Grant P. Strimel

Curriculum Vitae

Research & Engineering Interests

I am a Principal Scientist for Alexa Speech Recognition and Deep Learning groups at Amazon. I joined Alexa in 2018 to build the Alexa Pittsburgh organization, which has now grown to over fifty scientists where our mission is to deliver customer-delighting natural language processing experiences through both edge-first and cloud-centric solutions. My primary focus for Alexa has been on low-latency, real-time ML design, and I frequently work cross-discipline with many talented software engineers, product managers, language engineers, hardware architects, legal teams and executive leadership. Generally, my interests lie in the computational aspects of Machine Learning and Artificial Intelligence broadly construed: online learning, distributed learning, multi-task learning, edge machine learning, model compression, efficient data structures for ML and AI under resource constraints, beyond worst case analysis of algorithms, speech and spoken language understanding.

Employment

2018 - Present **Principal Scientist**, Amazon Alexa, Pittsburgh, PA.
- Building the Alexa Pittsburgh organization (from 5 to 50+ ML scientists and engineers).
- Delivering science for local voice control - enabling Alexa inference on device.
- Developing new modeling techniques for AI on the edge.
- End-to-end neural methods for speech recognition, natural language understanding and text-to-speech.

2014 - 2017 **Machine Learning Scientist**, Amazon.com, Seattle, WA.
- Research Lead - Amazon Consumer Payments.
- Advanced ML platform for advertisement targeting and recommendations.
- Focused on various theoretical and applied questions related to sampling, online matching, policy optimization and fairness.
- Large-scale computational methods for model building.

2013 - 2015 **Chief Scientist and Co-founder**, SCHAI LLC, Pittsburgh, PA.
- Built proprietary automated trading firm.
- Developed financial prediction models on state of the art AI and machine learning methods.
- Advanced a fully automated trading system that runs 24/5.
- Communicated technical efforts and results to investors and non-domain audience.

Summer 2013 **Software Development Internship**, Amazon.com, Seattle, WA.
- Project: "Genetic Banners and Advertisements".
- Investigated genetic algorithms and online machine learning models and their interactions for content optimization and advertisement targeting.
- Increased German division clicks by >7% over three month period.

Summer 2011 **Software Development Internship**, ANSYS Inc., Pittsburgh, PA.
- Used OpenCL to parallelize algorithms across GPUs used for scene graph processing.
- Implemented custom data structures similar to R-trees and Kd-trees for engineering applications.

Research Experience

2014 - Present **Machine Learning Scientist**, *Amazon.com*.

2013 **CORAL Artificial Intelligence Group**, *Carnegie Mellon University*.
- Advisor: Professor Manuela M. Veloso.
- Multiagent planning and learning with finite resources.
- Learning practical reachability maps for service robots in dynamic environments.
- Automatic online calibration of multiple RGBD Kinect devices using SLAM.

Fall 2013 **Machine Learning Theory**, *Carnegie Mellon University*.
- Advisor: Professor Avrim Blum.
- Worked to relax assumptions and extend previous results for use of ontology structures and unlabeled data for learning.
- Semi-supervised, multi-task learning theory.

Summer 2012 **CORAL Artificial Intelligence Group**, *Carnegie Mellon University*.
- Advisor: Professor Manuela M. Veloso.
- Multimodel person recognition systems for mobile robot services.

Education

2017 - 2018 **Doctoral Studies in Computer Science (DNF)**,
GPA 4.0, left for opportunity to build Amazon Alexa in Pittsburgh.
Carnegie Mellon University, Pittsburgh, PA.

2013 - 2014 **M.S. in Computer Science**,
Carnegie Mellon University, Pittsburgh, PA.
Advisor: Manuela M. Veloso.

2010 - 2013 **B.S. in Computer Science**,
Carnegie Mellon University, Pittsburgh, PA.
Minor in Robotics.
University Honors.

Thesis

Title *Map Learning and Coverage Planning for Robots in Large Unknown Environments.*

Supervisors Herbert A. Simon University Professor Manuela M. Veloso (chair) & Professor Avrim Blum.

Description This masters thesis presented methods for robotic mapping of large structured spaces and coverage planning under finite energy/fuel constraints. The research addressed both the theoretical and practical questions involved and techniques were evaluated on real service robots. (pdf)

Refereed Conference Proceedings

1. **Compute Cost Amortized Transformer for Streaming ASR.** Yi Xie, Jonathan Macoskey, Martin Radfar, Feng-Ju Chang, Brian King, Ariya Rastrow, Athanasios Mouchtaris, Grant P. Strimel. The 23rd Annual Conference of the International Speech Communication Association (**Interspeech**) 2022. (pdf)

2. **ConvRNN-T: Convolutional Augmented Recurrent Neural Network Transducers for Streaming Speech Recognition.** Martin Radfar, Rohit Barnwal, Rupak Vignesh Swaminathan, Feng-Ju Chang, Grant P. Strimel, Nathan Susanj, Athanasios Mouchtaris. The 23rd Annual Conference of the International Speech Communication Association (**Interspeech**) 2022. (pdf)
3. **Latency Control for Keyword Spotting.** Christin Jose, Joseph Wang, Grant P. Strimel, Mohammad Omar Khursheed, Yuriy Mishchenko, Brian Kulis. The 23rd Annual Conference of the International Speech Communication Association (**Interspeech**) 2022. (pdf)
4. **Caching Networks: Capitalizing on Common Speech for ASR.** Anastasios Alexandridis*, Grant P. Strimel*, Ariya Rastrow, Pavel Kveton, Jon Webb, Maurizio Omologo, Siegfried Kunzmann, Athanasios Mouchtaris. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2022. (pdf)
5. **Contextual Adapters for Personalized Speech Recognition in Neural Transducers.** Kanthashree Mysore Sathyendra, Thejaswi Muniyappa, Feng-Ju Chang, Jing Liu, Jinru Su, Grant P. Strimel, Athanasios Mouchtaris, Siegfried Kunzmann. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2022. (pdf)
6. **TinyS2I: A Small-footprint Utterance Classification Model with Contextual Support for On-device SLU.** Anastasios Alexandridis, Kanthashree Mysore Sathyendra, Grant P. Strimel, Pavel Kveton, Jon Webb, Athanasios Mouchtaris. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2022. (pdf)
7. **Multi-Task RNN-T with Semantic Decoder for Streamable Spoken Language Understanding.** Xuandi Fu, Feng-Ju Chang, Martin Radfar, Kai Wei, Jing Liu, Grant P. Strimel, Kanthashree Mysore Sathyendra. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2022. (pdf)
8. **A Neural Prosody Encoder for End-to-End Dialogue Act Classification.** Kai Wei, Dillon Knox, Martin Radfar, Thanh Tran, Markus Muller, Grant P. Strimel, Nathan Susanj, Athanasios Mouchtaris, Maurizio Omologo. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2022. (pdf)
9. **Adaptive Global-Local Context Fusion for Multi-Turn Spoken Language Understanding.** Thanh Tran*, Kai Wei*, Weitong Ruan, Ross McGowan, Nathan Susanj, Grant P. Strimel. Thirty-Fourth Annual Conference on Innovative Applications of Artificial Intelligence (**IAAI**) 2022. (pdf)
10. **Attentive Contextual Carryover for Multi-Turn End-to-End Spoken Language Understanding.** Kai Wei*, Thanh Tran*, Feng-Ju Chang, Kanthashree Mysore Sathyendra, Thejaswi Muniyappa, Jing Liu, Anirudh Raju, Ross McGowan, Nathan Susanj, Ariya Rastrow, Grant P. Strimel. IEEE Automatic Speech Recognition and Understanding Workshop (**ASRU**) 2021. (pdf)
11. **Amortized Neural Networks for Low-Latency Speech Recognition.** Jonathan Macoskey*, Grant P. Strimel*, Jinru Su, Ariya Rastrow. The 22nd Annual Conference of the International Speech Communication Association (**Interspeech**) 2021. (pdf)
12. **Learning a Neural Diff for Speech Models.** Jonathan Macoskey*, Grant P. Strimel*, Ariya Rastrow. The 22nd Annual Conference of the International Speech Communication Association (**Interspeech**) 2021. (pdf)
13. **CoDERT: Distilling Encoder Representations with Co-learning for Transducer-based Speech Recognition.** Rupak Vignesh Swaminathan, Brian King, Grant P. Strimel, Jasha Droppo, Athanasios Mouchtaris. The 22nd Annual Conference of the International Speech Communication Association (**Interspeech**) 2021. (pdf)

14. **SmallER: Scaling Neural Entity Resolution for Edge Devices.** Ross McGowan, Jinru Su, Vince DiCocco, Thejaswi Muniyappa, Grant P. Strimel. The 22nd Annual Conference of the International Speech Communication Association (**Interspeech**) 2021. (pdf)
15. **Bifocal Neural ASR: Exploiting Keyword Spotting for Inference Optimization.** Jonathan Macoskey*, Grant P. Strimel*, Ariya Rastrow. IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP**) 2021. (pdf)
16. **Rescore in a Flash: Compact, Cache Efficient Hashing Data Structures for N-gram Language Models.** Grant P. Strimel, Ariya Rastrow, Gautam Tiwari, Adrien Pierard, Jon Webb. The 21st Annual Conference of the International Speech Communication Association (**Interspeech**) 2020. (pdf)
17. **Semantic Complexity in End-to-End Spoken Language Understanding.** Joseph P. McKenna*, Samridhi Choudhary*, Michael Saxon*, Grant P. Strimel, Athanasios Mouchtaris (**Interspeech**) 2020. (pdf)
18. **Statistical Model Compression for Small-Footprint Natural Language Understanding.** Grant P. Strimel, Kanthashree Mysore Sathyendra, Stan Peshterliev. The 19th Annual Conference of the International Speech Communication Association (**Interspeech**) 2018. (pdf)
19. **Coverage Planning with Finite Resources.** Grant P. Strimel and Manuela M. Veloso. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**) 2014. (pdf)
20. **Toward Interactive Grounded Language Acquisition.** Thomas Kollar, Jayant Krishnamurthy, Grant P. Strimel. Proceedings of Robotics: Science and Systems (**RSS**) 2013. (pdf)

Issued Patents

Compression of machine learned models. Grant Strimel and Sachin Grover. US10558738B1. 2020. (link)

Science Posts

Sept. 2021 *How to make on-device speech recognition practical* (link), Amazon Science Blogs (Top 10 science posting 2021).

Aug. 2018 *Shrinking machine learning models for offline use* (link), Amazon Science Blogs.

Invited Talks

Jan. 2020 *Beyond Big Data AI/ML Summit Panelist*, Pittsburgh Technical Council.

Jan. 2020 *TVR: Beyond Big Data - Amazon Alexa* (link), TechVibe Radio Podcast.

March 2018 *Hey, Alexa, could you win the Olympics? - NLU Inference on a Budget*, CP Tech Talks, Amazon.com.

Feb. 2016 *Traffic Scheduling Under Eligibility Constraints*, CP Tech Talks, Amazon.com.

Dec. 2015 *Machine Learning on Distributed Data*, CP Tech Talks, Amazon.com.

Teaching

2015 - 2019 **Olympic Games**, *Amazon.com.*

- Developed biannual (Summer and Winter) programming competition for engineering organization.
- Educational programming and theory puzzles.
- Autograded submissions and live leaderboard.
- Taught participants how to develop parallel algorithms using Java8 functionality, interface with optimization suites to solve LP's and mixed integer programs, write scalable algorithms to process large graphs, write distributed communication-efficient protocols, etc.

2015 - 2016 **Amazon Payment Products Learning Lecture Series**, *Amazon.com.*

- Began learning series for engineers and analysts interested in data science, ML and algorithms.
- Planned and gave 32 biweekly lectures, with course notes and recordings.

Spring 2012 **CS Teaching Assistant**, *Carnegie Mellon University.*

- CS 15-210: Parallel and Sequential Data Structures and Algorithms.
  - Taught recitation, wrote and graded homework and tests.
  - Tutoring.

Activities

- I am an avid golfer. I was a Varsity Letterman Golfer at Carnegie Mellon and continue to play competitively at the amateur level.
- I enjoy listening to and playing music. I listen better than I play, but I get out tunes on the piano, guitar, harmonica, and saxophone.
- Intramural softball, dodgeball and kickball.