

Anna Seo Gyeong Choi

sc2359@cornell.edu · [Linkedin](#) · [Website](#)

Research interests

AI Ethics, Algorithmic Fairness, Speech/Signal Processing,

Education

- 2021 – Present **Cornell University**
Ph.D. in Information Science
Committee: Allison Koenecke (Chair), Matthew Wilkens, Marten van Schijndel
- 2019 – 2021 **Seoul National University**
M.A. in English Language and Literature
Thesis: *Analysis of Lexical Stress of Korean-accented English and Spanish-accented English for Accent Classification*
- 2015 – 2019 **Seoul National University**
B.A. in Hispanic Language and Literature, *cum laude*
B.A. in English Language and Literature, *cum laude*

Research experience

- 2022.08 - present **Department of Information Science, GRA**
Prof. Allison Koenecke (Cornell University)
Prof. Hilke Schellmann (New York University)
Prof. Mona Sloane (University of Virginia)
Processed speech datasets of non-“standard” English speech for speech-to-text transcription auditing with applications to hiring and healthcare domains; ran multiple commercial speech-to-text APIs, performed data analysis on transcription word error rates by linguistic dialect, and provided linguistic expertise on grammatical and phonological attributes of speech variants.
- 2019.07 – 2021.08 **Spoken Language Processing Lab**
Prof. Minhwa Chung (Seoul Nat’l Univ.)
Developed ASR system-based CALL tool, built large data speech corpora for specialized purposes, phonetically analyzed speech data, led multiple transcription projects.
- Acquiring and Utilizing Audio Big Data for Healthcare Service Design, supported by Korean Ministry of Science and ICT (Project manager)
 - Development of Free-running Speech Recognition Technologies for Embedded Robot System, supported by Korean Ministry of Trade, Industry and Energy
 - Development on Cognition of Professional Field-Specific Information and Learning Model Technology, supported by Korean Ministry of Science, ICT
 - Development of Content Creation and Entertainment Technologies Based on Intelligent Authoring Tool to Enhance Accessibility of Social Communication Disabilities, supported by Korean Ministry of Culture, Sports and Tourism

Publications

- 2024
(Upcoming) **Careless Whisper: Speech-to-Text Hallucination Harms**
Koenecke, A., **Choi, A. S. G.***, Mei, K.*, Schellmann, H.†, & Sloane, M.† (*, † equal contribution)
Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)
- 2024
(Upcoming) **KoFREN: Comprehensive Korean Word Frequency Norms Derived from Large Scale Free Speech Corpora**
Kim, J., **Choi, A. S. G.**, & Cho, S.
Proceedings of Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)
- 2024
(Upcoming) **Crosslinguistic Acoustic Feature-based Dementia Classification using Advanced Learning Architectures**
Choi, A. S. G., Kim, J., Kim, S., Baek, M. & Cho, S.
Proceedings of Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING) Workshop on Resources and Processing of Linguistic, Para-linguistic and Extra-linguistic Data from People with Various Forms of Cognitive/Psychiatric/Developmental Impairments (RaPID-5)
- 2023 **Augmented Datasheets for Speech Datasets and Ethical Decision-Making**
Papakyriakopoulos, O.*, **Choi, A. S. G.***, Thong, W., Zhao, D., Andrews, J., Bourke, R., Xiang, A.†, & Koenecke, A.† (*, † equal contribution)
Proceedings of ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)
- 2020 **Building Korean Conversational Speech Data in Emergency Medical Domain**
Kim, S., Lee, J., **Choi, S.G.**, Ji, S., Kang, J., Kim, J., et al.
Phonetics and Speech Sciences, 12 (4).

Presentations

- May 2024
(Upcoming) **Quantification of Automatic Speech Recognition System Performance on d/Deaf and Hard of Hearing Speech** (Oral)
American Laryngological Association (ALA) Spring Meeting at COSM 2024 Annual Meeting
- Jul 2023 **Auditing Korean Speech Datasets for Dialectal Fairness in Speech-to-Text Applications** (Poster)
9th International Conference on Computational Social Science (IC2S2 '23)
- Jul 2023 **Augmented Datasheets for Speech Datasets and Ethical Decision-Making** (Poster)
9th International Conference on Computational Social Science (IC2S2 '23)
- Apr 2022 **Phoneme-to-phoneme (P2P) lexicon expansion technology for Brazilian Portuguese** (Poster)
Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2022)
- Nov 2020 **Building Korean Conversational Speech Data in Emergency Medical Domain** (Oral)
Korean Conference on Speech Sciences

Oct 2019 **Stereotypical dialects in *Zootopia* and its implications** (Oral)
Coalition of English Departments in Asia

Awards

2023 Microsoft Azure Cloud Computing Grant (Cornell University)
2022 Cornell Center for Social Science Grant (Cornell University)
2021 Sage Fellowship (*Tuition + Stipend*) (Cornell University)
2019 - 2020 Graduate Student Instructor Scholarship (Seoul Nat'l Univ.)
2019 CORE Travel Grant for Conference (Seoul Nat'l Univ.)
2017 - 2019 CORE Undergraduate Scholarship (Seoul Nat'l Univ.)
2017 CORE Exchange Students Scholarship (Seoul Nat'l Univ.)
2015 Scholarship for Academic Excellence (Seoul Nat'l Univ.)

Teaching experience

Cornell Univ. **LING 1101: Introduction to Linguistics, SPAN 1210: Elementary Spanish I, INFO 2950: Introduction to Data Science**

Seoul Nat'l Univ. **Graduate Seminar on Applied Linguistics, Graduate Course on English Phonology, English Phonetics, Advanced English, College English 1, 2, English Foundation, English and Society**