Scott T. Wisdom, Ph.D.

Contact Information

E-mail: <first name>.thomas.<last name>@gmail.com   *Website:* stwisdom.github.io

Summary

I am passionate about machine learning, deep learning, statistical models, and signal processing for real-world problems, especially for separation, enhancement, detection, and classification of sound.

Education

University of Washington, Seattle, Washington, USA

Ph.D., Electrical Engineering, August 2017

- Dissertation Topic: "Improving and Unfolding Statistical Models of Nonstationary Signals"
- Summary: I proposed novel statistical models of complex-valued time-frequency representatinos of audio signals and created new principled deep and recurrent network architectures by unfolding statistical model inference algorithms. These approaches lead to improved performance on detection of speech and acoustic events, acoustic scene classification, pixel-by-pixel image classification, and audio source separation with single and multiple microphones.
- Committee: Les Atlas (co-advisor), James Pitton (co-advisor), John Hershey, Mari Ostendorf, Mike Seltzer, Marina Meila

University of Washington, Seattle, Washington, USA

M.S. Electrical Engineering, March 2014

- Thesis title: "Improved Statistical Signal Processing of Nonstationary Random Processes Using Time-Warping"
- Summary: I proposed a new method of processing nonstationary signals, where signals are time-warped to make them more stationary. My approach improved detection of sonar signals and enhancement of noisy and reverberated speech.
- Committee: Les Atlas (co-advisor), James Pitton (co-advisor), Rico Malvar

University of Colorado, Boulder, Colorado, USA

B.S. Electrical and Computer Engineering, May, 2010
B.A. English Literature, May, 2010

Professional and Academic Experience

Google, Cambridge, Massachusetts, USA

Senior Research Scientist   **July 2018 - present**

- Member of the Sound Understanding team within Google Research working on speech, audio, and audio-visual machine perception, with a focus on audio source separation.
- Pioneered multiple new tasks include universal sound separation (separating all sounds regardless of class), unsupervised sound separation (training separation models without isolated recordings of sounds), and unsupervised audio-visual separation (where input audio is accompanied by video).

Affectiva, Boston, Massachusetts, USA

Senior Speech Research Scientist   **September 2017 - June 2018**

- Formulated, implemented, and trained statistical models and deep networks for voice activity detection, multispeaker detection and speaker counting, and noise-robust speech emotion recognition.
- Supervised gathering and labeling of real-world audio data.
- Interfaced with data, product, engineering, marketing, and computer vision teams to define and implement requirements.

Interactive Systems Design Lab, University of Washington, Seattle, Washington, USA

Research Assistant   **June, 2012 - September 2017**

- Performed novel research in machine learning, deep learning, and statistical signal processing, especially applied to audio. Authored and co-authored over a dozen peer-reviewed papers.
- As most senior graduate student, mentored other graduate students in the lab and helped them develop their academic, technical, and professional skills.
- Maintained the lab Linux server for general-purpose GPU computing.

Jelinek Summer Workshop on Speech and Language, Seattle, Washington, USA

Graduate Student Team Member   **June, 2015 - August 2015**

- Member of the Far-Field Speech Recognition team, explored new methods to make speech recognition more robust to mismatched training data, especially in the presence of unknown noise and reverberation. Results are now published as a book.
- Created a new type of deep network architecture that directly processes complex-valued, frequency-domain, multi-microphone audio data for improved separation of speech sources.

Mitsubishi Electric Research Labs (MERL), Cambridge, Massachusetts, USA

Research Intern   **January, 2015 - March 2015**

- Worked with John Hershey, Jonathan Le Roux, and Shinji Watanabe on deep unfolding of generative speech models for multichannel source separation of reverberant and noisy speech.
- Derived and implemented a variety of existing and novel speech separation methods in Matlab.

Microsoft Research, Redmond, Washington, USA

Research Intern   **June, 2014 - September 2014**

- Worked with Jie Liu and the Sensing and Energy Research Group (SERG) on signal processing for wireless signal localization.
- Came up to to speed quickly on state-of-the-art methods for localization, designed a new method for indoor localization, and ran extensive experiments.

Sensor Systems Lab, University of Washington, Seattle, Washington, USA

Research Assistant   **June, 2011 - December, 2012**

- Worked with Josh Smith on the Wireless Resonant Energy Link (WREL) project, a wireless power system for moving devices.
- Helped design hardware, wrote embedded firmware in C, and authored a Python interface.
- Designed a robust tracking algorithm (now patented) to maintain maximum wireless energy transfer regardless of receivers position and orientation.

Fluke Networks, Everett, Washington, USA

Software Engineer in Advanced Engineering Program (AEP)   **July, 2010 - June, 2012**

- As part of AEP, a career acceleration program for high-potential recent graduates, worked on a wide variety of solo and team projects involving embedded systems, digital signal processing, wireless networking, and application development for innovative network test tools for copper, fiber, and wireless media.

SELECTED CONFERENCE PUBLICATIONS

E. Tzinis, S. Wisdom, T. Remez, J.R. Hershey, "AudioScopeV2: Audio-Visual Attention Architectures for Calibrated Open-Domain On-Screen Sound Separation", in *Proc. European Conference on Computer Vision (ECCV)*, Tel-Aviv, Israel, October 2022.

K. Patterson, K. Wilson, S. Wisdom, J.R. Hershey, "Distance-Based Sound Separation", in *Proc. Interspeech 2022*, Incheon, South Korea, September 2022.

K. Kilgour, B. Gfeller, Q. Huang, A. Jansen, <u>S. Wisdom</u>, M. Tagliasacchi, "Text-Driven Separation of Arbitrary Sounds", in *Proc. Interspeech 2022*, Incheon, South Korea, September 2022.

H. Muckenhirn, A. Safin, H. Erdogan, F. de Chaumont Quitry, M. Tagliasacchi, <u>S. Wisdom</u>, J. R. Hershey, "CycleGAN-Based Unpaired Speech Dereverberation", in *Proc. Interspeech 2022*, Incheon, South Korea, September 2022.

S. J. Yang, <u>S. Wisdom</u>, C. Gnegy, R. F. Lyon, S. Savla, "Listening with Googlears: Low-Latency Neural Multiframe Beamforming and Equalization for Hearing Aids", in *Proc. Interspeech 2022*, Incheon, South Korea, September 2022.

T. Denton, <u>S. Wisdom</u>, J. R. Hershey, "Improving Bird Classification with Unsupervised Sound Separation", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Singapore, May 2022.

A. Sivaraman, <u>S. Wisdom</u>, H. Erdogan, J. R. Hershey, "Adapting speech separation to real-world meetings using mixture invariant training", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Singapore, May 2022.

<u>S. Wisdom</u>, A. Jansen, R. J. Weiss, H. Erdogan, J. R. Hershey, "Sparse, Efficient, and Semantic Mixture Invariant Training: Taming In-the-Wild Unsupervised Sound Separation", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2021.

E. Fonseca, A. Jansen, D. P. W. Ellis, <u>S. Wisdom</u>, M. Tagliasacchi, J. R. Hershey, M. Plakal, S. Hershey, R. Channing Moore, X. Serra, "Self-supervised learning from automatically separated sound scenes", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2021. **(Best paper award)**

Y. Koizumi, S. Karita, <u>S. Wisdom</u>, H. Erdogan, J. R. Hershey, L. Jones, M. Bacchiani, "DF-Conformer: Integrated Architecture of Conv-TasNet and Conformer Using Linear Complexity Self-Attention for Speech Enhancement", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2021. **(Best paper award)**

<u>S. Wisdom</u>, H. Erdogan, D. P. W. Ellis, R. Serizel, N. Turpault, E. Fonseca, J. Salamon, P. Seetharaman, J. R. Hershey, "Whats All the FUSS About Free Universal Sound Separation Data?", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, Canada, June 2021.

S. Maiti, H. Erdogan, K. Wilson, <u>S. Wisdom</u>, S. Watanabe, J. R. Hershey, "End-To-End Diarization for Variable Number of Speakers with Local-Global Networks and Discriminative Speaker Embeddings", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Toronto, Canada, June 2021.

E. Tzinis, <u>S. Wisdom</u>, A. Jansen, S. Hershey, T. Remez, D.P.W. Ellis, J.R. Hershey, "Into the Wild with AudioScope: Unsupervised Audio-Visual Separation of On-Screen Sounds", in *International Conference on Learning Representations (ICLR)*, Vienna, Austria, May 2021.

Z. Q. Wang, H. Erdogan, <u>S. Wisdom</u>, K. Wilson, J. R. Hershey, "Sequential Multi-Frame Neural Beamforming for Speech Separation and Enhancement", in *Proc. IEEE Spoken Language Technology Workshop (SLT)*, virtual, January 2021.

<u>S. Wisdom</u>, E. Tzinis, H. Erdogan, R.J. Weiss, K. Wilson, J.R. Hershey, "Unsupervised Sound Sep-

aration Using Mixutre Invariant Training", in *Advances in Neural Information Processing Systems (NeurIPS)*, Vancouver, Canada, December 2020.

E. Tzinis, <u>S. Wisdom</u>, J.R. Hershey, A. Jansen, D.P.W. Ellis, "Improving Universal Sound Separation Using Sound Classification", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020.

S. Sonning, C. Schuldt, H. Erdogan, <u>S. Wisdom</u>, "Performance Study of a Convolutional Time-Domain Audio Separation Network for Real-Time Speech Denoising", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, May 2020.

I. Kavalerov, <u>S. Wisdom</u>, H. Erdogan, B. Patton, K. Wilson, J. Le Roux, and J.R. Hershey, "Universal Sound Separation", in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2019.

<u>S. Wisdom</u>, J.R. Hershey, K. Wilson, J. Thorpe, M. Chinen, B. Patton, and R.A. Saurous, "Differentiable Consistency Constraints for Improved Deep Speech Enhancement", *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

Le Roux, J., <u>S. Wisdom</u>, H. Erdogan, and J.R. Hershey, "SDR-half-baked or well done?", in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK, May 2019.

<u>S. Wisdom</u>, T. Powers, J. Pitton, and L. Atlas, Deep Recurrent NMF for Speech Separation by Unfolding Iterative Thresholding, in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2017.
(Best student paper award, WASPAA student travel grant awarded)

<u>S. Wisdom</u>, T. Powers, J. Pitton, and L. Atlas, Building Recurrent Networks by Unfolding Iterative Thresholding for Sequential Sparse Recovery, *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, Lousiana, USA, March 2017.

<u>S. Wisdom</u>, T. Powers, J.R. Hershey, J. Le Roux, and L. Atlas, "Full-Capacity Unitary Recurrent Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, Barcelona, Spain, Dec 2016. **(NeurIPS student travel grant awarded)**

<u>S. Wisdom</u>, J.R. Hershey, J. Le Roux, and S. Watanabe, "Deep Unfolding for Multichannel Source Separation," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Shanghai, China, March 2016. **(Travel grants awarded by NSF, IEEE SPS, UW)**

<u>S. Wisdom</u>, G. Okopal, L. Atlas, and J. Pitton, "Voice Activity Detection Using Subband Non-circularity," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brisbane, Austrailia, April 2015.

<u>S. Wisdom</u>, T. Powers, L. Atlas, and J. Pitton, "Enhancement of Reverberant and Noisy Speech by Extending Its Coherence," in *Proc. REVERB Challenge Workshop, Florence*, Italy, May 2014.

JOURNAL & BOOK PUBLICATIONS

J. R. Hershey, J. Le Roux, S. Watanabe, <u>S. Wisdom</u>, Z. Chen, Y. Isik, "Novel deep architectures in speech processing," in *New Era for Robust Speech Recognition: Exploiting Deep Learning*, S. Watanabe, M. Delcroix, F. Metze, J. R. Hershey, Eds. Springer, 2017.

G. Okopal, <u>S. Wisdom</u>, and L. Atlas "Speech Processing with the Strong Uncorrelating Transform," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Nov 2015.

A.P. Sample, B. Waters, <u>S. Wisdom</u>, and J.R. Smith, "Enabling Seamless Wireless Power Delivery in Dynamic Environments," in *Proc. IEEE*, vol. 101, no. 6, pp. 1343-1358, June 2013.

PATENTS E. Tzinis, <u>S. Wisdom</u>, A. Jansen, J. R. Hershey, "Audio-Visual Separation of On-Screen Sounds Based on Machine Learning Models."

A.P. Sample, J.R. Smith, B.H. Waters, <u>S. Wisdom</u>, "Enabling Seamless Wireless Power Delivery in Dynamic Environments."

TECHNICAL SKILLS
- Programming languages: Python, C/C++, bash scripting, Matlab.
- Other proficiencies: Keras, TensorFlow, Linux, Git, embedded systems.
- Algorithms: Experience deriving and programming implementations of deep learning, inference algorithms for probabilistic graphical models (exact and variational), speech processing, convex optimization, beamforming, statistical estimation and detection.

ACADEMIC SERVICE
- Elected member, IEEE Audio and Acoustic Signal Processing (AASP) Technical Committee (2019-2022).
- Area chair for ICASSP 2022, WASPAA 2021, ICASSP 2021.
- Session chair at Interspeech 2022, ICASSP 2022, WASPAA 2021, ICASSP 2021, ICASSP 2020, ICASSP 2019.
- Co-organizer for Detection and Classification of Acoustic Scenes (DCASE) Challenge on Sound Event Detection and Separation in Domestic Environments (2020, 2021).
- Reviewer for ICASSP, WASPAA, NeurIPS, ICLR, ICML, MLSP, CHI, IEEE/ACM Transactions on Signal Processing, IEEE/ACM Transactions on Audio, Speech, and Language Processing, IEEE Signal Processing Letters, Journal of the Acoustical Society of America, Computer Speech and Language, Neural Computation, EURASIP Journal on Advances in Signal Processing, Speech Communication, GlobalSIP, UbiComp.

INVITED TALKS
Speech and Audio in the Northeast (SANE) Workshop, Cambridge, Massachusetts   October 2022
"Unsupervised Sound Separation: A New Paradigm for Audio Processing"

Dolby SoHo, New York City, New York   April 2019
"Advances in Enhancing Noisy Speech, and Beyond"

Mitsubishi Electric Research Labs (MERL), Cambridge, Massachusetts   March, 2018
"Theory and Applications of Sparse Model-Based Recurrent Neural Networks"

Google, Cambridge, Massachusetts   November, 2017
"Unitary Recurrent Neural Networks and Deep Recurrent Nonnegative Matrix Factorization"

Pacific Northwest National Laboratory, Seattle, Washington   March, 2017
"Unfolding Statistical Inference Algorithms into Better Deep Networks"

Microsoft Research, Redmond, Washington   April, 2016
"Improving and Unfolding Statistical Models for Multichannel Audio"

Mitsubishi Electric Research Labs (MERL), Cambridge, Massachusetts   January, 2016
"Unfolding Generative Models for Multichannel Source Separation"

Amazon, Seattle, Washington   June, 2015
"Noncircularity: A New Feature for Speech Processing"

Mitsubishi Electric Research Labs (MERL), Cambridge, Massachusetts   September, 2014

"Beyond Locally Stationary and Circular"

NATO Centre for Maritime Research & Experimentation (CMRE), La Spezia, Italy   May, 2014
"Improved Statistical Signal Processing by Extending Coherence Time"