

Wei-Ning Hsu

617-852-8026 | wnhsu@fb.com | New York, NY, USA

EXPERIENCE

- 8/2021 - present **Senior Research Scientist**
Facebook AI Research, Facebook, New York, NY, USA
- 8/2020 - 7/2021 **Research Scientist**
Facebook AI Research, Facebook, New York, NY, USA
- 6/2019 - 8/2019 **Research Intern**
Facebook AI Research, Facebook, New York, NY, USA
- 6/2018 - 11/2018 **Research Intern**
Google Brain, Google, Mountain View, CA, USA
- 7/2016 - 8/2016 **Research Intern**
Mitsubishi Electric Research Lab, Cambridge, MA, USA

EDUCATION

- 5/2020 Ph.D. in Computer Science (GPA: 5.0/5.0)
Massachusetts Institute of Technology (MIT), Cambridge, MA, USA
- 6/2018 S.M. in Computer Science (GPA: 5.0/5.0)
Massachusetts Institute of Technology (MIT), Cambridge, MA, USA
- 6/2014 B.S. in Electrical Engineering (GPA: 3.91/4.00)
National Taiwan University (NTU), Taipei, Taiwan

RESEARCH INTERESTS

Speech Processing: speech recognition, speech synthesis, multi-modal learning

Machine Learning: semi- and self-supervised learning, representation learning, deep generative modeling

AWARDS AND HONORS

- 2020 | **ICML 2020 Top Reviewers**
- 2019 | **NeurIPS 2019 Best Reviewers**
- 2015 | **Top Universities Strategic Alliance Fellow**
- 2014 | **Dean's Award**

SERVICE AND ACTIVITIES

Senior Program Committee:

AAAI 2022

Expert Reviewers:

ICML 2021

Program Committee:

NeurIPS 2019-2021, ICML 2019-2021, AAAI 2020-2021, ICLR 2021-2022, ICASSP 2021-2022, Interspeech 2021, ACM MM 2021, ACL 2021, ARR 2021, EMNLP 2021, COLING 2018-2020

Journal Reviewer:

IEEE SPS, IEEE TASLP, IEEE PAMI, IEEE Open Journal of Signal Processing, Neural Networks, Computer Speech & Language

President, Taiwanese Student Association, Massachusetts Institute of Technology, 2016-2017

President, Electrical Engineering Student Association, National Taiwan University, 2012-2013

SKILLS

- Programming: Python, MATLAB, Shell scripts, C++, Go
- Toolkit: Tensorflow, PyTorch, CNTK, Kaldi, HTK, Theano, LibSVM
- Language: Mandarin (native), English (fluent), Taiwanese (fluent), Japanese (elementary)

PUBLICATIONS

Google Scholar: <https://scholar.google.com/citations?user=N5HDmqoAAAAJ>

Total Citation: 1671 / h-index: 20 / i10-index: 31 (Accessed 31 Oct 2021)

Journal

- [1] **Wei-Ning Hsu**, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. "HuBERT: Self-Supervised Speech Representation Learning by Masked Prediction of Hidden Units." *In Transactions on Audio Speech and Language Processing (TASLP)*, 2021.
- [2] Kushal Lakhotia, Evgeny Kharitonov, **Wei-Ning Hsu**, Yossi Adi, Adam Polyak, Benjamin Bolte, Tu-Anh Nguyen et al. "Generative spoken language modeling from raw audio." *In Transactions of the Association for Computational Linguistics (ACL)*, 2021.

Peer Reviewed Conference

- [3] Alexei Baevski, **Wei-Ning Hsu**, Alexis Conneau, and Michael Auli. "Unsupervised Speech Recognition." *In Neural Information Processing Systems (NeurIPS)*, 2021.
- [4] Changhan Wang, **Wei-Ning Hsu**, Yossi Adi, Adam Polyak, Ann Lee, Peng-Jen Chen, Jiatao Gu, and Juan Pino. "fairseq S²: A Scalable and Integrable Speech Synthesis Toolkit." *In Empirical Methods in Natural Language Processing (EMNLP)*, 2021.
- [5] Vimal Manohar, Tatiana Likhomanenko, Qiantong Xu, **Wei-Ning Hsu**, Ronan Collobert, Yatharth Saraf, Geoffrey Zweig, and Abdelrahman Mohamed. "Kaizen: Continuously improving teacher using Exponential Moving Average for semi-supervised speech recognition." *In IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2021.
- [6] **Wei-Ning Hsu**, Yao-Hung Hubert Tsai, Benjamin Bolte, Ruslan Salakhutdinov, and Abdelrahman Mohamed. "HuBERT: How much can a bad teacher benefit ASR pre-training?." *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.
- [7] Adam Polyak, Yossi Adi, Jade Copet, Eugene Kharitonov, Kushal Lakhotia, **Wei-Ning Hsu**, Abdelrahman Mohamed, and Emmanuel Dupoux. "Speech Resynthesis from Discrete Disentangled Self-Supervised Representations." *In Interspeech*, 2021.
- [8] **Wei-Ning Hsu**, Anuroop Sriram, Alexei Baevski, Tatiana Likhomanenko, Qiantong Xu, Vineel Pratap, Jacob Kahn et al. "Robust wav2vec 2.0: Analyzing Domain Shift in Self-Supervised Pre-Training." *In Interspeech*, 2021.
- [9] **Wei-Ning Hsu**, Ann Lee, Gabriel Synnaeve, and Awni Hannun. "Semi-Supervised end-to-end Speech Recognition via Local Prior Matching." *In IEEE Spoken Language Technology Workshop (SLT)*, 2021.
- [10] **Wei-Ning Hsu**, David Harwath, Christopher Song, and James Glass. "Text-free image-to-speech synthesis using learned segmental units." *In Annual Meeting of the Association for Computational Linguistics (ACL)*, 2021.
- [11] Sameer Khurana, Antoine Laurent, **Wei-Ning Hsu**, Jan Chorowski, Adrian Lancucki, Ricard Marxer, and James Glass. "A convolutional deep markov model for unsupervised speech representation learning." *In Interspeech*, 2020.
- [12] David Harwath*, **Wei-Ning Hsu***, James Glass. Learning Hierarchical Discrete Linguistic Units from Visually-Grounded Speech. *International Conference on Learning Representations (ICLR)*, 2020
- [13] **Wei-Ning Hsu**, David Harwath, James Glass. Transfer Learning from Audio-Visual Grounding to Speech Recognition. *In Interspeech*, 2019
- [14] Yu-An Chung, **Wei-Ning Hsu**, Hao Tang, James Glass. An Unsupervised Autoregressive Model for Speech Representation Learning. *In Interspeech*, 2019
- [15] **Wei-Ning Hsu**, Yu Zhang, Ron J. Weiss, Heiga Zen, Yonghui Wu, Yuxuan Wang, Yuan Cao, Ye Jia, Zhifeng Chen, Jonathan Shen, Patrick Nguyen, Ruoming Pang. Hierarchical Generative Modeling for Controllable Speech Synthesis. *International Conference on Learning Representations (ICLR)*, 2019
- [16] **Wei-Ning Hsu**, Yu Zhang, Ron J. Weiss, Yu-An Chung, Yuxuan Wang, Yonghui Wu, James Glass. Disentangling Correlated Speaker and Noise for Speech Synthesis via Data Augmentation and Adversarial Factorization. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019
- [17] Yu-An Chung, Yuxuan Wang, **Wei-Ning Hsu**, Yu Zhang, RJ Skerry-Ryan. Semi-Supervised Training for Improving Data Efficiency in End-to-End Speech Synthesis. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019
- [18] Suwon Shon, **Wei-Ning Hsu**, James Glass. Unsupervised Representation Learning of Speech for Dialect Identification. *In IEEE Spoken Language Technology Workshop (SLT)*, 2018.
- [19] **Wei-Ning Hsu**, James Glass. Scalable Factorized Hierarchical Variational Autoencoder Training. *In Interspeech*, 2018
- [20] **Wei-Ning Hsu**, Hao Tang, James Glass. Unsupervised Adaptation with Interpretable Disentangled Representations for Distant Conversational Speech Recognition. *In Interspeech*, 2018
- [21] Hao Tang, **Wei-Ning Hsu**, Francois Grondin, James Glass. A Study of Enhancement, Augmentation, and Autoencoder Methods for Domain Adaptation in Distant Speech Recognition. *In Interspeech*, 2018
- [22] Siqi Zheng, Jianzong Wang, Jing Xiao, **Wei-Ning Hsu**, James Glass. A Noise-Robust Self-Adaptive Multitarget Speaker Detection System. *In International Conference on Pattern Recognition (ICPR)*, 2018
- [23] **Wei-Ning Hsu**, James Glass. Extracting Domain Invariant Features by Unsupervised Learning for Robust Automatic Speech Recognition. *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018
- [24] **Wei-Ning Hsu**, Yu Zhang, James Glass. Unsupervised Learning of Disentangled and Interpretable Latent Representations from Sequential Data. *In Neural Information Processing Systems (NeurIPS)*, 2017
- [25] **Wei-Ning Hsu**, Yu Zhang, James Glass. Unsupervised Domain Adaptation for Robust Speech Recognition via Variational Autoencoder-Based Data Augmentation. *In IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2017

- [26] Maryam Najafian, **Wei-Ning Hsu**, Ahmed Ali, James Glass. Automatic Speech Recognition of Arabic Multi-Genre Broadcast Media. In *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2017
- [27] **Wei-Ning Hsu**, Yu Zhang, James Glass. Learning Latent Representations for Speech Generation and Transformation. In *Interspeech*, 2017.
- [28] **Wei-Ning Hsu**, Yu Zhang, James Glass. A Prioritized Grid Long Short-Term Memory RNN for Speech Recognition. In *IEEE Spoken Language Technology Workshop (SLT)*, 2016.
- [29] Tuka AlHanai, **Wei-Ning Hsu**, and James Glass. Development of the MIT ASR System for the 2016 Arabic Multi-Genre Broadcast Challenge. In *IEEE Spoken Language Technology Workshop (SLT)*, 2016.
- [30] Salvatore Romeo, Giovanni Da San Martino, Alberto Barrón-Cedeño, Alessandro Moschitti, Yonatan Belinkov, **Wei-Ning Hsu**, Yu Zhang, Mitra Mohtarami and James Glass. Neural Attention for Learning to Rank Questions in Community Question Answering. In *International Conference on Computational Linguistics (COLING)*, 2016.
- [31] Hakan Erdogan, Tomoki Hayashi, John Hershey, Takaaki Hori, Chiori Hori, **Wei-Ning Hsu**, Suyoun Kim, Jonathan Le Roux, Zhong Meng and Shinji Watanabe. Multi-Channel Speech Recognition: LSTMs All the Way Through. In *The 4th CHiME Speech Separation and Recognition Challenge*, 2016
- [32] **Wei-Ning Hsu**, Yu Zhang, Ann Lee and James Glass. Exploiting Depth and Highway Connections in Convolutional Recurrent Deep Neural Networks for Speech Recognition. in *Interspeech*, 2016.
- [33] Mitra Mohtarami, Yonatan Belinkov, **Wei-Ning Hsu**, Yu Zhang, Tao Lei, Scott Cyphers and James Glass. SLS: Neural-based Approaches to Answer Selection and Question Retrieval in Community Question Answering Systems. In *International Workshop on Semantic Evaluation*, 2016
- [34] **Wei-Ning Hsu** and Hsuan-Tien Lin. Active Learning by Learning. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2015.
- [35] Cheng-Tao Chung, **Wei-Ning Hsu**, Cheng-Yi Lee, and Lin-Shan Lee. Enhancing Automatically Discovered Multi-Level Acoustic Patterns Considering Context Consistency with Applications in Spoken Term Detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

Preprint

- [36] Bowen Shi, **Wei-Ning Hsu**, Kushal Lakhotia, Abdelrahman Mohamed. "Learning Audio-Visual Speech Representation by Masked Multimodal Cluster Prediction." *submitted to International Conference on Learning Representations (ICLR)*, 2022.
- [37] Xutai Ma, Hongyu Gong, Danni Liu, Ann Lee, Yun Tang, Peng-Jen Chen, **Wei-Ning Hsu**, Kenneth Heafield, Phillip Koehn, Juan Pino. "Direct simultaneous speech to speech translation." *arXiv preprint arXiv:2110.08250*, 2021.
- [38] Eugene Kharitonov, Ann Lee, Adam Polyak, Yossi Adi, Jade Copet, Kushal Lakhotia, Tu-Anh Nguyen, Morgane Rivière, Abdelrahman Mohamed, Emmanuel Dupoux, **Wei-Ning Hsu**. "Text-Free Prosody-Aware Generative Spoken Language Modeling." *arXiv preprint arXiv:2109.03264*, 2021.
- [39] Ann Lee, Peng-Jen Chen, Changhan Wang, Jiatao Gu, Xutai Ma, Adam Polyak, Yossi Adi, Qing He, Yun Tang, Juan Pino, **Wei-Ning Hsu**. "Direct speech-to-speech translation with discrete units." *arXiv preprint arXiv:2107.05604*, 2021.
- [40] Awni Hannun, Vineel Pratap, Jacob Kahn, and **Wei-Ning Hsu**. "Differentiable weighted finite-state transducers." *arXiv preprint arXiv:2010.01003*, 2020.
- [41] Jonathan Shen, Patrick Nguyen, Yonghui Wu, Zhifeng Chen, Mia X. Chen, Ye Jia, Anjuli Kannan, Tara Sainath, Yuan Cao, Chung-Cheng Chiu, Yanzhang He, Jan Chorowski, Smit Hinsu, Stella Laurenzo, James Qin, Orhan Firat, Wolfgang Macherey, Suyog Gupta, Ankur Bapna, Shuyuan Zhang, Ruoming Pang, Ron J. Weiss, Rohit Prabhavalkar, Qiao Liang, Benoit Jacob, Bowen Liang, HyoukJoong Lee, Ciprian Chelba, Sébastien Jean, Bo Li, Melvin Johnson, Rohan Anil, Rajat Tibrewal, Xiaobing Liu, Akiko Eriguchi, Navdeep Jaitly, Naveen Ari, Colin Cherry, Parisa Haghani, Otavio Good, Youlong Cheng, Raziq Alvarez, Isaac Caswell, **Wei-Ning Hsu**, Zongheng Yang, Kuan-Chieh Wang, Ekaterina Gonina, Katrin Tomanek, Ben Vanik, Zelin Wu, Llion Jones, Mike Schuster, Yanping Huang, Dehao Chen, Kazuki Irie, George Foster, John Richardson, Klaus Macherey, Antoine Bruguier, Heiga Zen, Colin Raffel, Shankar Kumar, Kanishka Rao, David Rybach, Matthew Murray, Vijayaditya Peddinti, Maxim Krikun, Michiel A. U. Bacchiani, Thomas B. Jablin, Rob Suderman, Ian Williams, Benjamin Lee, Deepti Bhatia, Justin Carlson, Semih Yavuz, Yu Zhang, Ian McGraw, Max Galkin, Qi Ge, Golan Pundak, Chad Whipkey, Todd Wang, Uri Alon, Dmitry Lepikhin, Ye Tian, Sara Sabour, William Chan, Shubham Toshniwal, Baohua Liao, Michael Nirschl, Pat Rondon. "Lingvo: a modular and scalable framework for sequence-to-sequence modeling." *arXiv preprint arXiv:1902.08295*, 2019.
- [42] **Wei-Ning Hsu**, James Glass. Disentangling by Partitioning: A Representation Learning Framework for Multimodal Sensory Data. *arXiv preprint arXiv:1805.11264*, 2018.
- [43] **Wei-Ning Hsu**, Yu Zhang, James Glass. Recurrent Neural Network Encoder with Attention for Community Question Answering. *arXiv preprint arXiv:1603.07044* 2016.

Thesis

- [44] **Wei-Ning Hsu**. "Unsupervised learning of disentangled representations for speech with neural variational inference models." SM diss., Massachusetts Institute of Technology, 2018.
- [45] **Wei-Ning Hsu**. "Speech processing with less supervision: learning from weak labels and multiple modalities." PhD diss., Massachusetts Institute of Technology, 2020.