

DUC LE

CONTACT

Email: lehoangduc1990@gmail.com
LinkedIn: <https://www.linkedin.com/in/duchoangle/>
Google Scholar: <https://scholar.google.com/citations?user=Zorb5T4AAAAJ&hl=en>

WORK EXPERIENCE

TikTok - Music Machine Learning January 2023 - Present
Senior Staff Research Scientist San Jose, CA

Leading an applied research team working on generative AI, specializing in controllable end-to-end music generation and large-scale music understanding.

Meta - AI Speech January 2022 - January 2023
Research Scientist Manager Menlo Park, CA

Led an applied research team to push and productionize the state-of-the-art in end-to-end ASR, rare word and entity recognition, contextual biasing, language model fusion, large-scale multilingual speech recognition, and spoken language understanding.

Meta - AI Speech September 2017 - January 2022
Staff Research Scientist Menlo Park, CA

Tech Lead in the Meta AI Speech Team, specializing in end-to-end ASR and acoustic modeling. Developed the foundational ASR technology that is now deployed and used across the company and led multiple high-impact R&D efforts.

Meta - Language Technology Group Summer 2016 & 2015
Research Intern Menlo Park, CA

Worked on multi-task CTC acoustic model training, model compression, and i-vector based adaptation.

EDUCATION

University of Michigan, Ann Arbor, MI 2012 - 2017
Ph.D. in Computer Science
Dissertation: *Towards Automatic Speech-Language Assessment for Aphasia Rehabilitation*

University of Michigan, Ann Arbor, MI 2012 - 2014
M.S. in Computer Science

University of Texas - Dallas, Richardson, TX 2008 - 2012
B.S. in Computer Science, *summa cum laude*
Senior design topic: *Resource-Constrained Data Mining*

PUBLICATIONS

Music Generation and Understanding

1. Julian D. Parker, Janne Spijkervet, Katerina Kosta, Furkan Yesiler, Boris Kuznetsov, Ju-Chiang Wang, Matt Avent, Jitong Chen, and **Duc Le**. “StemGen: A music generation model that listens.” *ICASSP*. Seoul, South Korea. April, 2024. (*submitted*)
2. Minz Won, Yun-Ning Hung, and **Duc Le**. “A Foundation Model for Music Informatics.” *ICASSP*. Seoul, South Korea. April, 2024. (*submitted*)

3. Yun-Ning Hung, Ju-Chiang Wang, Minz Won, and **Duc Le**. “Scaling Up Music Information Retrieval Training with Semi-Supervised Learning.” *ICASSP*. Seoul, South Korea. April, 2024. (*submitted*)
4. Martin Kukla, Minz Won, Yun-Ning Hung, and **Duc Le**. “MusT3: Unified Multi-Task Model for Fine-Grained Music Understanding.” *NeurIPS Machine Learning for Audio Workshop*. New Orleans, USA. December, 2023.

Automatic Speech Recognition - Contextual Biasing and Language Model Fusion

1. Suyoun Kim, Ke Li, Lucas Kabela, Rongqing Huang, Jiedan Zhu, Ozlem Kalinli, and **Duc Le**. “Joint Audio/Text Training for Transformer Rescoring of Streaming Speech Recognition.” *Findings of EMNLP*. Abu Dhabi, UAE. December, 2022.
2. Antoine Bruguier, **Duc Le**, Rohit Prabhavalkar, Dangna Li, Zhe Liu, Bo Wang, Eun Chang, Fuchun Peng, Ozlem Kalinli, and Michael L. Seltzer. “Neural-FST Class Language Model for End-to-End Speech Recognition.” *ICASSP*. Singapore. May, 2022.
3. **Duc Le**, Mahaveer Jain, Gil Keren, Suyoun Kim, Yangyang Shi, Jay Mahadeokar, Julian Chan, Yuan Shangguan, Christian Fuegen, Ozlem Kalinli, Yatharth Saraf, and Michael L. Seltzer. “Contextualized Streaming End-to-End Speech Recognition with Trie-Based Deep Biasing and Shallow Fusion.” *INTER-SPEECH*. Brno, Czech Republic. September, 2021.
4. Suyoun Kim, Yuan Shangguan, Jay Mahadeokar, Antoine Bruguier, Christian Fuegen, Michael L. Seltzer, and **Duc Le**. “Improved Neural Language Model Fusion for Streaming Recurrent Neural Network Transducer.” *ICASSP*. Virtual. June, 2021.
5. **Duc Le**, Gil Keren, Julian Chan, Jay Mahadeokar, Christian Fuegen, and Michael L. Seltzer. “Deep Shallow Fusion for RNN-T Personalization.” *SLT*. January, 2021.
6. **Duc Le**, Thilo Koehler, Christian Fuegen, and Michael L. Seltzer. “G2G: TTS-Driven Pronunciation Learning for Graphemic Hybrid ASR.” *ICASSP*. Barcelona, Spain. May, 2020.

Automatic Speech Recognition - Novel Model Architecture

1. **Duc Le**, Frank Seide, Yuhao Wang, Yang Li, Kjell Schubert, Ozlem Kalinli, and Michael L. Seltzer. “Factorized Blank Thresholding for Improved Runtime Efficiency of Neural Transducers.” *ICASSP*. Rhodes Island, Greece. June, 2023.
2. Ke Li, Jay Mahadeokar, Jinxi Guo, Yangyang Shi, Gil Keren, Ozlem Kalinli, Michael L. Seltzer, and **Duc Le**. “Improving Fast-Slow Encoder based Transducer with Streaming Deliberation.” *ICASSP*. Rhodes Island, Greece. June, 2023.
3. Jay Mahadeokar, Yangyang Shi, Ke Li, **Duc Le**, Jiedan Zhu, Vikas Chandra, Ozlem Kalinli, and Michael L. Seltzer. “Streaming Parallel Transducer Beam Search with Fast-Slow Cascaded Encoders.” *INTER-SPEECH*. Incheon, South Korea. September, 2022.
4. Yangyang Shi, Yongqiang Wang, Chunyang Wu, Ching-Feng Yeh, Julian Chan, Frank Zhang, **Duc Le**, and Michael L. Seltzer. “Emformer: Efficient Memory Transformer Based Acoustic Model For Low Latency Streaming Speech Recognition.” *ICASSP*. Virtual. June, 2021.
5. Yangyang Shi, Yongqiang Wang, Chunyang Wu, Christian Fuegen, Frank Zhang, **Duc Le**, Ching-Feng Yeh, and Michael L. Seltzer. “Weak-Attention Suppression For Transformer Based Speech Recognition.” *INTERSPEECH*. Shanghai, China. October, 2020.
6. Yongqiang Wang, Abdelrahman Mohamed, **Duc Le**, Chunxi Liu, Alex Xiao, Jay Mahadeokar, Hongzhao Huang, Andros Tjandra, Xiaohui Zhang, Frank Zhang, Christian Fuegen, Geoffrey Zweig, and Michael L. Seltzer. “Transformer-based Acoustic Modeling for Hybrid Speech Recognition.” *ICASSP*. Barcelona, Spain. May, 2020.
7. Ching-Feng Yeh, Jay Mahadeokar, Kaustubh Kalgaonkar, Yongqiang Wang, **Duc Le**, Mahaveer Jain, Kjell Schubert, Christian Fuegen, and Michael L. Seltzer. “Transformer-Transducer: End-to-End Speech Recognition with Self-Attention.” *arXiv*. 2019.

Automatic Speech Recognition - Latency Optimization

1. Yangyang Shi, Varun Nagaraja, Chunyang Wu, Jay Mahadeokar, **Duc Le**, Rohit Prabhavalkar, Alex Xiao, Ching-Feng Yeh, Julian Chan, Christian Fuegen, Ozlem Kalinli, and Michael L. Seltzer. “Dynamic Encoder Transducer: A Flexible Solution For Trading Off Accuracy For Latency.” *INTERSPEECH*. Brno, Czech Republic. September, 2021.
2. Jay Mahadeokar, Yangyang Shi, Yuan Shangguan, Chunyang Wu, Alex Xiao, Hang Su, **Duc Le**, Ozlem Kalinli, Christian Fuegen, and Michael L. Seltzer. “Flexi-Transducer: Optimizing Latency, Accuracy and Compute for Multi-Domain On-Device Scenarios.” *INTERSPEECH*. Brno, Czech Republic. September, 2021.
3. Yuan Shangguan, Rohit Prabhavalkar, Hang Su, Jay Mahadeokar, Yangyang Shi, Jiatong Zhou, Chunyang Wu, **Duc Le**, Ozlem Kalinli, Christian Fuegen, and Michael L. Seltzer. “Dissecting User-Perceived Latency of On-Device E2E Speech Recognition.” *INTERSPEECH*. Brno, Czech Republic. September, 2021.
4. Jay Mahadeokar, Yuan Shangguan, **Duc Le**, Gil Keren, Hang Su, Thong Le, Ching-Feng Yeh, Christian Fuegen, and Michael L. Seltzer. “Alignment Restricted Streaming Recurrent Neural Network Transducer.” *SLT*. January, 2021.

Automatic Speech Recognition - Other Topics

1. Mu Yang, Andros Tjandra, Chunxi Liu, David Zhang, **Duc Le**, and Ozlem Kalinli. “Learning ASR Pathways: A Sparse Multilingual ASR Model.” *ICASSP*. Rhodes Island, Greece. June, 2023.
2. Andros Tjandra, Nayan Singhal, David Zhang, Ozlem Kalinli, Abdelrahman Mohamed, **Duc Le**, and Michael L. Seltzer. “Massively Multilingual ASR on 70 Languages: Tokenization, Architecture, and Generalization Capabilities.” *ICASSP*. Rhodes Island, Greece. June, 2023.
3. Niko Moritz, Frank Seide, **Duc Le**, Jay Mahadeokar, Christian Fuegen. “An Investigation of Monotonic Transducers for Large-Scale Automatic Speech Recognition.” *SLT*. Doha, Qatar. January, 2023.
4. Alex Xiao, Weiye Zheng, Gil Keren, **Duc Le**, Frank Zhang, Christian Fuegen, Ozlem Kalinli, Yatharth Saraf, and Abdelrahman Mohamed. “Scaling ASR Improves Zero and Few Shot Learning.” *INTERSPEECH*. Incheon, South Korea. September, 2022.
5. Chunxi Liu, Frank Zhang, **Duc Le**, Suyoun Kim, Yatharth Saraf, and Geoffrey Zweig. “Improving RNN Transducer Based ASR with Auxiliary Tasks.” *SLT*. January, 2021.
6. **Duc Le**, Xiaohui Zhang, Weiye Zheng, Christian Fuegen, Geoffrey Zweig, and Michael L. Seltzer. “From Senones to Chenones: Tied Context-Dependent Graphemes for Hybrid Speech Recognition.” *ASRU*. Singapore. December, 2019.

Spoken Language Understanding

1. Suyoun Kim, Akshat Shrivastava, **Duc Le**, Ju Lin, Ozlem Kalinli, and Michael L. Seltzer. “Modality Confidence Aware Training for Robust End-to-End Spoken Language Understanding.” *INTERSPEECH*. Dublin, Ireland. August, 2023.
2. Paden Tomasello, Akshat Shrivastava, Daniel Lazar, Po-Chun Hsu, **Duc Le**, et al. “STOP: A Dataset for Spoken Task Oriented Semantic Parsing.” *SLT*. Doha, Qatar. January, 2023.
3. **Duc Le**, Akshat Shrivastava, Paden Tomasello, Suyoun Kim, Aleksandr Livshits, Ozlem Kalinli, and Michael L. Seltzer. “Deliberation Model for On-Device Spoken Language Understanding.” *INTERSPEECH*. Incheon, South Korea. September, 2022.
4. Suyoun Kim, **Duc Le**, Weiye Zheng, Tarun Singh, Abhinav Arora, Xiaoyu Zhai, Christian Fuegen, Ozlem Kalinli, and Michael L. Seltzer. “Evaluating User Perception of Speech Recognition System Quality with Semantic Distance Metric.” *INTERSPEECH*. Incheon, South Korea. September, 2022.
5. Suyoun Kim, Abhinav Arora, **Duc Le**, Ching-Feng Yeh, Christian Fuegen, Ozlem Kalinli, and Michael L. Seltzer. “Semantic Distance: A New Metric for ASR Performance Analysis Towards Spoken Language Understanding.” *INTERSPEECH*. Brno, Czech Republic. September, 2021.

Speech-Language Assessment

1. Matthew Perez, Wenyu Jin, **Duc Le**, Noelle Carlozzi, Praveen Dayalu, Angela Roberts, and Emily Mower Provost. "Classification of Huntington Disease using Acoustic and Lexical Features." *INTERSPEECH*. Hyderabad, India. September, 2018.
2. **Duc Le**, Keli Licata, and Emily Mower Provost, "Automatic Quantitative Analysis of Spontaneous Aphasic Speech." *Speech Communication*. June, 2018.
3. **Duc Le**, Keli Licata, and Emily Mower Provost. "Automatic Paraphasia Detection from Aphasic Speech: A Preliminary Study." *INTERSPEECH*. Stockholm, Sweden. August, 2017.
4. **Duc Le** and Emily Mower Provost. "Improving Automatic Recognition of Aphasic Speech with Aphasia-Bank." *INTERSPEECH*. San Francisco, USA. September, 2016.
5. **Duc Le**, Keli Licata, Carol Persad, and Emily Mower Provost. "Automatic Assessment of Speech Intelligibility for Individuals with Aphasia." *IEEE/ACM TASLP*. August, 2016.
6. **Duc Le** and Emily Mower Provost. "Modeling Pronunciation, Rhythm, and Intonation for Automatic Assessment of Speech Quality in Aphasia Rehabilitation." *INTERSPEECH*. Singapore. September, 2014.
7. **Duc Le**, Keli Licata, Elizabeth Mercado, Carol Persad, and Emily Mower Provost. "Automatic Analysis of Speech Quality for Aphasia Treatment." *ICASSP*. Florence, Italy. May, 2014.

Speech Emotion Recognition

1. **Duc Le**, Zakaria Aldeneh, and Emily Mower Provost. "Discretized Continuous Speech Emotion Recognition with Multi-Task Deep Recurrent Neural Network." *INTERSPEECH*. Stockholm, Sweden. August, 2017.
2. John Gideon, Biqiao Zhang, Zakaria Aldeneh, Yelin Kim, Soheil Khorram, **Duc Le**, and Emily Mower Provost. "Wild Wild Emotion: A Multimodal Ensemble Approach." *ICMI*. Tokyo, Japan. November, 2016.
3. **Duc Le** and Emily Mower Provost. "Data Selection for Acoustic Emotion Recognition: Analyzing and Comparing Utterance and Sub-Utterance Selection Strategies." *ACII*. Xi'an, China. September, 2015.
4. **Duc Le** and Emily Mower Provost. "Emotion Recognition from Spontaneous Speech using Hidden Markov Models with Deep Belief Networks." *ASRU*. Olomouc, Czech Republic. December, 2013.
5. Je Hun Jeon, **Duc Le**, Rui Xia, and Yang Liu. "A Preliminary Study of Cross-lingual Emotion Recognition from Speech: Automatic Classification versus Human Perception." *INTERSPEECH*. Lyon, France. August, 2013.

SERVICES

Session Chair	ICASSP (2021, 2022, 2023), INTERSPEECH (2022)
Reviewer	ICASSP, INTERSPEECH, SLT, ASRU, TASLP, Speech Communication

TECHNICAL SKILLS

Programming Languages	Python, C/C++, Java, Bash, Lua, C#, Ruby, SQL
Software Frameworks	PyTorch, Caffe2, NumPy, Torch, Theano, Hive, Kaldi, HTK
Specializations	Speech Recognition, Music Generation, Music Understanding

LANGUAGES

English, Vietnamese