Gerasimos Potamianos

Associate Professor, Dept. of Electrical and Computer Engineering
University of Thessaly
37, Glavani Street, Room G-3/2
GR 38221, Volos, Greece

Tel: +30 24210 74928   *Fax*: +30 24210 74997
E-mail: gpotam@ieee.org   *URL*: www.e-ce.uth.gr/~gpotamianos

Research

Multimodal speech processing with applications to human-computer interaction and ambient intelligence. Particular emphasis on audio-visual speech processing, far-field automatic speech recognition, multichannel acoustic scene analysis, multimedia signal processing and fusion, sign language recognition, computer vision for human tracking and activity recognition, incorporating deep learning methods.

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

Feb. 2012 – Present : **Associate Professor**, *Department of Electrical and Computer Engineering*, UNIVERSITY OF THESSALY, Volos, Greece.

Has been *teaching* courses in the areas of signal and speech processing, machine learning, and computer vision. *Research activities* include work on far-field automatic speech recognition, multichannel acoustic and audio-visual scene analysis, human-computer and child-robot interaction, as part of recently completed FP7 EU project "DIRHA" and H2020 EU project "BabyRobot" in collaboration with the Robot Perception and Interaction Unit of the Athena Research and Innovation Center in Greece, work on audio-visual speech processing in collaboration with the IBM T. J. Watson Research Center in Yorktown Heights, New York and the Liopa startup at Belfast, Ireland and as part of recently completed PEOPLE-IRG project "AVISPIRE", in collaboration with the Institute of Informatics and Telecommunications of the National Centre for Scientific Research (NCSR), "Demokritos", in Greece, as well as work on video-based sign language recognition as part of the recently awarded SL-ReDu National Greek NSF Project in collaboration with the Athena Research and Innovation Center in Greece. Has been *supervising* numerous students as part of their Diploma, Master, or Ph.D. Theses. Further, from Oct. 2014 till Sept. 2017 has *served* as Member of the Research Committee of the University, since Oct. 2015 has been Director or Deputy Director of Graduate Studies at the ECE Dept., and since Oct. 2018 Vice-Chair at the Computer Science Dept. of the University.

Apr. 2009 – Jan. 2012 : **Director of Research** (Rank A Researcher), *Institute of Informatics and Telecommunications*, NATIONAL CENTRE FOR SCIENTIFIC RESEARCH (NCSR), "DEMOKRITOS", Athens, Greece.

Has been a member of the Software and Knowledge Engineering Laboratory of the Institute, conducting *research* on multimodal speech processing, bridging existing lab expertise on text, image, and video processing technologies, as recipient of Marie Curie Grant (PEOPLE-IRG) for project "AVISPIRE". Has also *served* as the Technical Coordinator of FP7 EU project "SYNC3" and contributed to FP6 EU project "INDIGO". Has also *taught* courses at the Department of Electrical and Computer Engineering of the National Technical University of Athens, Greece (Spring 2011) and of the University of Thessaly, Volos, Greece (Fall 2011).

Oct. 2008 – March 2009 : **Senior Researcher** (Rank B), *Institute of Computer Science*, FOUNDATION FOR RESEARCH AND TECHNOLOGY HELLAS (FORTH), Heraklion, Crete, Greece.

Has been a member of the Telecommunications and Networks Laboratory at FORTH, working within the ambient intelligence program of the Institute.

Sept. 1999 – Sept. 2008 : **Research Staff Member** and **Manager** (Feb. 2007 – Sept. 2008), *Multimodal Conversational Technologies Department*, IBM T. J. WATSON RESEARCH CENTER, Yorktown Heights, New York.

Has led 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. From July 2008 till Sept. 2008, has also led efforts in the speech-to-speech translation area, supervising a combined group of over twenty researchers.

Own research work has focused on audio-visual speech technologies with applications to human-computer interaction and ambient intelligence. Effort has concentrated on a variety of problems including visual speech representation, fusion algorithms for large-vocabulary continuous audio-visual speech recognition, bimodal speech enhancement, speaker recognition, audio-visual speech activity detection, far-field automatic speech recognition, 2D and 3D person tracking, and real-time algorithmic implementation, with emphasis on applications to the automobile and smart rooms/homes. Most recent efforts have been driven by three FP6 European research projects ("CHIL", "NETCARITY", and "DICIT").

Aug. 1996 – July 1998 and Sept. 1998 – 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.

Publications	• Twenty-One (21) Articles in Journals and Books (Additional One (1) Under Review). • One-Hundred-Fifteen (115) Articles in Conference and Workshop Proceedings. • Three (3) Volume Edited Book Series. • Three (3) Editorials in Special Journal Issues. • Seven (7) Granted Patents. • Two (2) Short Courses / Tutorials. • Eight (8) Technical Reports. • Over 5.5k Works Citing these Publications. • Publication Impact: h-index = 38; i10-index = 97.

Presentations

- Invited speaker at the 2018 LISTEN Summer School, Bonn, Germany.
- Invited speaker at ACCVW 2016, Taipei, Taiwan.
- Lecturer at the ISCA 2014 Summer School (SPCC), Crete, Greece.
- Invited speaker at ASRU 2009, Merano, Italy.
- Lecturer at the NCSR "Demokritos" Summer School 2009 and 2010, Athens, Greece.
- 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.
- Lecturer at the ELSNET 2001 Summer School, Prague, Czech Republic.
- Fifty-nine (59) conference paper presentations (1989 – 2019).
- Numerous invited presentations at Universities and Industrial Research Labs.

Professional Activities

- Organizing Committee Member of Upcoming ICASSP 2023.
- Organizing Committee Member (one of two Publication Chairs) of SLT 2018.
- Organizing Committee Member (one of four Technical Program Chairs) of EUSIPCO 2017.
- Area Editor of Three-Volume Handbook of Multimodal-Multisensor Interfaces (Eds.: S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and A. Krüger), Spring 2017 – 2018.
- Member of the IEEE Speech and Language Technical Committee (SLTC), 2009 – 2011.
- Guest Editor of the special issue on Interactive Multimedia in the IEEE Transactions on Multimedia, Sept. 2011.
- Guest Editor of the special issue on Multimodal Speech Processing in the IEEE Transactions on Audio, Speech and Language Processing, March 2009.
- 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, Nov. 2002.
- Area Chair for the EUSIPCO 2008, 2012, MLMI 2008, ICMI-MLMI 2009, ICMI 2015, and ACL 2020 Conferences.
- Area Coordinator, Interspeech 2008, 2009 Conferences.
- Member of the Christian Benoit Award Committee (2007, 2009, 2011, 2013, 2015, 2017).
- Member of the European Biometrics Forum Award Committee (2006 – 2008).
- 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; Computer Speech and Language; Computer Vision and Image Understanding; IEEE Signal Processing Letters; 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; IEEE Signal Processing Magazine; Information Fusion; Multimedia Systems).
- Reviewer for numerous conferences (ACM MM, ASRU, AVSP, DCASE, EUSIPCO, ICASSP, ICME, ICMI, Interspeech, IS ADEPT, MLMI, MMSP, MUM, PETRA, SLT).

Administration   *Acting Chair* of the Ethics Committee of the Athena Research and Innovation Center, Athens, Greece, Fall 2017 – to date.

Vice Chairman of the Department of Computer Science, University of Thessaly, Lamia, Greece, Fall 2018 – to date.

Deputy Director of Graduate Studies at the Electrical and Computer Engineering Department, University of Thessaly, Volos, Greece, Fall 2018 – to date.

Director of Graduate Studies at the Electrical and Computer Engineering Department, University of Thessaly, Volos, Greece, Fall 2015 – 2018.

Member of the Research Committee, University of Thessaly, Volos, Greece, Fall 2014 – 2017.

Skills   *Research:* In-depth knowledge of signal and image processing, speech recognition, language modeling, machine learning, multimedia signal processing, and computer vision. Eight years academic and twelve years of industrial research experience in the U.S., as well as ten years of experience in research centers and academia in Greece. Recent work focuses on audio-visual speech processing, far-field automatic speech recognition, multichannel acoustic scene analysis, and computer vision problems incorporating deep learning approaches in the framework of human-computer interaction and ambient intelligence applications.

Supervision: While at IBM, he has been manager of the Multimodal Conversational Technologies Department (2007 – 2008), having supervised over twenty researchers. Also, a team lead for audio-visual speech technologies research work (2003 – 2008). and a mentor and supervisor of nine summer students (2000 – 2008). While at NCSR "Demokritos", he has coordinated a four-member team as part of the "SYNC3" EU-funded project activities, and supervised one Ph.D. student (degree awarded) under a joint NCSR "Demokritos"/ University of Texas at Arlington Ph.D. program. As Associate Professor at the University of Thessaly, he has supervised / been supervising over twenty students on their Diploma Theses, five students on their Master Theses, and four (ongoing) Ph.D. students. Has also served in the Committees of ten awarded Ph.D.'s at various Universities in Greece, Scandinavia, U.S, and Australia. Additionally, has co-supervised a team of seven researchers as part of the "DIRHA" and "BabyRobot" EU projects.

Projects, Proposals, Collaborations: While at IBM, he has participated in various proposal writing efforts resulting in funding by DARPA, the European Union (FP6), internal IBM divisions, and external IBM customers, and he has successfully completed collaborations within multi-partner consortia including the "NETCARITY" and "DICIT" FP6 projects, as well as the "CHIL" FP6 integrated project, where he led the speech-technology work-package. As a Research Director at NCSR "Demokritos" he has successfully concluded the Technical Coordination of the "SYNC3" FP7 project, has been the beneficiary researcher of recently completed PEOPLE-IRG project "AVISPIRE", and has also contributed to the "INDIGO" FP6 project. While at the University of Thessaly, in collaboration with the Athena Research and Innovation Center in Greece, he has lead work-package tasks of the recently completed "DIRHA" FP7 and "BabyRobot" H2020 projects and is Technical Coordinator of the newly awarded SL-ReDu National Greek project. Additional current and past collaborations with faculty at U.S. and European Universities and Research Centers within the framework of research proposals.

Teaching: Instructor and teaching assistant at the Johns Hopkins University (1990, 1994); short-course and tutorial speaker (ELSNET 2001; ICIP 2003; ISCA SPCC 2014); has given invited lectures at Columbia University as part of a course on "*Automatic Speech Recognition*" (S'2004; F'2005); has taught "*Signals and Systems*" at the National Technical University of Athens as a part-time instructor and at the University of Thessaly (S'2011, F'2011, F'2012, S'2014, S'2015, S'2016, S'2017, S'2018, S'2019), "*Speech Processing*" at the latter (F'2011, F'2012, F'2013, F'2014, F'2015, F'2016, F'2017, F'2018, F'2019), "*Digital Signal Processing*" (S'2012, F'2012, F'2013, F'2014, F'2015, F'2016, F'2017, F'2018, F'2019), "*Advanced Topics in Signal Processing*" (S'2012), "*Pattern Recognition*" (S'2013, S'2014, S'2015, S'2016, S'2017, S'2018, S'2019), and "*Computer Vision*" (S'2013, S'2014).

Computer Skills: C, C++, PERL, Python, Matlab, Octave, HTK, Kaldi, and Torch7.

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

Honors

Recipient of *Marie-Curie International Reintegration Grant* for Project "AVISPIRE" (Audio-Visual Speech Processing in Realistic Environments), Oct. 2009 – Feb. 2013.

Elected *Member* of the *IEEE Speech and Language Technical Committee* (2009 – 2011).

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).

Memberships

Member of the International Institute of Electrical and Electronics Engineers (IEEE).
Member of the International Speech Communication Association (ISCA).
Member of the European Association for Signal Processing (EURASIP).
Member of the National Technical Chamber of Greece (TEE).

Personal

Born in 1965.
Marital Status: Married.
Hobbies: Sailing, Sea Kayaking, Hiking.

Journal Articles

(——) N. Efthymiou, P. P. Filntisis, P. Koutras, A. Tsiami, J. Hadfield, **G. Potamianos**, and P. Maragos, "ChildBot: Multi-robot perception and interaction with children," (Under Review:) *International Journal of Social Robotics* (Submitted in Mar. 2019; Revised in Oct. 2019).

(J12) S. Thermos, G. T. Papadopoulos, P. Darras, and **G. Potamianos**, "Deep sensorimotor learning for RGB-D object recognition," *Computer Vision and Image Understanding*, vol. 190, 2020.

(J11) P. Giannoulis, **G. Potamianos**, and P. Maragos, "Room-localized speech activity detection in multi-microphone smart homes," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 15, pp. 1–23, 2019.

(J10) P. P. Filntisis, N. Efthymiou, P. Koutras, **G. Potamianos**, and P. Maragos, "Fusing body posture with facial expressions for joint recognition of affect in child-robot interaction," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4011–4018, 2019.

(J09) I. Rodomagoulakis, A. Katsamanis, **G. Potamianos**, P. Giannoulis, A. Tsiami, and P. Maragos, "Room-localized spoken command recognition in multi-room, multi-microphone environments," *Computer Speech and Language*, vol. 46, pp. 419–443, 2017.

(J08) 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," *Journal of Language Resources and Evaluation*, vol. 41, pp. 389–407, 2007.

(J07) 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.

(J06) 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.

(J05) **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.

(J04) **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.

(J03) **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.

(J02) **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.

(J01) **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

(B09) **G. Potamianos**, E. Marcheret, Y. Mroueh, V. Goel, A. Koumbaroulis, A. Vartholomaios, and S. Thermos, "Audio and visual modality combination in speech processing applications", *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*, S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, G. Potamianos, and A. Krüger (Eds.), Ch. 12, pp. 489–543, ACM Books / Morgan-Claypool Publishers, San Rafael, CA, 2017.

(B08) **G. Potamianos**, C. Neti, J. Luettin, and I. Matthews, "Audio-visual automatic speech recognition: An overview," *Audio-Visual Speech Processing*, E. Vatikiotis-Bateson, G. Bailly, and P. Perrier (Eds.), Ch. 9, Cambridge University Press, 2012.

(B07) K. Kumar, **G. Potamianos**, J. Navratil, E. Marcheret, and V. Libal, "Audio-visual speech synchrony detection by a family of bimodal linear prediction models," *Multibiometrics for Human Identification*, B. Bhanu and V. Govindaraju (Eds.), Ch. 2, pp. 31–50, Cambridge University Press, 2011.

(B06) A. Waibel, R. Stiefelhagen, R. Carlson, J. Casas, J. Kleindienst, L. Lamel, O. Lanz, D. Mostefa, M. Omologo, F. Pianesi, L. Polymenakos, **G. Potamianos**, J. Soldatos, G. Sutschet, and J. Terken, "Computers in the human interaction loop," *Handbook of Ambient Intelligence and Smart Environments*, H. Nakashima, H. Aghajan, and J.C. Augusto (Eds.), Part IX, pp. 1071–1116, Springer, 2010.

(B05) P. Lucey, **G. Potamianos**, and S. Sridharan, "Visual speech recognition across multiple views," *Visual Speech Recognition: Lip Segmentation and Mapping*, A. Wee-Chung Liew and S. Wang (Eds.), ch. X, pp. 294–325, Information Science Reference (IGI), 2009.

(B04) **G. Potamianos**, L. Lamel, M. Wölfel, J. Huang, E. Marcheret, C. Barras, X. Zhu, J. McDonough, J. Hernando, D. Macho, and C. Nadeu, "Automatic speech recognition," *Computers in the Human Interaction Loop*, A. Waibel and R. Stiefelhagen (Eds.), ch. 6, pp. 43–59, Springer, 2009.

(B03) K. Bernardin, R. Stiefelhagen, A. Pnevmatikakis, O. Lanz, A. Brutti, J.R. Casas, and **G. Potamianos**, "Person tracking," *Computers in the Human Interaction Loop*, A. Waibel and R. Stiefelhagen (Eds.), ch. 3, pp. 11–22, Springer, 2009.

(B02) **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.

(B01) 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. Bovic (Ed.), ch. 10.8, pp. 1263–1289, Elsevier Academic Press, Burlington, MA, ISBN: 0-12-119792-1, 2005.

Edited Books

(EB03) S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, **G. Potamianos**, and A. Krüger (Eds.), *The Handbook of Multimodal-Multisensor Interfaces, Volume 3: Language Processing, Software and Commercialization, Emerging Directions*, ACM Books / Morgan-Claypool Publishers, San Rafael, CA, 2019.

(EB02) S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, **G. Potamianos**, and A. Krüger (Eds.), *The Handbook of Multimodal-Multisensor Interfaces, Volume 2: Signal Processing, Architectures, and Detection of Emotion and Cognition*, ACM Books / Morgan-Claypool Publishers, San Rafael, CA, 2018.

(EB01) S. Oviatt, B. Schuller, P. Cohen, D. Sonntag, **G. Potamianos**, and A. Krüger (Eds.), *The Handbook of Multimodal-Multisensor Interfaces, Volume 1: Foundations, User Modeling, and Common Modality Combinations*, ACM Books / Morgan-Claypool Publishers, San Rafael, CA, 2017.

Editorials in Journal Special Issues

(EJ03) S.-H. G. Chan, J. Li, P. Frossard, and **G. Potamianos**, "Special section on interactive multimedia," *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 841–843, 2011.

(EJ02) H. Meng, S. Oviatt, **G. Potamianos**, and G. Rigoll, "Introduction to the special issue on multimodal processing in speech-based interactions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 3, pp. 409–410, 2009.

(EJ01) C. Neti, **G. Potamianos**, J. Luettin, and E. Vatikiotis-Bateson, "Editorial of the special issue on joint audio-visual speech processing," *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 11, 2002.

Publications in Conference and Workshop Proceedings

(——) A. Koumparoulis, **G. Potamianos**, S. Thomas, and E. da Silva Morais, "Audio-assisted image inpainting for talking faces," (Submitted to:) *Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020.

(——) K. Papadimitriou and **G. Potamianos**, "A fully convolutional sequence learning approach for recognizing cued speech and fingerspelling in videos," (Submitted to:) *Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020.

(——) S. Thermos, P. Darras, and **G. Potamianos**, "A deep learning approach to object affordance segmentation," (Submitted to:) *Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Barcelona, Spain, 2020.

(C115) S. P. Chytas and **G. Potamianos**, "Hierarchical detection of sound events and their localization using convolutional neural networks with adaptive thresholds," *Proc. Detection and Classification of Acoustic Scenes and Events Workshop (DCASE)*, pp. 50–54, New York, NY, 2019.

(C114) A. Koumparoulis and **G. Potamianos**, "MobiLipNet: Resource-efficient deep learning based lipreading," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 2763–2767, Graz, Austria, 2019.

(C113) K. Papadimitriou and **G. Potamianos**, "End-to-end convolutional sequence learning for ASL fingerspelling recognition," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 2315–2319, Graz, Austria, 2019.

(C112) K. Papadimitriou and **G. Potamianos**, "Fingerspelled alphabet sign recognition in upper-body videos," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, La Coruna, Spain, 2019.

(C111) A. Koumparoulis and **G. Potamianos**, "Deep view2view mapping for view-invariant lipreading," *Proc. IEEE Spoken Language Technology Works. (SLT)*, pp. 588–594, Athens, Greece, 2018.

(C110) K. Papadimitriou and **G. Potamianos**, "A hybrid approach to hand detection and type classification in upper-body videos," *Proc. Europ. Work. Visual Information Process. (EUVIP)*, Tampere, Finland, 2018.

(C109) J. Hadfield, P. Koutras, N. Efthymiou, **G. Potamianos**, C. S. Tzafestas, and P. Maragos, "Object assembly guidance in child-robot interaction using RGB-D based 3D tracking," *Proc Int. Conf. Intell. Robots Systems (IROS)*, pp. 347–354, Madrid, Spain, 2018.

(C108) P. Giannoulis, **G. Potamianos**, and P. Maragos, "Multi-channel non-negative matrix factorization for overlapped acoustic event detection," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, pp. 857–861, Rome, Italy, 2018.

(C107) N. Efthymiou, P. Koutras, P. P. Filntisis, **G. Potamianos**, and P. Maragos, "Multi-view fusion for action recognition in child-robot interaction," *Proc. Int. Conf. Image Process. (ICIP)*, pp. 455–459, Athens, Greece, 2018.

(C106) S. Thermos, G. T. Papadopoulos, P. Darras, and **G. Potamianos**, "Attention-enhanced sensorimotor object recognition," *Proc. Int. Conf. Image Process. (ICIP)*, pp. 336–340, Athens, Greece, 2018.

(C105) A. Tsiami, P. Koutras, N. Efthymiou, P. P. Filntisis, **G. Potamianos**, and P. Maragos, "Multi3: Multi-sensory perception system for multi-modal child interaction with multiple robots," *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, Brisbane, Australia, 2018.

(C104) A. Tsiami, P. P. Filntisis, N. Efthymiou, P. Koutras, **G. Potamianos**, and P. Maragos, "Far-field audio-visual scene perception of multi-party human-robot interaction for children and adults," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 6568–6572, Calgary, Canada, 2018.

(C103) A. Koumparoulis, **G. Potamianos**, Y. Mroueh, and S. J. Rennie, "Exploring ROI size in deep learning based lipreading," *Proc. Int. Conf. on Auditory-Visual Speech Process. (AVSP)*, pp. 64–69, Stockholm, Sweden, 2017.

(C102) P. Giannoulis, **G. Potamianos**, and P. Maragos, "On the joint use of NMF and classification for overlapping acoustic event detection," *Proc. Int. Works. on Computational Intelligence for Multimedia Understanding (IWCIM)*, Held in Conjunction with *Europ. Conf. Signal Process. (EUSIPCO)*, Kos Island, Greece, 2017.

(C101) S. Thermos, G. T. Papadopoulos, P. Darras, and **G. Potamianos**, "Deep affordance-grounded sensorimotor object recognition," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 49–57, Honolulu, HI, 2017.

(C100) S. Thermos and **G. Potamianos**, "Audio-visual speech activity detection in a two-speaker scenario incorporating depth information from a profile or frontal view," *Proc. IEEE Spoken Language Technology Works. (SLT)*, pp. 579–584, San Diego, CA, 2016.

(C099) P. Giannoulis, **G. Potamianos**, P. Maragos, and A. Katsamanis, "Improved dictionary selection and detection schemes in sparse-CNMF-based overlapping acoustic event detection," *Proc. Detection and Classification of Acoustic Scenes and Events Works. (DCASE)*, pp. 25–29, Budapest, Hungary, 2016.

(demo) A. Tsiami, A. Katsamanis, I. Rodomagoulakis, **G. Potamianos**, and P. Maragos, "Sweet home listen: A distant speech recognition system for home automation control," *"Show and Tell" Session Demonstration*, *Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Shanghai, China, 2016.

(C098) E. Marcheret, **G. Potamianos**, J. Vopicka, and V. Goel, "Scattering vs. discrete cosine transform features in visual speech processing," *Proc. Int. Joint Conf. on Facial Analysis, Animation and Auditory-Visual Speech Process. (FAAVSP)*, pp. 175–180, Vienna, Austria, 2015.

(C097) E. Marcheret, **G. Potamianos**, J. Vopicka, and V. Goel, "Detecting audio-visual synchrony using deep neural networks," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 548–552, Dresden, Germany, 2015.

(C096) P. Giannoulis, A. Brutti, M. Matassoni, A. Abad, A. Katsamanis, M. Matos, **G. Potamianos**, and P. Maragos, "Multi-room speech activity detection using a distributed microphone network in domestic environments," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, pp. 1271–1275, Nice, France, 2015.

(C095) Z. I. Skordilis, A. Tsiami, P. Maragos, **G. Potamianos**, L. Spelgatti, and R. Sannino, "Multichannel speech enhancement using MEMS microphones," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 2729–2733, Brisbane,

Australia, 2015.

(C094) G. Floros, K. Kyritsis, and **G. Potamianos**, "Database and baseline system for detecting degraded traffic signs in urban environments," *Proc. Europ. Work. Visual Information Process. (EUVIP)*, Paris, France, 2014.

(C093) A. Tsiami, I. Rodomagoulakis, P. Giannoulis, A. Katsamanis, **G. Potamianos**, and P. Maragos, "ATHENA: A Greek multi-sensory database for home automation control," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 1608–1612, Singapore, 2014.

(C092) P. Giannoulis, **G. Potamianos**, A. Katsamanis, and P. Maragos, "Multi-microphone fusion for detection of speech and acoustic events in smart spaces," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, pp. 2375–2379, Lisbon, Portugal, 2014.

(C091) A. Tsiami, A. Katsamanis, P. Maragos, and **G. Potamianos**, "Experiments in acoustic source localization using sparse arrays in adverse indoors environments, *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, pp. 2390–2394, Lisbon, Portugal, 2014.

(C090) P. Giannoulis, A. Tsiami, I. Rodomagoulakis, A. Katsamanis, **G. Potamianos**, and P. Maragos, "The Athena-RC system for speech activity detection and speaker localization in the DIRHA smart home," *Proc. Joint Work. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, pp. 167–171, Nancy, France, 2014.

(C089) A. Katsamanis, I. Rodomagoulakis, **G. Potamianos**, P. Maragos, and A. Tsiami, "Robust far-field spoken command recognition for home automation combining adaptation and multichannel processing," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 5547–5551, Florence, Italy, 2014.

(C088) I. Rodomagoulakis, **G. Potamianos**, and P. Maragos, "Advances in large vocabulary continuous speech recognition in Greek: Modeling and nonlinear features," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, Marrakech, Morocco, 2013.

(C087) I. Rodomagoulakis, P. Giannoulis, Z.-I. Skordilis, P. Maragos, and **G. Potamianos**, "Experiments on far-field multichannel speech processing in smart homes," *Proc. Int. Conf. Digital Signal Process. (DSP)*, pp. 1–6, Santorini, Greece, 2013.

(C086) G. Galatas, **G. Potamianos**, and F. Makedon, "Robust multi-modal speech recognition in two languages utilizing video and distance information from the Kinect," *Proc. Int. Conf. Human-Computer Interaction (HCII)*, Las Vegas, NV, 2013.

(C085) G. Galatas, **G. Potamianos**, and F. Makedon, "Audio-visual speech recognition incorporating facial depth information captured by the Kinect," *Proc. Europ. Conf. Signal Process. (EUSIPCO)*, pp. 2714–2717, Bucharest, Romania, 2012.

(C084) G. Galatas, **G. Potamianos**, and F. Makedon, "Audio-visual speech recognition using depth information from the Kinect in noisy video conditions," *Proc. Int. Conf. Pervasive Technologies Related to Assistive Environments (PETRA)*, Crete, Greece, 2012.

(C083) P. Giannoulis and **G. Potamianos**, "A hierarchical approach with feature selection for emotion recognition from speech," *Proc. Int. Conf. Language Resources and Evaluation (LREC)*, Istanbul, Turkey, 2012.

(C082) G. Galatas, **G. Potamianos**, D. Kosmopoulos, C. McMurrough, and F. Makedon, "Bilingual corpus for AVASR using multiple sensors and depth information," *Proc. Int. Conf. Auditory-Visual Speech Process. (AVSP)*, pp. 103–106, Volterra, Italy, 2011.

(C081) N. Sarris, **G. Potamianos**, J.-M. Renders, C. Grover, E. Karstens, L. Kallipolitis, V. Tountopoulos, G. Petasis, A. Krithara, M. Gallé, G. Jacquet, B. Alex, R. Tobin, and L. Bounegru, "A system for synergistically structuring news content from traditional media and the blogosphere," *Proc. E-Challenges Conference*, Florence, Italy, 2011.

(C080) G. Galatas, **G. Potamianos**, A. Papangelis, and F. Makedon, "Audio visual speech recognition in noisy visual environments," *Proc. Int. Conf. Pervasive Technologies Related to Assistive Environments (PETRA)*, Crete, Greece, 2011.

(C079) L.-H. Kim, M. Hasegawa-Johnson, **G. Potamianos**, and V. Libal, "Joint estimation of DOA and speech based on EM beamforming," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 121–124, Dallas, TX, 2010.

(C078) K. Kumar, J. Navratil, E. Marcheret, V. Libal, and **G. Potamianos**, "Robust audio-visual speech synchrony detection by generalized bimodal linear prediction," *Proc. Conf. Int. Speech Comm. Assoc. (Interspeech)*, pp. 2251–2254, Brighton, United Kingdom, 2009.

(C077) K. Kumar, J. Navratil, E. Marcheret, V. Libal, G. Ramaswamy, and **G. Potamianos**, "Audio-visual speech synchronization detection using a bimodal linear prediction model," *Proc. IEEE Comp. Soc. Works. Biometrics*, Held in Conjunction with *CVPR*, Miami Beach, FL, 2009.

(C076) V. Libal, B. Ramabhadran, N. Mana, F. Pianesi, P. Chippendale, O. Lanz, and **G. Potamianos**, "Multimodal classification of activities of daily living inside smart homes," *Int. Works. Ambient Assisted Living (IWAAL)*, LNCS vol. 5518, Part II, pp. 687–694, Salamanca, Spain, 2009.

(C075) X. Zhuang, J. Huang, **G. Potamianos**, and M. Hasegawa-Johnson, "Acoustic fall detection using Gaussian mixture models and GMM supervectors," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 69–72, Taipei, Taiwan, 2009.

(C074) J. Huang, X. Zhuang, V. Libal, and **G. Potamianos**, "Long-time span acoustic activity analysis from far-field sensors in smart homes," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 4173–4176, Taipei, Taiwan, 2009.

(C073) **G. Potamianos**, J. Huang, E. Marcheret, V. Libal, R. Balchandran, M. Epstein, L. Seredi, M. Labsky, L. Ures, M. Black, and P. Lucey, "Far-field multimodal speech processing and conversational interaction in smart spaces," *Proc. Joint Work. Hands-Free Speech Communication and Microphone Arrays (HSCMA)*, Trento, Italy, 2008.

(C072) A Leone, G. Diraco, C. Distante, P. Siciliano, M. Grassi, A. Lombardi, G. Rescio, P. Malcovati, M. Malfatti, L. Gonzo, V. Libal, J. Huang, and **G. Potamianos**, "A multi-sensor approach for people fall detection in home environment," *Proc. 10th Europ. Conf. Computer Vision (ECCV)*, Marseille, France, 2008.

(C071) P. Lucey, **G. Potamianos**, and S. Sridharan, "Patch-based analysis of visual speech from multiple views," *Proc. Int. Conf. Auditory-Visual Speech Process. (AVSP)*, pp. 69–73, Tangalooma, Australia, 2008.

(C070) R. Balchandran, M. Epstein, **G. Potamianos**, and L. Seredi, "A multi-modal spoken dialog system for interactive TV," *Proc. Int. Conf. Multimodal Interfaces (ICMI) – Demo Papers*, pp. 191–192, Chania, Greece, 2008.

(C069) M. Grassi, A. Lombardi, G. Rescio, P. Malcovati, A Leone, G. Diraco, C. Distante, P. Siciliano, M. Malfatti, L. Gonzo, V. Libal, J. Huang, and **G. Potamianos**, "A hardware-software framework for high-reliability people fall detection," *Proc. 7th IEEE Conf. on Sensors (SENSORS)*, pp. 1328–1331, Lecce, Italy, 2008.

(C068) A. Tyagi, J.W. Davis, and **G. Potamianos**, "Steepest descent for efficient covariance tracking," *Proc. IEEE Work. Motion and Video Computing (WMVC)*, Copper Mountain, Colorado, 2008.

(C067) J. Huang, E. Marcheret, K. Visweswariah, and **G. Potamianos**, "The IBM RT07 evaluation systems for speaker diarization on lecture meetings," in *Multimodal Technologies for Perception of Humans: International Evaluation Workshops CLEAR 2007 and RT 2007, Baltimore, Maryland, May 2007*, LNCS vol. 4625, pp. 497–508, Springer, Berlin, 2008.

(C066) J. Huang, E. Marcheret, K. Visweswariah, V. Libal, and **G. Potamianos**, "The IBM Rich Transcription Spring 2007 speech-to-text systems for lecture meetings," in *Multimodal Technologies for Perception of Humans: International Evaluation Workshops CLEAR 2007 and RT 2007, Baltimore, Maryland, May 2007*, LNCS vol. 4625, pp. 429–441, Springer, Berlin, 2008.

(C065) V. Libal, J. Connell, **G. Potamianos**, and E. Marcheret, "An embedded system for in-vehicle visual speech activity detection," *Proc. Int. Work. Multimedia Signal Process. (MMSP)*, pp. 255–258, Chania, Greece, 2007.

(C064) 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.

(C063) 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.

(C062) P. Lucey, **G. Potamianos**, and S. Sridharan, "Pose-invariant audio-visual automatic speech recognition," *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 176–180, Hilvarenbeek, The Netherlands, 2007.

(C061) 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.

(C060) 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.

(C059) 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.

(C058) **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.

(C057) 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.

(C056) P. Lucey and **G. Potamianos**, "Lipreading using profile versus frontal views," *Proc. IEEE Work. Multimedia Signal Process. (MMSP)*, pp. 24–28, Victoria, Canada, 2006.

(C055) 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.

(C054) **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.

(C053) 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.

(C052) 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.

(C051) 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.

(C050) **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.

(C049) 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.

(C048) 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.

(C047) 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.

(C046) J. Jiang, **G. Potamianos**, and G. Iyengar, "Improved face finding in visually challenging environments," *Proc. Int. Conf. Multimedia Expo. (ICME)*, Amsterdam, The Netherlands, 2005.

(C045) 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. (ICME)*, Amsterdam, The Netherlands, 2005.

(C044) P. Scanlon, **G. Potamianos**, V. Libal, and S.M. Chu, "Mutual information based visual feature selection for lipreading," *Proc. Int. Conf. Spoken Lang. Process. (ICSLP)*, Jeju Island, Korea, 2004.

(C043) 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. (ICSLP)*, Jeju Island, Korea, 2004.

(C042) **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. (ICASSP)*, vol. 3, pp. 777–780, Montreal, Canada, 2004.

(C041) 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. (ICASSP)*, vol. 5, pp. 873–876, Montreal, Canada, 2004.

(C040) 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. (ICME)*, Taipei, Taiwan, 2004.

(C039) **G. Potamianos**, C. Neti, and S. Deligne, "Joint audio-visual speech processing for recognition and enhancement," *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 95–104, St. Jorioz, France, 2003.

(C038) J. Huang, **G. Potamianos**, and C. Neti, "Improving audio-visual speech recognition with an infrared headset," *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 175–178, St. Jorioz, France, 2003.

(C037) **G. Potamianos** and C. Neti, "Audio-visual speech recognition in challenging environments," *Proc. Europ. Conf. Speech Comm. Technol. (Eurospeech)*, pp. 1293–1296, Geneva, Switzerland, 2003.

(C036) 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. (ICME)*, vol. II, pp. 469-472, Baltimore, MD, 2003.

(C035) 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. (ICME)*, pp. 9–12, Baltimore, MD, 2003.

(C034) 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. (ICASSP)*, vol. I, pp. 24–27, Hong Kong, China, 2003.

(C033) 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. (ICASSP)*, vol. V, pp. 712–715, Hong Kong, China, 2003.

(C032) 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. (ICSLP)*, vol. 3, pp. 1449–1452, Denver, CO, 2002.

(C031) R. Goecke, **G. Potamianos**, and C. Neti, "Noisy audio feature enhancement using audio-visual speech data," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 2025–2028, Orlando, FL, 2002.

(C030) 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. (ICASSP)*, pp. 853–856, Orlando, FL, 2002.

(C029) G. Gravier, **G. Potamianos**, and C. Neti, "Asynchrony modeling for audio-visual speech recognition," *Proc. Human Lang. Techn. Conf. (HLT)*, pp. 1–6, San Diego, CA, 2002.

(C028) **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. (Eurospeech)*, pp. 1027–1030, Aalborg, Denmark, 2001.

(C027) **G. Potamianos** and C. Neti, "Automatic speechreading of impaired speech," *Proc. Work. Audio-Visual Speech Process. (AVSP)*, pp. 177–182, Aalborg, Denmark, 2001.

(C026) **G. Potamianos** and C. Neti, "Improved ROI and within frame discriminant features for lipreading," *Proc. Int. Conf. Image Process. (ICIP)*, pp. 250–253, Thessaloniki, Greece, 2001.

(C025) C. Neti, **G. Potamianos**, J. Luettin, 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. (MMSP)*, pp. 619–624, Cannes, France, 2001.

(C024) 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. (MMSP)*, pp. 79–84, Cannes, France, 2001.

(C023) I. Matthews, **G. Potamianos**, C. Neti, and J. Luettin, "A comparison of model and transform-based visual features for audio-visual LVCSR," *Proc. Int. Conf. Multimedia Expo. (ICME)*, Tokyo, Japan, 2001.

(C022) **G. Potamianos**, J. Luettin, and C. Neti, "Hierarchical discriminant features for audio-visual LVCSR," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 1, pp. 165–168, Salt Lake City, UT, 2001.

(C021) J. Luettin, **G. Potamianos**, and C. Neti, "Asynchronous stream modeling for large-vocabulary audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 1, pp. 169–172, Salt Lake City, UT, 2001.

(C020) H. Glotin, D. Vergyri, C. Neti, **G. Potamianos**, and J. Luettin, "Weighting schemes for audio-visual fusion in speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 1, pp. 173–176, Salt Lake City, UT, 2001.

(C019) **G. Potamianos** and C. Neti, "Stream confidence estimation for audio-visual speech recognition," *Proc. Int. Conf. Spoken Language Process. (ICSLP)*, vol. III, pp. 746–749, Beijing, China, 2000.

(C018) 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. (ICSLP)*, vol III, pp. 11-14, Beijing, China, 2000.

(C017) **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. (ICME)*, vol. II, pp. 1097–1100, New York, NY, 2000.

(C016) E. Cosatto, **G. Potamianos**, and H.P. Graf, "Audio-visual unit selection for the synthesis of photo-realistic talking-heads," *Proc. IEEE Int. Conf. Multimedia Expo. (ICME)*, vol. II, pp. 619–622, New York, NY, 2000.

(C015) **G. Potamianos** and A. Potamianos, "Speaker adaptation for audio-visual automatic speech recognition," *Proc. Europ. Speech Comm. Technol. (Eurospeech)*, Budapest, Hungary, vol. 3, pp. 1291–1294, 1999.

(C014) **G. Potamianos** and H.P. Graf, "Linear discriminant analysis for speechreading," *Proc. IEEE Work. Multimedia Signal*

Process. (MMSP), Los Angeles, CA, pp. 221–226, 1998.

(C013) **G. Potamianos**, H.P. Graf, and E. Cosatto, "An image transform approach for HMM based automatic lipreading," *Proc. Int. Conf. Image Process. (ICIP)*, Chicago, IL, pp. 173–177, 1998.

(C012) **G. Potamianos** and H.P. Graf, "Discriminative training of HMM stream exponents for audio-visual speech recognition," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Seattle, WA, vol. 6, pp. 3733–3736, 1998.

(C011) 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.

(C010) **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. (AVSP)*, Rhodes, Greece, pp. 65–68, 1997.

(C009) 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. (ICSMC)*, Orlando, FL, pp. 2034–2039, 1997.

(C008) **G. Potamianos**, "Efficient Monte Carlo estimation of partition function ratios of Markov random field images," *Proc. Conf. Inform. Sci. Systems (CISS)*, Princeton, NJ, vol. II, pp. 1212–1215, 1996.

(C007) **G. Potamianos** and J. Goutsias, "A unified approach to Monte Carlo likelihood estimation of Gibbs random field images," *Proc. Conf. Inform. Sci. Systems (CISS)*, Princeton, NJ, vol. I, pp. 84–90, 1994.

(C006) **G. Potamianos** and J. Goutsias, "An analysis of Monte Carlo methods for likelihood estimation of Gibbsian images," *Proc. Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Minneapolis, MN, vol. V, pp. 519–522, 1993.

(C005) **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 (CISS)*, Princeton, vol. I, pp. 357–362, 1992.

(C004) **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. (ICASSP)*, Toronto, Canada, vol. 4, pp. 2325–2328, 1991.

(C003) **G. Potamianos** and J. Diamessis, "Frequency sampling design of 2-D IIR filters using continued fractions," *Proc. Int. Symp. Circuits Systems (ISCAS)*, New Orleans, LA, pp. 2454–2457, 1990.

(C002) J. Diamessis and **G. Potamianos**, "A novel method for designing IIR filters with nonuniform samples," *Proc. Conf. Inform. Sci. Systems (CISS)*, Princeton, NJ, vol. 1, pp. 192–195, 1990.

(C001) J. Diamessis and **G. Potamianos**, "Modeling unequally spaced 2-D discrete signals by rational functions," *Proc. Int. Symp. Circuits Systems (ISCAS)*, 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

(T08) P. P. Filntisis, N. Efthymiou, P. Koutras, **G. Potamianos**, and P. Maragos, "Fusing body posture with facial expressions for joint recognition of affect in child-robot interaction," *Computing Research Repository*, arXiv:1901.01805, 2019.

(T07) 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.

(T06) **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.

(T05) **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.

(T04) **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.

(T03) **G. Potamianos**, A. Gatherer, and G. Wyman, "Coaxial cable channel impulse response characterization," *Technical Activity Report*, Integrated Systems Laboratory, Texas Instruments, Dallas, 1994.

(T02) **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.

(T01) **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.

U.S. Patents

(P07) S.M. Chu, V. Goel, E. Marcheret, and **G. Potamianos**, *Method for Likelihood Computation in Multi-Stream HMM Based Speech Recognition*, Patent No.: US007480617B2, Jan. 20, 2009.

(P06) S. Deligne, C.V. Neti, and **G. Potamianos**, *Audio-Visual Codebook Dependent Cepstral Normalization*, Patent No.: US007319955B2, Jan. 15, 2008.

(P05) 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.

(P04) 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.

(P03) 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.

(P02) 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.

(P01) E. Cosatto, H.P. Graf, and **G. Potamianos**, *Robust Multi-Modal Method for Recognizing Objects*, Patent No.: US006118887A, Sep. 12, 2000.