphy, "A coupled HMM for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2013–2016, 2002.
    - 7) S. Gurbuz, Z. Tufekci, E. Patterson, and J.N. Gowdy, "Multi-stream product modal audio-visual integration strategy for robust adaptive speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2021–2024, 2002.
    - 8) O. Cetin and M. Ostendorf, "Cross-stream observation dependencies for multi-stream speech recognition," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 2517–2520, 2003.
    - 9) M. Nagarajan and T.V. Sreenivas, "Product HMM – A novel class of HMMs for sub-sequence modelling," *Proc. Work. Spoken Language Process.*, Mumbai, India, Jan. 9–11, 2003.

- 10) H.-X. Yao, W. Gao, W. Shan, and M.-H. Xu, "Visual features extracting and selecting for lipreading," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 251–259, 2003.
  - 11) O. Cetin, *Multi-Rate Modeling, Model Inference, and Estimation for Statistical Classifiers*, Ph.D. Thesis, University of Washington, 2004.
  - 12) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, The Australian National University, Jan. 2004.
  - 13) S. Lucey, T. Chen, S. Sridharan, and C. Chandran, "Integration strategies for audio-visual speech processing: Applied to text dependent speaker identification/verification," *IEEE Trans. Multimedia*, 7(3): 496–506, 2005.
  - 14) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
  - 15) A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos, "Multimodal fusion by adaptive compensation for feature uncertainty with application to audiovisual speech recognition," *Proc. Europ. Signal Process. Conf. (Eusipco)*, 2006.
- **12 citations of:** G. Potamianos and C. Neti, "Stream confidence estimation for audio-visual speech recognition," *Proc. Intern. Conf. Spoken Language Process.*, pp. 746–749, Beijing, China, 2000.
    - 1) M.T. Chan, "HMM-based audio-visual speech recognition integrating geometric- and appearance-based visual features," *Proc. IEEE Work. Multimedia Signal Process.*, pp. 9–14, 2001.
    - 2) M. Heckmann, T. Wild, F. Berthommier, and K. Kroschel, "Comparing audio- and a-posteriori-probability-based stream confidence measures for audio-visual speech recognition," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1023–1026, 2001.
    - 3) M. Heckmann, F. Berthommier, and K. Kroschel, "A hybrid ANN/HMM audio-visual speech recognition system," *Proc. Work. Audio-Visual Speech Process.*, 2001.
    - 4) M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1260–1273, 2002.
    - 5) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
    - 6) S. Gurbuz, Z. Tufekci, E. Patterson, and J.N. Gowdy, "Multi-stream product modal audio-visual integration strategy for robust adaptive speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 2021–2024, 2002.
    - 7) S. Oviatt, "Advances in robust multimodal interface design," *IEEE Computer Graphics and Applications*, 23(5): 62–68, 2003.
    - 8) S. Oviatt, "User-centered modeling and evaluation of multimodal interfaces," *Proceedings of the IEEE*, 91(9): 1457–1468, 2003.
    - 9) T.W. Lewis and D.M.W. Powers, "Audio-visual speech recognition using red exclusion and neural networks," *J. of Research and Practice in Information Theory*, 35(1): 41–64, 2003.
    - 10) M. Heckmann, F. Berthommier, C. Savariaux, and K. Kroschel, "Effects of image distortions on audio-visual speech recognition," *Proc. Work. Audio-Visual Speech Process.*, pp. 163–168, 2003.
    - 11) T.W. Lewis and D.M.W. Powers, "Sensor fusion weighting measures in audio-visual speech recognition," *Proc. 27th conference on Australasian computer science*, pp. 305–314, 2004.
    - 12) L. Xie, R.-C. Zhao, and Z.-Q. Liu, "Adaptive stream reliability modeling based on local dispersion measures for audio visual speech recognition," *Proc. Int. Conf. Machine Learning Cybernetics*, vol. 8, pp. 4852–4857, 2005.
  - **12 citations of:** G. Potamianos, C. Neti, G. Iyengar, and E. Helmuth, "Large-vocabulary audio-visual speech recognition by machines and humans," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1027–1030, Aalborg, Denmark, 2001.
    - 1) T.E. Starner, "The role of speech input in wearable computing," *IEEE Pervasive Computing J.*, 1(3): 89–93, 2002.
    - 2) M. Heckmann, F. Berthommier, and K. Kroschel, "Noise adaptive stream weighting in audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1260–1273, 2002.
    - 3) E.K. Patterson, S. Gurbuz, Z. Tufekci, and J.N. Gowdy, "Moving talker, speaker-independent feature study, and baseline results using the CUAVE multimodal speech corpus," *EURASIP J. Appl. Signal Process.*, 2002(11): 1189–1201, 2002.
    - 4) P. Lamere, P. Kwok, W. Walker, E. Gouvea, R. Singh, B. Raj, and P. Wolf, "Design of the CMU Sphinx-4 decoder," *Proc. Europ. Conf. Speech Comm. Technol. (Eurospeech)*, pp. 1181–1184, 2003.
    - 5) K. Murai and S. Nakamura, "Real time face detection for multimodal speech recognition," *Proc. Int. Conf. Multimedia Expo*, vol. 2, pp. 373–376, 2003.
    - 6) E.K. Patterson and J.N. Gowdy, "An audio-visual approach to simultaneous-speaker speech recognition," *Proc. Int. Conf. Acoustics Speech Signal Process.*, vol. 5, pp. 780–783, 2003.
    - 7) K. Murai and S. Nakamura, "A robust bimodal speech section detection," *J. VLSI Signal Process.*, 36(2–3): 81–90, 2004.
    - 8) J. Kratt, F. Metze, R. Stiefelhagen, and A. Waibel, "Large vocabulary audio-visual speech recognition using the Janus speech recognition toolkit," *Proc. 26th DAGM Symposium on Pattern Recognition*, pp. 488–495, 2004.
    - 9) T. Tsunekawa, K. Hotto, and H. Takahashi, "Lipreading using recurrent neural prediction model," *Proc. Int. Conf. Image Analysis and Recognition*, pp. 405–412, 2004.
    - 10) M. McClain, K. Brady, M. Brandstein, and T. Quatieri, "Automated lip-reading for improved speech intelligibility," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 1, pp. 701–704, 2004.
    - 11) R. Prasad, H. Saruwatari, and K. Shikano, "Robots that hear, understand and talk," *Advanced Robotics*, 18(5): 533–564, 2004.
    - 12) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
  - **11 citations of:** I. Matthews, G. Potamianos, C. Neti, and J. Luetin, "A comparison of model and transform-based visual features for audio-visual LVCSR," *Proc. Intern. Conf. Multimedia Expo*, Tokyo, Japan, 2001.
    - 1) H.-X. Yao, W. Gao, W. Shan, and M.-H. Xu, "Visual features extracting and selecting for lipreading," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 251–259, 2003.
    - 2) Z. Wen and T.S. Huang, "Capturing subtle facial motions in 3D face tracking," *Proc. Int. Conf. Computer Vision (ICCV)*, vol. 2, pp. 1343–1350, 2003.
    - 3) P.S. Aleksic and A.K. Katsaggelos, "Comparison of low- and high-level visual features for audio-visual continuous automatic speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process.*, vol. 5, pp. 917–920, 2004.
    - 4) P.S. Aleksic and A.K. Katsaggelos, "Comparison of MPEG-4 facial animation parameter groups with respect to audio-visual speech recognition performance," *Proc. Int. Conf. Image Process.*, 2005.
    - 5) J. Chaloupka, "Extraction of the visual features by discrete cosine transform for audio-visual speech recognition," *Proc. Int. Conf. Radioelektronika*, 2005.

- 6) N.A. Fox, R. Gross, J.F. Cohn, and R.B. Reilly, "Robust automatic human identification using face, mouth, and acoustic information," *Proc. Works. Analysis and Modelling of Faces and Gestures (AMFG)*, Springer LNCS 3723, pp. 264–278, 2005.
  - 7) N.A. Fox, B.A. O'Mullane, and R.B. Reilly, "VALID: A new practical audio-visual database, and comparative results," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 3546, pp. 777–786, 2005.
  - 8) N.A. Fox, *Audio and Video Based Person Identification*, Ph.D. Thesis, Dept. of Electronic and Electrical Eng., University College Dublin, National University of Ireland, 2005.
  - 9) F. Shafait, R. Kricke, I. Shdaifat, and R.-R. Grigat, "Real time lip motion analysis for a person authentication system using near infrared illumination," *Proc. Int. Conf. Image Process.*, pp. 1957–1960, 2006.
  - 10) U. Saeed, F. Matta, and J.-L. Dugelay, "Person recognition based on head and mouth dynamics," *Int. Works. Multimedia Signal Process.*, 2006.
  - 11) N.A. Fox, R. Gross, J.F. Cohn, and R.B. Reilly, "Robust biometric person identification using automatic classifier fusion of speech, mouth, and face experts," To Appear: *IEEE Trans. Multimedia*, 2007.
- **10 citations of:** G. Potamianos, A. Verma, C. Neti, G. Iyengar, and S. Basu, "A cascade image transform for speaker independent automatic speechreading," *Proc. IEEE Int. Conf. Multimedia Expo*, vol. II, pp. 1097–1100, New York, NY, 2000.
    - 1) T. Chen, "Audiovisual speech processing," *Signal Process. Magazine*, 18(1): 9–21, 2001.
    - 2) C.C. Chibelushi, F. Deravi, and J.S.D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. Multimedia*, 4(1): 23–37, 2002.
    - 3) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
    - 4) M. Heckmann, F. Berthommier, C. Savariaux, and K. Kroschel, "Effects of image distortions on audio-visual speech recognition," *Proc. Work. Audio-Visual Speech Process.*, pp. 163–168, 2003.
    - 5) G.F. Meyer, J.B. Mulligan, and S.M. Wuergler, "Continuous audio-visual digit recognition using N-best decision fusion," *Information Fusion*, 5(2): 91–101, 2004.
    - 6) M.J. Sánchez Martínez and J.P. de la Cruz Gutiérrez, "Audio-visual speech recognition using motion based lipreading," *Proc. Int. Conf. Spoken Language Process.*, 2004.
    - 7) J. Kratt, F. Metz, R. Stiefelhagen, and A. Waibel, "Large vocabulary audio-visual speech recognition using the Janus speech recognition toolkit," *Proc. 26th DAGM Symposium on Pattern Recognition*, LNCS, pp. 488–495, 2004.
    - 8) R. Goecke, *A Stereo Vision Lip Tracking Algorithm and Subsequent Statistical Analyses of the Audio-Video Correlation in Australian English*, Ph.D. Thesis, Research School of Information Sciences and Engineering, The Australian National University, Canberra, Australia, 2004.
    - 9) D. Dean, P. Lucey, S. Sridharan, and T. Wark, "Comparing audio and visual information for speech processing," *Proc. Int. Symp. Signal Process. and Applications*, pp. 58–61, 2005.
    - 10) X. Hong, H. Yao, Y. Wan, and R. Chen, "A PCA based visual DCT feature extraction method for lip-reading," *Proc. Int. Conf. Intelligent Information Hiding and Multimedia*, pp. 321–326, 2006.
  - **10 citations of:** G. Gravier, G. Potamianos, and C. Neti, "Asynchrony modeling for audio-visual speech recognition," *Proc. Human Lang. Techn. Conf.*, pp. 1–6, 2002.
    - 1) A.V. Nefian, L. Liang, X. Pi, X. Liu, and K. Murphy, "Dynamic Bayesian networks for audio-visual speech recognition," *EURASIP J. Appl. Signal Process.*, 2002(11): 1274–1288, 2002.
    - 2) T.E. Starner, "The role of speech input in wearable computing," *IEEE Pervasive Computing J.*, 1(3): 89–93, 2002.
    - 3) H.-X. Yao, W. Gao, W. Shan, and M.-H. Xu, "Visual features extracting and selecting for lipreading," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, pp. 251–259, 2003.
    - 4) S.W. Foo, Y. Lian, and L. Dong, "Recognition of visual speech elements using adaptively boosted hidden Markov models," *IEEE Trans. Circuits Systems Video Techn.*, 14(5): 693–705, 2004.
    - 5) J. Kratt, F. Metz, R. Stiefelhagen, and A. Waibel, "Large vocabulary audio-visual speech recognition using the Janus speech recognition toolkit," *Proc. 26th DAGM Symposium on Pattern Recognition*, pp. 488–495, 2004.
    - 6) L. Dong, S.W. Foo, and Y. Lian, "A two-channel training algorithm for hidden Markov model and its application to lip reading," *EURASIP J. Applied Signal Process.*, 2005(9): 1382–1399, 2005.
    - 7) K. Saenko, K. Livescu, M. Siracusa, K. Wilson, J. Glass, and T. Darrell, "Visual speech recognition with loosely synchronized feature streams," *Proc. Int. Conf. Comp. Vision*, vol. 2, pp. 1424–1431, 2005.
    - 8) K. Saenko, K. Livescu, J. Glass, and T. Darrell, "Production domain modeling of pronunciation for visual speech recognition," *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, vol. 5, pp. 473–476, 2005.
    - 9) X. Li, *Combination and Generation of Parallel Feature Streams for Improved Speech Recognition*, Ph.D. Thesis, Electrical and Comp. Eng. Dept., Carnegie Mellon Univ., Pittsburgh, PA, 2005.
    - 10) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
  - **9 citations of:** C. Neti, G. Iyengar, G. Potamianos, A. Senior, and B. Maison, "Perceptual interfaces for information interaction: Joint processing of audio and visual information for human-computer interaction", *Proc. Int. Conf. Spoken Language Process.*, vol III, pp. 11-14, Beijing, China, 2000.
    - 1) C. Apte, L. Morgenstern, and S.J. Hong, "AI at IBM Research," *Intelligent Systems and their Applications*, 15(6): 51–57, 2000.
    - 2) S. Oviatt, "Advances in the robust processing of multimodal speech and pen systems," In *Multimodal Interfaces for Human Machine Communication*, P.C. Yuen and T.Y. Yan (Eds.), World Scientific Publisher: London, UK, 2001.
    - 3) P.S. Aleksic, J.J. Williams, Z. Wu, and A.K. Katsaggelos, "Audio-visual speech recognition using MPEG-4 compliant visual features," *EURASIP J. Appl. Signal Process.*, 2002(11): 1213–1227, 2002.
    - 4) S. Nakamura, "Statistical multimodal integration for audio-visual speech processing," *IEEE Trans. Neural Networks*, 13(4): 854–866, 2002.
    - 5) S. Oviatt, "User-centered modeling and evaluation of multimodal interfaces," *Proceedings of the IEEE*, 91(9): 1457–1468, 2003.
    - 6) L. Deng and X. Huang, "Challenges in adopting speech recognition," *Communications of the ACM*, 47(1): 69–75, 2004.
    - 7) S. Oviatt and R. Lunsford, "Multimodal interfaces for cell phones and mobile technology," *J. of Sol-Gel Science and Technology*, 8(2): 127–132, 2005.
    - 8) R. Lunsford, S. Oviatt, and R. Coulston, "Audio-visual cues distinguishing self- from system-directed speech in younger and older adults," *Proc. Int. Conf. Multimodal Interfaces*, pp. 167–174, 2005.

- 9) R. Lunsford, S. Oviatt, and A.M. Arthur, "Toward open-microphone engagement for multiparty interactions," *Proc. Int. Conf. Multimodal Interfaces*, pp. 273–280, 2006.
- **8 citations of:** E. Cosatto, G. Potamianos, and H.P. Graf, "Audio-visual unit selection for the synthesis of photo-realistic talking-heads," *Proc. IEEE Intern. Conf. Multimedia Expo.*, vol. II, pp. 619–622, New York, NY, 2000.
    - 1) J. Ostermann and M. Kampmann, "Source models for content-based video coding," *Proc. Int. Conf. Image Process.*, vol. 3, pp. 62–65, 2000.
    - 2) P.K. Kakumanu, *Audio-visual processing for speech-driven facial animation*, M.Sc. Thesis, Wright State University, 2002.
    - 3) K.F. Moore, *Use of Mouth Position and Mouth Movement to Filter Noise from Speech in a Hearing Aid*, U.S. Patent 6,707,921 B2, Mar. 2004.
    - 4) Z.-M. Wang, L.-H. Cai, and H.-Z. Ai, "Text-to-visual speech in Chinese based on data-driven approach," *Ruan Jian Xue Bao (J. Softw.)*, 16(6): 1054–1063, 2005.
    - 5) J. Ma, R. Cole, B. Pellom, W. Ward, and B. Wise, "Accurate visible speech synthesis based on concatenating variable length motion capture data," *IEEE Trans. Visualization Graphics*, 12(2): 266–276, 2006.
    - 6) Z. Wang and J. Tao, "Review of text-to-visual speech synthesis," *Computer Research and Development*, 43(1): 145–152, 2006.
    - 7) Z. Deng and U. Neumann, "eFASE: Expressive facial animation synthesis and editing with phoneme-isomap controls," *Proc. ACM SIGGRAPH Symp. Computer Animation*, 2006.
    - 8) S. Fagel, "Joint audio-visual unit selection – the JAVUS speech synthesizer," *Proc. Int. Conf. Speech and Computer (SPECOM)*, 2006.
  - **7 citations of:** G. Potamianos and C. Neti, "Audio-visual speech recognition in challenging environments," *Proc. Europ. Conf. Speech Comm. Technol.*, pp. 1293–1296, Geneva, Switzerland, 2003.
    - 1) P. Motlicek, L. Burget, J. Cernocky, and I. Potucek, "Phoneme recognition of meetings using audio-visual data," *Proc. Joint AMI/PASCAL/IM2/M4 Works.*, pp. 6–11, 2004.
    - 2) X. Zhang, K. Takeda, J.H.L. Hansen, and T. Maeno, "Audio-visual speaker localization for car navigation systems," *Proc. Conf. Spoken Lang. Process. (Interspeech)*, 2004.
    - 3) T.J. Hazen, K. Saenko, C.-H. La, and J.R. Glass, "A segment-based audio-visual speech recognizer: Data collection, development, and initial experiments," *Proc. Int. Conf. Multimodal Interfaces*, pp. 235–242, 2004.
    - 4) J. Huang and K. Visweswariah, "Improving lip-reading with feature space transforms for multi-stream audio-visual speech recognition," *Proc. Interspeech*, pp. 1221–1224, 2005.
    - 5) J. Huang and D. Povey, "Discriminatively trained features using fMPE for multi-stream audio-visual speech recognition," *Proc. Interspeech*, pp. 777–780, 2005.
    - 6) J. Huang, E. Marcheret, and K. Visweswariah, "Rapid feature space speaker adaptation for multi-stream HMM-based audio-visual speech recognition," *Proc. Int. Conf. Multimedia Expo*, pp. 338–341, 2005.
    - 7) T.J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," *IEEE Trans. Audio, Speech, and Language Process.*, 14(3): 1082–1088, 2006.
  - **5 citations of:** G. Potamianos and H.P. Graf, "Linear discriminant analysis for speechreading," *Proc. IEEE Work. Multimedia Signal Process.*, pp. 221–226, 1998.
    - 1) S.W. Foo and L. Dong, "A supervised two-channel learning method for hidden Markov model and application on lip reading," *Proc. Int. Conf. Advanced Learning Technologies (ICALT)*, 2002.
    - 2) E. Cosatto, *Sample-Based Talking-Head Synthesis*, Ph.D. Thesis, EPFL, Lausanne, Switzerland, 2002.
    - 3) S. Lucey, "An evaluation of visual speech features for the tasks of speech and speaker recognition," *Proc. Int. Conf. Audio Video-based Biometric Person Authentication*, Springer LNCS 2688, pp. 260–267, 2003.
    - 4) A. Sharma, K.K. Paliwal, and G.C. Onwubolu, "Class-dependent PCA, MDC, and LDA: A combined classifier for pattern classification," *Pattern Recogn.*, 39: 1215–1229, 2006.
    - 5) X. Hong, H. Yao, Y. Wan, and R. Chen, "A PCA based visual DCT feature extraction method for lip-reading," *Proc. Int. Conf. Intelligent Information Hiding and Multimedia*, pp. 321–326, 2006.
  - **12 citations of:** G. Potamianos and F. Jelinek, "A study of n-gram and decision tree letter language modeling methods," *Speech Communication*, vol. 24, no. 3, pp. 171–192, 1998.
    - 1) A. Corduneanu, "A pylonic decision-tree language model with optimal question selection," *Proc. Ann. Meeting Assoc. Comp. Linguistics*, pp. 606–609, 1999.
    - 2) D.H. Milone and A.J. Rubio, "Including prosodic cues in ASR systems," *Proc. Conf. Systemics, Cybernetics and Informatics (SCI)*, 2001.
    - 3) D.H. Milone and A.J. Rubio, "Prosodic and accentual information for automatic speech recognition," *IEEE Trans. Speech Audio Process.*, 11(4): 321–333, 2003.
    - 4) X. Zhao and T.P. Speed, "Probability models for short DNA motifs," *Proc. Conf. Española de Biometría*, 2003.
    - 5) R.K. Azad and M. Borodovsky, "Effects of choice of DNA sequence model structure on gene identification accuracy," *Bioinformatics*, 20(7): 993–1005, 2004.
    - 6) R.K. Azad and M. Borodovsky, "Probabilistic methods of identifying genes in prokaryotic genomes: Connections to the HMM theory," *Briefings in Bioinformatics*, 5(2): 118–130, 2004.
    - 7) L. Zhuang, T. Bao, X. Zhu, C. Wang, and S. Naoi, "A Chinese OCR spelling check approach based on statistical language models," *Proc. Int. Conf. Systems, Man, and Cybernetics*, 2004.
    - 8) X. Zhao, H. Huang, and T.P. Speed, "Finding short DNA motifs using permuted Markov models," *Proc. Int. Conf. Computational Molecular Biology*, pp. 68–75, 2004.
    - 9) C. Zhong and M. Seligman, "Toward practical spoken language translation," *Machine Translation*, 19(2): 113–137, 2005.
    - 10) L. Barari and B. Quasemi Zadeh, "CloniZER spell checker: Adaptive, language independent spell checker," *Proc. AIML Conf.*, 2005.
    - 11) P. Xu and L. Mangu, "Using random forest language models in the IBM RT-04 CTS system," *Proc. Interspeech*, pp. 741–744, 2005.
    - 12) M. Topkara, G. Riccardi, D. Hakkani-Tür, and M.J. Atallah, "Natural language watermarking: Challenges in building a practical system," *Proc. SPIE Int. Conf. Security, Stenography, and Watermarking of Multimedia Contents*, 2006.

- **11 citations of:** G.G. Potamianos and J. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simulations," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1322–1332, 1993.
  - 1) H. Lucke, "Improved acoustic modeling for speech recognition using 2D Markov random fields," *Proc. Int. Conf. Acoustics, Speech, Signal Process.*, vol. 1, pp. 540–543, 1995.
  - 2) J.-S. Wang, "Cluster Monte Carlo algorithms and their applications," In *Lecture Notes in Computer Science: Invited Session Papers from the Second Asian Conference on Computer Vision: Recent Developments in Computer Vision*, vol. 1035, pp. 307–315, 1995.
  - 3) B.J. Frey, F.R. Kschischang, and P.G. Gulak, "Early detection and trellis splicing: Reduced-complexity soft iterative coding," *Proc. of the 1996 Turbo-coding Seminar*, T. Maseng (Ed.), Lund, Sweden, 1996.
  - 4) Z.Y. Zhou, R.M. Leahy, and J.Y. Qi, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," *IEEE Trans. Image Process.*, 6(6): 844–861, 1997.
  - 5) S. DellaPietra, V. DellaPietra, and J. Lafferty, "Inducing features of random fields," *IEEE Trans. Pattern Analysis Machine Intell.*, 19(4): 380–393, 1997.
  - 6) S.S. Saquib, *Edge Preserving Models and Efficient Algorithms for Ill-Posed Inverse Problems in Image Processing*, Ph.D. Thesis, Purdue University, May 1997.
  - 7) S.S. Saquib, C.A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields with applications to Bayesian tomography," *IEEE Trans. Image Process.*, 7(7): 1029–1044, 1998.
  - 8) N. Cressie and C. Liu, "Binary Markov mesh models and symmetric Markov random fields: Some results on their equivalence," *Methodology and Computing in Applied Probability*, 3(1), 2001.
  - 9) S.C. Zhu and X. Liu, "Learning in Gibbsian fields: How accurate and how fast can it be?" *IEEE Trans. Pattern Analysis Machine Intell.*, 24(7): 1001–1006, 2002.
  - 10) A. Baddeley, G. Nair, and N. Cressie, "Directed Markov point processes – characterisation and construction," *Research Report 2002/14*, Dept. of Mathematics and Statistics, Univ. Western Australia, July 2002.
  - 11) M.V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Systems, Man, Cybernetics – Part B*, 35(3), 527–537, 2005.
- **10 citations of:** G. Potamianos and J. Goutsias, "Stochastic approximation algorithms for partition function estimation of Gibbs random fields," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1948–1965, 1997.
  - 1) F. Champagnat, J. Idier, and Y. Goussard, "Stationary Markov random fields on a finite rectangular lattice," *IEEE Trans. Information Theory*, 44(7): 2901–2916, 1998.
  - 2) M.J. Turmon and S. Mukhtar, "Representing solar active regions with triangulations," *Proc. Computational Statistics*, 1998.
  - 3) R.Y. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and Computing in Applied Probability*, 2: 127–190, 1999.
  - 4) F. Huang and Y. Ogata, "Comparison of two methods for calculating the partition functions of various spatial statistical models," *Australian and New Zealand Journal of Statistics*, 43(1): 47–65, 2001.
  - 5) R.Y. Rubinstein, "Combinatorial optimization, cross-entropy, ants and rare events," In: *Stochastic Optimization, Algorithms, and Applications*, S. Uryasev and P.M. Pardalos (Eds.), Kluwer Academic Publishers, 2001.
  - 6) M. Wainwright, T. Jaakkola, and A. Willsky, "Tree-based parametrization for approximate estimation of stochastic processes on graphs with cycles," *LIDS Technical Report P-2510*, MIT, Cambridge, MA, 2001.
  - 7) S.C. Zhu and X. Liu, "Learning in Gibbsian fields: How accurate and how fast can it be?" *IEEE Trans. Pattern Analysis Machine Intell.*, 24(7): 1001–1006, 2002.
  - 8) F. Forbes and N. Peyrard, "Hidden Markov random field model selection criteria based on mean field-like approximations," *IEEE Trans. Pattern Analysis Machine Intell.*, 25(9): 1089–1101, 2003.
  - 9) M.V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Trans. Systems, Man, Cybernetics – Part B*, 35(3), 527–537, 2005.
  - 10) M. Wainwright, T. Jaakkola, and A. Willsky, "A new class of upper bounds on the log partition function," *IEEE Trans. Information Theory*, 51(7): 2313–2335, 2005.