

Athanasios Katsamanis

Parmenidou 28
Athens 11636, Greece

nkatsam@sipi.usc.edu
<http://sipi.usc.edu/~nkatsam>

Research Interests

Multimodal Behavioral Analysis and Signal Processing Saliency detection in multimodal behavioral observations. Statistical audiovisual speech synthesis and recognition. Context modeling for emotion recognition. Statistical multimodal interaction and integration. Audiovisual speech inversion exploiting multimodal articulatory data. Multimodal human computer interfaces. Multimodal fusion for sign language recognition. Dynamic Bayesian networks for multimodal fusion.

Speech Production and Synthesis Speech aeroacoustics and articulatory synthesis. Modeling the production of fricatives. Computational speech modeling for acoustic-to-articulatory inversion. Statistical articulatory modeling based on real data. Articulatory feature extraction from real data. Improved acoustic modeling for speech. Statistical speech synthesis. Biomedical applications and second language tutoring.

Speech Recognition Exploiting articulatory knowledge for speech recognition using dynamic Bayesian networks. Statistical speaker articulation adaptation. Dynamic speech modeling for speech enhancement and recognition. Applications for large vocabulary continuous speech recognition.

Education

PhD in Electrical and Computer Engineering 2003-2009
National Technical University of Athens, School of E.C.E Athens, Greece
Subject: Non-linear computational speech modeling exploiting aerodynamics in the vocal tract
Advisor: Prof. Petros Maragos (URL: <http://cvsp.cs.ntua.gr/maragos>)

Diploma/M.Eng. in Electrical and Computer Engineering 1998 - 2003
National Technical University of Athens, School of E.C.E Athens, Greece
Final Grade: 9.49/10 (highest honors)
Diploma Thesis: Statistical Multiband Demodulation and Speech Analysis Using Kalman Filters and Energy Related Methods (Advisor: Prof. Petros Maragos)

Research Experience

Postdoctoral Research Associate 2009-2012
University of Southern California, School of Electrical Engineering Los Angeles, California, USA

As a member of the Signal Analysis and Interpretation Laboratory I have participated in the following projects:

Museum of Science: NSF-funded project on the development of interactive virtual guides at the museum of science in Boston.

Speech production and analysis: NIH-funded project on speech production modeling using real-time Magnetic Resonance Imaging of the vocal tract.

CreativeIT: NSF-funded project to study and develop computational models of human improvisation.

RATS: DARPA-funded project on Robust Automatic Transcription of Speech.

DCAPS: DARPA-funded project on Detection and Computational Analysis of Psychological Signals.

Graduate Research Assistant 2004-2009
National Technical University of Athens, School of E.C.E. *Athens, Greece*

As a member of the Computer Vision, Speech Communication & Signal Processing (CVSP) Group, I participated in European and national research projects:

ΠΕΝΕΔ: Research on computational speech production modeling based on aeroacoustics.

ASPI (Audiovisual to Articulatory Speech Inversion): Research on statistical audiovisual to articulatory speech inversion exploiting multimodal articulatory data.

MUSCLE (Multimedia Understanding via Semantics, Computation & Learning): Research on multimodal fusion for audiovisual speech recognition and event detection in audio streams.

HIWIRE (Human Input that Works in Real Environments): Research on speaker normalization and adaptation for user-robust continuous speech recognition.

Protagoras: Research on speech demodulation and formant tracking.

Visiting Researcher summer 2007
École Nationale Supérieure des Télécommunications *Paris, France*

Project: Improved acoustic modeling for speech inversion.

Mentor: Prof. Shinji Maeda.

Research Assistant July 2000-May 2002
Institute for Language and Speech Processing *Athens, Greece*

Designed and implemented an educational environment in MATLAB for the “Signals and Systems” course taught at the E.C.E. School at NTUA (Mentor: Prof. G. Carayiannis.)

Worked on a web interface for text-to-speech synthesis (Mentor: Dr. Ath. Protopapas)

Participated in the design and preparatory group for the European research project **METIS** on machine translation based on monolingual corpora. (Mentor: Dr. G. Tampouratzis)

Research Assistant summer 2002
Hong Kong Polytechnic University, School of Computer Science *Hong Kong, China*

In the frame of the student exchange program IAESTE I worked under the supervision of Assoc. Prof. Robert Luk on the design and implementation of an automatic Cantonese speech recognizer using the Hidden Markov Model Toolkit (HTK). Worked on duration modeling and on general system optimization.

Awards-Scholarships

PhD Studies Scholarship 2005-2009
Onassis Public Benefit Foundation

PhD Studies Scholarship 2006-2008
Greek National Scholarship Foundation

Research Assistantship Scholarship 2000-2002
Institute for Language and Speech Processing

Papakyriakopoulou and Kritikou Awards for excellent performance in Mathematics 2000
National Technical University of Athens

Award for excellent performance at the E.C.E. School 2000
Technical Chamber of Greece

Undergraduate Studies Scholarship 1998-1999
Greek National Scholarship Foundation

Teaching Experience

Guest Lecturer

2010, 2011

University of Southern California, School of E.E.

Los Angeles, U.S.A.

Gave three invited lectures for the following graduate-level classes:

EE 519 Speech Recognition and Processing for Multimedia (Fall Semester, 2010),

Lecture on Short-Time Fourier Transform

Lecture on Speech Enhancement

EE 619 Advanced Subjects in Speech Recognition (Spring Semester, 2011),

Lecture on Noise-Robust Speech Recognition

Graduate Teaching Assistant

2004-2008

National Technical University of Athens, School of E.C.E

Athens, Greece

Served as a lab and teaching assistant (lab preparation, presentation and grading) for the following courses, instructed by Prof. Petros Maragos:

Pattern Recognition with Emphasis on Speech Recognition (9th Semester, 2004-2006),

Signals and Systems (4th Semester, 2007),

Digital Signal Processing (6th Semester, 2008).

Academic Activities

Reviewer

IEEE Transactions on Audio, Speech and Language Processing

IEEE Signal Processing Letters

Journal of the Acoustical Society of America

IEEE Transactions on Affective Computing

Speech Communication

Journal of Engineering in Medicine

Interspeech

International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Invited Talks/ Posters/ Seminars

- [1] **Prosodic characterization of reading styles using audiobook corpora**
at the 161st Meeting of the Acoustical Society of America, San Diego, 2011.
- [2] **Multipulse articulatory modeling in the Wisconsin X-ray microbeam speech production database**
at the meeting of the Acoustical Society of America in Seattle, May 2011.
- [3] **Tongue Tracking in Ultrasound Images with Active Appearance Models**
at Ultrafest, New Haven, 2010.
- [4] **Audiovisual Speech: Recognition and Inversion**
at GIPSA-lab in Grenoble, France, Apr. 2009.
- [5] **Audio-Visual Speech Analysis and Recognition**
in MUSCLE joint with VITALAS Conference, Cannes, France, Feb. 2008.

- [6] **Fricative synthesis investigations using the transmission line matrix method**
at the meeting of the Acoustical Society of America in Paris, July 2008.
- [7] **Towards Automatic Speech Recognition in Adverse Environments**
in the Nonlinear Speech Processing Workshop (WNSP05), Heraklion, Greece, 2005.
- [8] **Statistical Speech Analysis and Nonlinear Modeling**
in COST277 Seminar in Limerick, 2004.
- [9] **Speech Technologies for E-Commerce and Transactional Services**
in EFTEHNOS series of seminars on supportive information and telecommunication technologies for people with special needs, University of Athens, 2004.

Open-Source Research Software

SailAlign http://sipi.usc.edu/~nkatsam/nassos/sail_align.html
Long Speech-Text Alignment Software *First release: January 2011*

SailAlign is an open-source software toolkit for robust long speech-text alignment implementing an adaptive, iterative speech recognition and text alignment scheme that allows for the processing of very long (and possibly noisy) audio and is robust to transcription errors. It is mainly written as a perl library and it depends upon freely available software and libraries where necessary, namely HTK, srilm, the Edinburgh Speech Tools Library and sclite.

Professional Experience in Industry

KEPYES, Hellenic Army <http://www.army.gr>
Software Engineer *June 2012-February 2013*

In the frame of my compulsory military service for the army, I designed, developed and tested a school information system for the management of the school of nursing officers of the hellenic army.

Qualia S.A. <http://www.qualia.gr>
Software Engineer *June 2006-June 2007*

Worked on the design and development of a broadcast news automatic speech recognition system integrated in a multimedia information retrieval platform.

Voice-In S.A. <http://www.voice-in.gr>
Developer *October 2002-July 2003*

Designed and implemented an Interactive Voice Responce (IVR) application optionally integrating automatic speech recognition and speech synthesis technologies.

Professional Memberships

Institute of Electrical and Electronics Engineers <i>Signal Processing Society</i>	2003-present
Acoustical Society of America <i>Member of the Speech Technical Committee</i>	2010-present
Technical Chamber of Greece	2004-present

Software Development, System & Network Administration Experience

Programming Languages C/C++, Perl, Python, Ruby, MATLAB, Java

Operating Systems Linux, Mac, Windows

Contributed in the design, development and administration of the CVSP group website and the CVSP network and computer system infrastructure (<http://cvsp.cs.ntua.gr/>).

Languages

- **English:** Proficient level (Cambridge Proficiency, CEFR level C2)
- **German:** Intermediate level (Mittelstufe, CEFR level B2)
- **French:** Basic level (CEFR level A2)
- **Greek:** Native

References

References available upon request

List of Publications

Journal Papers

- [1] A. Metallinou, A. Katsamanis, and S. Narayanan,
Tracking continuous emotional trends of participants during affective dyadic interactions using body language and speech information,
Image and Vision Computing (IMAVIS), Special Issue on Affect Analysis in Continuous Input, vol. 3(1): 137-152, Feb. 2013.
- [2] C.-C. Lee, A. Katsamanis, M. Black, B. Baucom, A. Christensen, P. Georgiou, and S. Narayanan,
Computing Vocal Entrainment: A Signal-derived PCA-based Quantification with Application to Affect Recognition in Married Couples' Interactions,
Computer Speech and Language, 2013.
- [3] M. Black, A. Katsamanis, B. R. Baucom, C.-C. Lee, A. C. Lammert, A. Christensen, P. G. Georgiou, and S. Narayanan
Toward Automating a Human Behavioral Coding System for Married Couples Interactions using Acoustic Features,
Speech Communication, vol. 55(1): 1-21, 2013.
- [4] A. Metallinou, M. Wöllmer, A. Katsamanis, F. Eyben, B. Schuller, and S. Narayanan,
Context-Sensitive Learning for Enhanced Audiovisual Emotion Classification,
IEEE Trans. on Affective Computing, vol. 3(2): 184-198, 2012.
- [5] A. Katsamanis, G. Papandreou, and P. Maragos,
Face Active Appearance Modeling and Speech Acoustic Information to Recover Articulation,
IEEE Trans. on Audio, Speech and Language Processing, vol. 17(3): 411-422, Mar. 2009.

- [6] G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos,
Adaptive Multimodal Fusion by Uncertainty Compensation with Application to Audio-Visual Speech Recognition,
IEEE Trans. on Audio, Speech and Language Processing, vol. 17(3): 423-435, Mar. 2009.

Papers in International Refereed Conferences

- [1] C.C. Lee, A. Katsamanis, B. Baucom, P. Georgiou, and S. Narayanan,
Using Measures of Vocal Entrainment to Inform Outcome-Related Behaviors in Marital Conflicts,
Proc. of the Asia-Pacific Signal and Information Processing Association Conference (APSIPA), 2012.
- [2] D. Traum, P. Aggarwal, R. Artstein, S. Foutz, J. Gerten, A. Katsamanis, D. Noren, W. Swartout,
Ada and Grace: Direct Interaction with Museum Visitors,
Proc. of Intelligent Virtual Agents Conference (IVA), 2012.
- [3] P. Aggarwal, R. Artstein, J. Gerten, A. Katsamanis, S. Narayanan, A. Nazarian, D. Traum,
The Twins corpus of museum visitor questions,
Proc. of the Language Resources and Evaluation Conference (LREC), 2012.
- [4] C.C. Lee, A. Katsamanis, P. Georgiou, and S. Narayanan,
Based on Isolated Saliency or Causal Integration? Toward a Better Understanding of Human Annotation Process using Multiple Instance Learning and Sequential Probability Ratio Test,
Proc. Int'l Conf. on Speech Communication and Technology, 2012.
- [5] A. Metallinou, A. Katsamanis and S. Narayanan,
A hierarchical framework for modeling multimodality and emotional evolution in affective dialogs,
Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing (ICASSP), 2012.
- [6] M. Wöllmer, A. Metallinou, A. Katsamanis, B. Schuller and S. Narayanan,
Analyzing the memory of BLSTM neural networks for enhanced emotion classification in dyadic spoken interactions,
Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing (ICASSP), 2012.
- [7] T. Chaspari, E. Mower Provost, A. Katsamanis, S. Narayaan,
An acoustic analysis of shared enjoyment in ECA interactions of children with autism,
Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing (ICASSP), 2012.
- [8] A. Katsamanis, J. Gibson, M. Black, and S. Narayanan,
Multiple instance learning for classification of human behavior observations,
Proc. of International Conference on Affective Computing and Intelligent Interactions (ACII), 2011.
- [9] C.-C. Lee, A. Katsamanis, Black M., P. Georgiou, and S. Narayanan,
Affective state recognition in married couples' interactions using PCA-based vocal entrainment measures with multiple instance learning,
Proc. of International Conference on Affective Computing and Intelligent Interactions (ACII), 2011.
- [10] M. Black, P. Georgiou, A. Katsamanis, B. Baucom, and S. Narayanan,
"You made me do it": Classification of blame in married couples' interaction by fusing automatically derived speech and language information,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.

- [11] J. Gibson, A. Katsamanis, M. Black, and S. Narayanan,
Automatic identification of salient acoustic instances in couples' behavioral interactions using diverse density support vector machines,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [12] A. Katsamanis, E. Bresch, V. Ramanarayanan, and S. Narayanan,
Validating rt-MRI based articulatory representations via articulatory recognition,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [13] A. Lammert, M. Proctor, A. Katsamanis, and S. Narayanan,
Morphological variation in the adult vocal tract: A modeling study of its potential acoustic impact,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [14] C.-C. Lee, A. Katsamanis, M. Black, B. Baucom, P. Georgiou, and S. Narayanan,
An analysis of pca-based vocal entrainment measures in married couples' affective spoken interactions,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [15] S. Narayanan, E. Bresch, P. Ghosh, L. Goldstein, A. Katsamanis, Y. Kim, A. Lammert, M. Proctor, V. Ramanarayanan, and Y. Zhu,
A multimodal real-time mri articulatory corpus for speech research,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [16] M. Proctor, A. Lammert, A. Katsamanis, L. Goldstein, C. Hagedorn, and S. Narayanan,
Direct estimation of articulatory kinematics from real-time magnetic resonance image sequences,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [17] V. Ramanarayanan, A. Katsamanis, and S. Narayanan,
Automatic data-driven learning of articulatory primitives from real-time MRI data using convolutive NMF with sparseness constraints,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [18] B. Xiao, V. Rozgic, A. Katsamanis, B. Baucom, P. Georgiou, and S. Narayanan,
Acoustic and visual cues of turn-taking dynamics in dyadic interactions,
Proc. Int'l Conf. on Speech Communication and Technology, 2011.
- [19] A. Metallinou, A. Katsamanis, Y. Wang and S. S. Narayanan,
Tracking Changes in Continuous Emotion States using Body Language and Prosodic Cues, Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing, 2011.
- [20] V. Rozgic, B. Xiao, A. Katsamanis, B. Baucom, P. G. Georgiou and S. S. Narayanan,
Estimation of ordinal approach-avoidance labels in dyadic interaction: ordinal logistic regression approach,
Proc. IEEE Int'l Conf. Acous., Speech, and Signal Processing, 2011.
- [21] A. Katsamanis, M. P. Black, P. G. Georgiou, L. Goldstein and S. S. Narayanan,
SailAlign: Robust long speech-text alignment,
Proc. of Workshop on New Tools and Methods for Very-Large Scale Phonetics Research, 2011.
- [22] M. Black, A. Katsamanis, C.-C. Lee, A. Lammert, B. Baucom, A. Christensen, P. G. Georgiou and S. S. Narayanan,
Automatic Classification of Married Couples' Behavior using Audio Features,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 27-30, Tokyo, Japan, Interspeech 2010.
- [23] E. Bresch, A. Katsamanis, L. Goldstein and S. S. Narayanan,
Statistical multi-stream modeling of real-time MRI articulatory speech data,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 27-30, Tokyo, Japan, Interspeech 2010.
- [24] V. Rozgic, B. Xiao, A. Katsamanis, B. Baucom, P. G. Georgiou and Shrikanth S. Narayanan,

- A new multichannel multimodal dyadic interaction database**,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 27-30, Tokyo, Japan, Interspeech 2010.
- [25] C.-C. Lee, M. Black, A. Katsamanis, A. Lammert, B. Baucom, A. Christensen, P. G. Georgiou and S. S. Narayanan,
Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 27-30, Tokyo, Japan, Interspeech 2010.
- [26] M. I. Proctor, D. Bone, A. Katsamanis and S. S. Narayanan,
Rapid Semi-automatic Segmentation of Real-time Magnetic Resonance Images for Parametric Vocal Tract Analysis,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 27-30, Tokyo, Japan, Interspeech 2010.
- [27] A. Roussos, A. Katsamanis, P. Maragos, **Tongue Tracking in Ultrasound Images with Active Appearance Models**,
Proc. IEEE Int'l Conf. on Image Processing (ICIP-09), Cairo, Egypt, Nov. 7-11, 2009.
- [28] S. Theodorakis, A. Katsamanis, P. Maragos,
Product-HMMs for automatic sign language recognition,
Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP-2009), Taipei, Taiwan, Apr. 2009.
- [29] A. Katsamanis, T. Roussos, P. Maragos, M. Aron and M.-O. Berger,
Inversion from Audiovisual Speech to Articulatory Information by Exploiting Multimodal Data,
Proc. Int'l Seminar on Speech Production (ISSP 2008), Strasbourg, France, Dec. 2008.
- [30] A. Katsamanis, G. Ananthakrishnan, G. Papandreou, P. Maragos, O. Engwall
Audiovisual Speech Inversion by Switching Dynamical Modeling Governed by a Hidden Markov Process,
Proc. European Signal Processing Conf. (EUSIPCO 2008), Lausanne, Switzerland, Aug. 2008.
- [31] A. Katsamanis, G. Papandreou, and P. Maragos,
Audiovisual-to-Articulatory Speech Inversion Using Active Appearance Models for the Face and Hidden Markov Models for the Dynamics,
Proc. IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, NV, U.S.A., Mar.-Apr. 2008.
- [32] S. Lefkimmatis, P. Maragos, A. Katsamanis,
Multisensor Multiband Cross-Energy Tracking for Feature Extraction and Recognition,
Proc. IEEE Int'l Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008), Las Vegas, NV, U.S.A., Mar.-Apr. 2008.
- [33] G. Papandreou, A. Katsamanis, V. Pitsikalis, and P. Maragos,
Multimodal Fusion and Learning with Uncertain Features Applied to Audiovisual Speech Recognition,
Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), pp. 264-267, Chania, Greece, Oct. 1-3, 2007.
- [34] A. Katsamanis, G. Papandreou, and P. Maragos,
Audiovisual-to-Articulatory Inversion Using Hidden Markov Models,
Proc. IEEE Workshop on Multimedia Signal Processing (MMSP-2007), pp. 457-460, Chania, Greece, Oct. 1-3, 2007.
- [35] A. Katsamanis, P. Tsiakoulis, P. Maragos, A. Potamianos,
Investigations in Articulatory Synthesis,
Proc. Int'l Congress of Phonetic Sciences, ICPhS 2007, Saarbruecken, Germany, Aug. 2007.

- [36] V. Pitsikalis, A. Katsamanis, G. Papandreou, and P. Maragos,
Adaptive Multimodal Fusion by Uncertainty Compensation,
Proc. Int'l Conf. on Spoken Language Processing, Sep. 17-21, Pittsburgh PA, USA,
Interspeech-ICSLP 2006.
- [37] A. Katsamanis, G. Papandreou, V. Pitsikalis, and P. Maragos,
**Multimodal Fusion by Adaptive Compensation for Feature Uncertainty with
Application to Audiovisual Speech Recognition**,
Proc. 14th European Signal Processing Conf. (EUSIPCO-2006), Florence, Italy, Sept. 4-8
2006.
- [38] A. Katsamanis and P. Maragos,
Advances in Statistical Estimation and Tracking of AM-FM Speech Components,
Proc. Interspeech 2005 - Eurospeech – 9th European Conf. on Speech Communication and
Technology, Lisbon, Portugal, Sept. 2005.

Book Chapters

- [1] G. Papandreou, A. Katsamanis, V. Pitsikalis and P. Maragos,
**Adaptive Multimodal Fusion by Uncertainty Compensation with Application to
Audio-Visual Speech Recognition**,
in Multimodal Processing and Interaction: Audio, Video, Text, edited by P. Maragos, A.
Potamianos, and P. Gros, Springer-Verlag, New York, 2008.
- [2] P. Maragos, P. Gros, A. Katsamanis and G. Papandreou,
Cross-Modal Integration for Performance Improving in Multimedia: A Review,
in Multimodal Processing and Interaction: Audio, Video, Text, edited by P. Maragos, A.
Potamianos, and P. Gros, Springer-Verlag, New York, 2008.