David Harwath

curriculum vitae

77 Massachusetts Avenue, 32-G438
Cambridge, MA 02139
USA

Email: harwath@cs.utexas.edu
Homepage: http://people.csail.mit.edu/dharwath

Citizenship: USA

Employment

The University of Texas at Austin 2020 - Present
Assistant Professor, Computer Science Department

Massachusetts Institute of Technology 2018 - 2020
Research Scientist, Computer Science and Artificial Intelligence Lab (CSAIL)

Google, Inc. Summer 2013
Intern, Speech Team

MIT Lincoln Laboratory Summer 2010, 2011, 2012
Intern, Human Language Technologies Group

Education

Massachusetts Institute of Technology
Ph.D., Computer Science, 2018
Thesis advisor: James Glass
Thesis committee: James Glass, Antonio Torralba, Victor Zue

Massachusetts Institute of Technology
S.M., Computer Science, 2013
Thesis advisor: James Glass

University of Illinois at Urbana-Champaign
B.S., Electrical Engineering, 2010
Thesis advisor: Mark Hasegawa-Johnson

Honors & awards

George M. Sprowls Award for best Ph.D. dissertation in computer science, MIT, 2018
Best paper nomination, ASRU 2015
E.C. Jordan Award, UIUC, 2010
Bitzer and Slottow Award, UIUC, 2009
1^{st} Place in the "Real World" category at Engineering Open House, UIUC, 2009
Tau Beta Pi, member
Eta Kappa Nu, member

Theses supervised

Andrew Roudichenko, M.Eng., Massachusetts Insitute of Technology, 2020 (expected) (co-advised with James Glass)
Angie Boggust, M.Eng., Massachusetts Insitute of Technology, 2020 (co-advised with James Glass)
Emmanuel Azuh, M.Eng., Massachusetts Insitute of Technology, 2019 (co-advised with James Glass)
Ken Leidal, M.Eng., Massachusetts Insitute of Technology, 2018 (co-advised with James Glass)

Teaching

Instruction

MIT 6.345: Automatic Speech Recognition, Teaching assistant, spring 2017
MIT 6.036: Introduction to Machine Learning, Teaching assistant, spring 2016

Invited lectures

MIT 6.345: Automatic Speech Recognition, spring 2019
MIT 6.UAR: Seminar in Undergraduate Research, spring 2019
MIT 6.864: Advanced Natural Language Processing, fall 2017

Refereed Journal Publications

[J2] David Harwath, Adrià Recasens, Dídac Surís, Galen Chuang, Antonio Torralba, and James Glass, "Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input," International Journal of Computer Vision, 2019. https://doi.org/10.1007/s11263-019-01205-0

[J1] Stephen H. Shum, David Harwath, Najim Dehak, and James R. Glass, "On the Use of Acoustic Unit Discovery for Language Recognition," IEEE/ACM Transactions on Audio, Speech, and Language Processing, September 2016, Vol. 24, No. 9, pp. 1665-1676

Refereed Conference Proceedings

[C21] Yasunori Ohishi, Akisato Kimura, Takahito Kawanishi, Kunio Kashino, David Harwath and James Glass, "Pair Expansion for Learning Multilingual Semantic Embeddings using Disjoint Visually-grounded Speech Audio Datasets," to appear at Interspeech, 2020

[C20] Yasunori Ohishi, Akisato Kimura, Takahito Kawanishi, Kunio Kashino, David Harwath, James Glass, "Trilingual Semantic Embeddings of Visually Grounded Speech with Self-Attention Mechanisms," Proc. ICASSP, 2020

[C19] David Harwath, Wei-Ning Hsu, and James Glass, "Learning Hierarchical Discrete Linguistic Units from Visually-Grounded Speech," Proc. ICLR, 2020

[C18] Wei-Ning Hsu, David Harwath, and James Glass, "Transfer Learning from Audio-Visual Grounding to Speech Recognition," Proc. Interspeech, 2019

[C17] Emmanuel Azuh, David Harwath, and James Glass, "Towards Bilingual Lexicon Discovery From Visually Grounded Speech Audio," Proc. Interspeech, 2019

[C16] Dídac Surís, Adrià Recasens, David Bau, David Harwath, James Glass, and Antonio Torralba, "Learning Words by Drawing Images," Proc. CVPR, 2019

[C15] David Harwath and James Glass, "Towards Visually Grounded Sub-Word Speech Unit Discovery," Proc. ICASSP, 2019

[C14] Angie Boggust, Kartik Audhkhasi, Dhiraj Joshi, David Harwath, Samuel Thomas, Rogerio Feris, Danny Gutfreund, Yang Zhang, Antonio Torralba, Michael Picheny, James Glass, "Grounding Spoken Words in Unlabeled Video," Proc. CVPR Workshops, 2019

[C13] David Harwath, Adrià Recasens, Dídac Surís, Galen Chuang, Antonio Torralba, and James Glass, "Jointly Discovering Visual Objects and Spoken Words from Raw Sensory Input," Proc. ECCV, 2018 *(invited to appear in the IJCV special issue for the best papers of ECCV 2018)*

[C12] David Harwath, Galen Chuang, and James Glass, "Vision as an Interlingua: Learning Multilingual Semantic Embeddings of Untranscribed Speech," Proc. ICASSP 2018

[C11] Kenneth Leidal, David Harwath, and James Glass, "Learning Modality-Invariant Representations for Speech and Images," Proc. ASRU, 2017

[C10] David Harwath and James Glass, "Learning Word-Like Units from Joint Audio-Visual Analysis," Proc. ACL, 2017

[C9] David Harwath, Antonio Torralba, and James Glass, "Unsupervised Learning of Spoken Language with Visual Context," Proc. NeurIPS, 2016

[C8] Felix Sun, David Harwath, and James Glass, "Look, Listen, and Decode: Multimodal Speech Recognition with Images," Proc. SLT, 2016

[C7] David Harwath and James Glass, "Deep Multimodal Semantic Embeddings for Speech and Images," Proc. ASRU, 2015 *(nominated for best paper award)*

[C6] David Harwath and James R. Glass, "Speech Recognition Without a Lexicon - Bridging the Gap Between Graphemic and Phonetic Systems," Proc. Interspeech, 2014

[C5] David Harwath, Alexander Gruenstein, and Ian McGraw, "Choosing Useful Word Alternates for Automatic Speech Recognition Correction Interfaces," Proc. Interspeech, 2014

[C4] David Harwath, Timothy J. Hazen, and James Glass, "Zero Resource Spoken Audio Corpus Analysis," Proc. ICASSP, 2013

[C3] Aren Jansen, Emmanuel Dupoux, Sharon Goldwater, Mark Johnson, Sanjeev Khudanpur, Kenneth Church, Naomi Feldman, Hynek Hermansky, Florian Metze, Richard Rose, Mike Seltzer, Pascal Clark, Ian McGraw, Balakrishnan Varadarajan, Erin Bennett, Benjamin Borschinger, Justin Chiu, Ewan Dunbar, Abdellah Fourtassi, David Harwath, Chia-ying Lee, Keith Levin, Atta Norouzian, Vijay Peddinti, Rachael Richardson, Thomas Schatz, Samuel Thomas, "A Summary of the 2012 JHU CLSP Workshop on Zero Resource Speech Technologies and Models of Early Language Acquisition," Proc. ICASSP, 2013

[C2] David Harwath and Timothy J. Hazen, "Topic Identification Based Extrinsic Evaluation of Summarization Techniques Applied to Conversational Speech," Proc. ICASSP, 2012

[C1] David Harwath and Mark Hasegawa-Johnson, "Phonetic Landmark Detection for Automatic Language Identification," Proc. Speech Prosody, 2010

Patents

[P2] David Harwath and James Glass, "Joint Acoustic and Visual Processing," US Patent App. 15/623,682, 2018

[P1] Alexander Gruenstein, David Harwath, Ian McGraw, "Selecting Alternates for Speech Recognition," Us Patent App. 20150127346, 2015

Theses

[T3] David Harwath, "Learning Spoken Language Through Vision," Ph.D. Thesis, Massachusetts Institute of Technology, 2018

[T2] David Harwath, "Unsupervised Modeling of Latent Topics and Lexical Units in Speech Audio," S.M. Thesis, Massachusetts Institute of Technology, 2013

[T1] David Harwath, "Phonetic Landmark Detection for Automatic Language Identification," B.S. Thesis, University of Illinois at Urbana-Champaign, 2010

Invited Talks

CVPR Sight & Sound Workshop, June 2020
CVPR Workshop on Visual Learning with Limited Labels, June 2020
Google Research, April 2020
Facebook AI Research, April 2020
University of Texas at Austin, Computer Science Department, April 2020
University of Maryland at College Park, Computer Science Department, April 2020
University of Michigan - Ann Arbor, Computer Science and Engineering Department, March 2020
Johns Hopkins University, Electrical Engineering Department, November 2018
Toyota Technological Institute at Chicago, January 2020
Google Speech Team, July 2019
MIT Lincoln Laboratory, May 2019
Analog Devices, January 2019
University of Texas at Austin, Computer Science Department, November 2018
Boston Deep Learning Summit, May 2018
Speech and Audio in the Northeast (SANE), October 2012

Media Coverage

Ars Technica, "AI learns to decipher images based on spoken words—almost like a toddler," September 2018
MIT News, "Machine-learning system tackles speech and object recognition, all at once," September 2018
New Scientist, "Neural net learns words like a child, by looking and listening," February 2017
Fast Company, "AI For Matching Images With Spoken Word Gets A Boost From MIT," February 2017
MIT News, "Learning words from pictures," December 2016

Service

Organizing committee member for the Multimodal Video Analysis and Moments in Time workshop at ICCV 2019

Reviewer for IEEE Trans. ASLP, IEEE Trans. PAMI, CSL, JASA, ACM MM, ICCV, CVPR, NeurIPS, Interspeech, ICASSP, ACL BlackboxNLP, CoNLL, and EMNLP.

Last updated on September 2, 2020