Jeongsoo Choi

Ph.D. candidate, Multimodal AI Lab.
Korea Advanced Institute of Science and Technology (KAIST)

Contact

Email	jeongsoo.choi@kaist.ac.kr
Webpage	https://choijeongsoo.github.io
Phone	+82-10-2106-5748

Research Interest

My research interests lie in Multimodal Machine Learning, Speech Processing, and applications.
- Multilingual and Multimodal Machine Translation
- Audio-Visual Speech Recognition and Generation

Work Experience

Microsoft, Beijing, China   May. 2024 - Oct. 2024
Research Intern, Speech Team
- Advisors: Shujie Liu and Jinyu Li

Education

Ph.D. in Electrical Engineering, **KAIST**   Sep. 2020 - Present
- Advisor: Prof. Joon Son Chung
B.S. in Electrical Engineering, **KAIST**   Mar. 2015 - Aug. 2020

Publications   <C: Conference, J: Journal, P: Preprint, ∗: Equal Contribution>

[P3] **V2SFlow: Video-to-Speech Generation with Speech Decomposition and Rectified Flow**
Jeongsoo Choi*, Ji-Hoon Kim*, Jinyu Li, Joon Son Chung, Shujie Liu
Under review

[P2] **ARLON: Boosting Diffusion Transformers with Autoregressive Models for Long Video Generation**
Zongyi Li*, Shujie Hu*, Shujie Liu, Long Zhou, Jeongsoo Choi, Lingwei Meng, Xun Guo, Jinyu Li, Hefei Ling, Furu Wei
arXiv 2024, *Under review*

[P1] **Accelerating Codec-based Speech Synthesis with Multi-Token Prediction and Speculative Decoding**
Tan Dat Nguyen, Ji-Hoon Kim, Jeongsoo Choi, Shukjae Choi, Jinseok Park, Younglo Lee, Joon Son Chung
arXiv 2024, *Under review*

[J2] **Textless Unit-to-Unit training for Many-to-Many Multilingual Speech-to-Speech Translation**
Minsu Kim*, Jeongsoo Choi*, Dahun Kim, Yong Man Ro
IEEE/ACM TASLP (2024)

[C9] **AV2AV: Direct Audio-Visual Speech to Audio-Visual Speech Translation with Unified Audio-Visual Speech Representation**
Jeongsoo Choi*, Se Jin Park*, Minsu Kim*, Yong Man Ro
CVPR 2024 Highlight Presentation

[C8] **Text-driven Talking Face Synthesis by Reprogramming Audio-driven Models**
Jeongsoo Choi, Minsu Kim, Se Jin Park, Yong Man Ro
ICASSP 2024

[C7] **Towards Practical and Efficient Image-to-Speech Captioning with Vision-Language Pre-training and Multi-modal Tokens**
Minsu Kim, Jeongsoo Choi, Soumi Maiti, Jeong Hun Yeo, Shinji Watanabe, Yong Man Ro
ICASSP 2024

[C6] **Exploring Phonetic Context-Aware Lip-Sync For Talking Face Generation**
Se Jin Park, Minsu Kim, Jeongsoo Choi, Yong Man Ro
ICASSP 2024

[J1] **AKVSR: Audio Knowledge Empowered Visual Speech Recognition by Compressing Audio**
Jeong Hun Yeo, Minsu Kim, Jeongsoo Choi, Dae Hoe Kim, Yong Man Ro
IEEE TMM (2024)

[C5] **DiffV2S: Diffusion-based Video-to-Speech Synthesis with Vision-guided Speaker Embedding**
Jeongsoo Choi*, Joanna Hong*, Yong Man Ro
ICCV 2023

[C4] **Lip Reading for Low-resource Languages by Learning and Combining General Speech Knowledge and Language-specific Knowledge**
Minsu Kim*, Jeong Hun Yeo*, Jeongsoo Choi, Yong Man Ro
ICCV 2023

[C3] **Intelligible Lip-to-Speech Synthesis with Speech Units**
Jeongsoo Choi, Minsu Kim, Yong Man Ro
Interspeech 2023

[C2] **Watch or Listen: Robust Audio-Visual Speech Recognition With Visual Corruption Modeling**
Joanna Hong*, Minsu Kim*, Jeongsoo Choi, Yong Man Ro
CVPR 2023

[C1] **SyncTalkFace: Talking Face Generation with Precise Lip-Syncing via Audio-Lip Memory**
Se Jin Park, Minsu Kim, Joanna Hong, Jeongsoo Choi, Yong Man Ro
AAAI 2022 **Oral Presentation**

Programming Skills

Languages
- Python, C/C++, and MATLAB

Deep Learning Libraries
- Pytorch, Fairseq, and HuggingFace

References

Prof. Joon Son Chung
- School of Electrical Engineering, KAIST
- Email: joonson@kaist.ac.kr
- Homepage: https://mmai.io/joon
- Relationship: Ph.D. advisor at KAIST