

Byeonggeun Kim

Senior Applied Scientist · Amazon AGI

bungkun1349@gmail.com
byeonggeunkim.github.io
Google Scholar | LinkedIn

RESEARCH PROFILE

I am a Senior Applied Scientist at Amazon AGI working on audio and speech modeling. My work spans efficient on-device audio, few-shot and robust learning, and audio foundation and generative models. Earlier Amazon projects included audio encoders, neural codecs, audio and music generation, and contributions to Amazon Nova; I currently focus on full-duplex speech-to-speech modeling.

RESEARCH EXPERIENCE

Amazon AGI Seattle, WA

Senior Applied Scientist Apr. 2025 – Present

Current focus: full-duplex speech-to-speech modeling for natural, real-time spoken interactions.

Selected talks: Yale University, CPSC 7760 Seminar — “Efficient and High Fidelity Text-to-Audio Generation: From Parallel Decoding to Causal Streaming” (Feb. 2026); Amazon Media & Entertainment ML+AI Summit — “Autoregressive text-to-audio generation approaches” (Sep. 2025).

Applied Scientist II Feb. 2023 – Mar. 2025

Earlier work: audio understanding encoders, neural codecs, and audio and music generation for Amazon AGI projects, including contributions to Amazon Nova’s multimodal models.

Recognition: Alexa Perceptual Technology Award (Q3/Q4 2023).

Qualcomm AI Research Seoul, South Korea

Staff AI Researcher Nov. 2022 – Jan. 2023

Senior AI Researcher Dec. 2019 – Nov. 2022

AI Researcher Apr. 2018 – Nov. 2019

Developed efficient architectures and few-shot, open-set, and domain-robust methods for on-device keyword spotting and acoustic scene classification.

SELECTED OUTCOMES

Amazon Nova 2. Audio and music generation, audio encoder, and neural codec contributions to Amazon’s multimodal models.

BC-ResNet. Efficient keyword spotting architecture published at INTERSPEECH 2021.

DCASE 2021. First-place low-complexity acoustic scene classification system.

Research output. 20 publications and 9 granted U.S. patents across audio and machine learning.

PUBLICATIONS

(* denotes equal contribution; self-name in bold.)

1. **Interspeech 2026 (Oral)** Improving Text-to-Audio Instruction Following via Fine-Grained Feedback from Audio-Aware Large Language Models [PDF]
Chun-Yi Kuan, Siwon Kim, **Byeonggeun Kim**, et al. (*mentored internship*)
2. **arXiv preprint, 2026** Fast Text-to-Audio Generation with One-Step Sampling via Energy-Scoring and Auxiliary Contextual Representation Distillation [PDF]
Kuan-Po Huang, Bo-Ru Lu, **Byeonggeun Kim**, et al. (*mentored internship*)
3. **ICML 2025** Generative Audio Language Modeling with Continuous-valued Tokens and Masked Next-Token Prediction [PDF]
Shu-wen Yang, **Byeonggeun Kim**, Kuan-Po Huang, Qingming Tang, Huy Phan, Bo-Ru Lu, et al. (*mentored internship*)
4. **ICML 2025** IMPACT: Iterative Mask-based Parallel Decoding for Text-to-Audio Generation with Diffusion Modeling [PDF]
Kuan-Po Huang, Shu-wen Yang, Huy Phan, Bo-Ru Lu, **Byeonggeun Kim**, et al. (*mentored internship*)
5. **ICASSP 2025 (Oral)** Effective Techniques for Scaling Audio Encoder Pretraining [PDF]
Byeonggeun Kim*, Andrew Bydlon*, Qingming Tang, Huy Phan, Chieh-Chi Kao, Tao Zhang, Chao Wang.
6. **NeurIPS 2025 Workshop** Unlocking Transfer Learning for Open-World Few-Shot Recognition [PDF]
Byeonggeun Kim*, Jun-Tae Lee*, Kyuhong Shim, Simyung Chang. (*Reliable ML Workshop*)
7. **Amazon Tech Report 2025** Amazon Nova 2: Multimodal reasoning and generation models [PDF]
Amazon AGI (*contributor: audio/music generation, audio encoder & codec*).
8. **ICASSP 2024** Cross-Triggering Issue in Audio Event Detection and Mitigation [PDF]
Huy Phan, **Byeonggeun Kim**, Vu Nguyen, Andrew Bydlon, Qingming Tang, et al.
9. **ICIP 2023** Task-Agnostic Open-Set Prototype for Few-Shot Open-Set Recognition [PDF]
Byeonggeun Kim*, Jun-Tae Lee*, Kyuhong Shim, Simyung Chang.
10. **ICASSP 2023 (Oral)** Scalable Weight Reparametrization for Efficient Transfer Learning [PDF]
Byeonggeun Kim*, Jun-Tae Lee*, Seunghan Yang, Simyung Chang.
11. **Interspeech 2023 (Oral)** Improving Small Footprint Few-shot Keyword Spotting with Supervision on Auxiliary Data [PDF]
Seunghan Yang, **Byeonggeun Kim**, Kyuhong Shim, Simyung Chang.
12. **ICLR 2023** TTN: A Domain-Shift Aware Batch Normalization in Test-Time Adaptation [PDF]
Hyesu Lim, **Byeonggeun Kim**, Jaegul Choo, Sungha Choi. (*mentored internship*)

13. **Interspeech 2022 (Oral)** Personalized Keyword Spotting through Multi-task Learning [PDF]
Seunghan Yang, **Byeonggeun Kim**, Inseop Chung, Simyung Chang.
14. **Interspeech 2022** Dummy Prototypical Networks for Few-shot Open-set Keyword Spotting [PDF]
Byeonggeun Kim, Seunghan Yang, Inseop Chung, Simyung Chang.
15. **Interspeech 2022** Domain Generalization with Relaxed Instance Frequency-wise Normalization for Multi-device Acoustic Scene Classification [PDF]
Byeonggeun Kim, Seunghan Yang, Jangho Kim, Hyunsin Park, Juntae Lee, Simyung Chang.
16. **Interspeech 2021** Broadcasted Residual Learning for Efficient Keyword Spotting (BC-ResNet) [PDF] [Code]
Byeonggeun Kim*, Simyung Chang*, Jinkyu Lee, Dooyong Sung.
17. **DCASE 2021 Workshop** Domain Generalization on Efficient Acoustic Scene Classification Using Residual Normalization [PDF]
Byeonggeun Kim, Seunghan Yang, Jangho Kim, Simyung Chang.
18. **DCASE 2021 Challenge — 1st Place** QTI Submission to DCASE 2021: Residual Normalization for Device-Imbalanced Acoustic Scene Classification with Efficient Design [PDF]
Byeonggeun Kim, Seunghan Yang, Jangho Kim, Simyung Chang.
19. **ASRU 2019** Query-by-Example On-Device Keyword Spotting [PDF] [Dataset]
Byeonggeun Kim, Mingu Lee, Jinkyu Lee, Yeonseok Kim, Kyuwoong Hwang.
20. **ASRU 2019** Orthogonality Constrained Multi-Head Attention for Keyword Spotting [PDF]
Mingu Lee, Jinkyu Lee, Hye Jin Jang, **Byeonggeun Kim**, Wonil Chang, Kyuwoong Hwang.

GRANTED PATENTS

1. *Acoustic event detection* [PDF] **US 12,646,502**
Quoc Huy Phan, **Byeonggeun Kim**, Andrew Thomas Bydlon, Qingming Tang, Chieh-Chi Kao, Chao Wang, Tien Vu Nguyen. (*Jun. 2026*)
2. *Multi-task learning for personalized keyword spotting* [PDF] **US 12,347,439**
Seunghan Yang, **Byeonggeun Kim**, Inseop Chung, Simyung Chang. (*Jul. 2025*)
3. *Relaxed instance frequency normalization for neural-network-based audio processing* [PDF] **US 12,266,379**
Byeonggeun Kim, Seunghan Yang, Hyunsin Park, Juntae Lee, Simyung Chang. (*Apr. 2025*)
4. *Target Keyword Selection* [PDF] **US 12,039,968**
Wonil Chang, Jinseok Lee, . . . , **Byeonggeun Kim**, et al. (*Jul. 2024*)
5. *Task Agnostic Open-set Prototypes for Few-shot Open-set Recognition* [PDF] **US 12,019,641**
Byeonggeun Kim, Juntae Lee, Simyung Chang. (*Jun. 2024*)

6. *Systems and methods of image processing based on gaze detection* [PDF] US 11,798,204
Hyunsin Park, Juntae Lee, Simyung Chang, **Byeonggeun Kim**, Jaewon Choi, Kyu Woong Hwang. (Oct. 2023)
7. *On-device self training in two-stage wakeup system. . .* [PDF] US 11,664,012
Young Mo Kang, Sungrack Yun, Kyu Woong Hwang, Hye Jin Jang, **Byeonggeun Kim**. (May 2023)
8. *Activating speech recognition based on hand patterns detected using plurality of filters* [PDF] US 11,437,031
Sungrack Yun, Young Mo Kang, Hye Jin Jang, **Byeonggeun Kim**, Kyu Woong Hwang. (Sep. 2022)
9. *Method and apparatus for activating speech recognition* [PDF] US 11,205,433
Byeonggeun Kim, Young Mo Kang, Sungrack Yun, Kyu Woong Hwang, Hye Jin Jang. (Dec. 2021)

PROFESSIONAL ACTIVITIES

Area Chair: ICASSP 2026.

Session Chair: INTERSPEECH 2022.

Reviewer: ICASSP (2023–2026), INTERSPEECH (2023–2026), CVPR (2024–2026), ICCV 2025, ECCV 2026, ACL 2025, AAAI (2025–2026), WACV (2025, 2026), BMVC 2026, NeurIPS 2026, EMNLP 2025.

EDUCATION

Korea Advanced Institute of Science & Technology (KAIST) Daejeon, South Korea
M.S., Electrical and Electronic Engineering 2018
B.S., Electrical and Electronic Engineering & Business Administration 2016

LANGUAGES

Korean (native), English (full professional proficiency), Japanese (conversational).