Karan Thakkar

kthakka2@jhu.edu | +1(667)-900-2879 | github/carankt | GScholar | LinkedIn | carankt.github.io

EDUCATION

Johns Hopkins University Baltimore, MD
Ph.D. student in Electrical and Computer Engineering; GPA: 3.98 *Aug. 2021 – Present*

Vellore Institute of Technology Vellore, India
Bachelor of Technology in Electrical and Electronics; GPA: 3.72 (9.33/10.0) *Aug. 2016 – July. 2020*

RESEARCH EXPERIENCE

Laboratory for Computational Audio Perception Ph.D. Candidate
Python, MATLAB, PyTorch, PsychoPy, EEGLAB *Aug 2021 - Present*

- **State Space Modeling of Top-Down and Bottom-Up Auditory Attention**: Developing dynamic and explainable models of Auditory Attention, leveraging audio stimuli to analyze and interpret brain states from EEG data.
- **Self Supervised Features for Auditory Attention Decoding**: Evaluated 12 self-supervised and 2 shallow representations for Auditory Attention Decoding in EEG data from 57 subjects, revealing deep features' superiority in decoding across multiple languages.
- **Interdisciplinary Neural Systems Research**: Collaborated with researchers from the Neural Systems Laboratory at the University of Maryland to investigate auditory segregation in ferrets and compared neural responses with neural activations, revealing enhanced processing of attended audio in both experiments.
- **Generative Model for Sound Separation**: Innovated generative diffusion probabilistic model for target sound extraction, outperforming discriminative methods in perceptual qualities by 10%. [project link]
- **Bio-inspired Framework for Speech and Music Separation**: Engineered the Explicit Memory Multiresolution Adaptive framework drawing inspiration from neuro-auditory principles of temporal coherence, memory and adaptation.

DeepSync Technologies (now DubPro.ai) Machine Learning Engineer
Bash, Git, Python, PyTorch, Azure *Jul 2020 - Jul 2021*

- **Signal Pre-Processing Automation**: Engineered and streamlined an automated system for audio data acquisition, validation, and preprocessing from diverse artists, effectively managing over 10,000 hours of raw audio content.
- **Scalable Text to Speech (TTS) System Development**: Crafted a high-performance end-to-end text-to-speech system, modeled after FastSpeech2, which attained a speech inference speed 7x faster than Google's Tacotron2.
- **Fast and Efficient Vocoding**: Innovated a mel-spectrogram to waveform vocoder inspired by HiFi-GAN and VocGAN.

VIT Vellore and NUS Singapore Undergraduate Research Fellow
Python, MATLAB, Keras, Tensorflow, GCP, Hadoop *Jul 2018 - Apr 2020*

- **Deep Learning in Wireless Communication Systems**: Developed and optimized an OFDM system with an $O(n)$ deep signal detection algorithm, outperforming traditional MMSE and LS methods. Integrated a deep learning-based recovery and denoising block enhancing noise robustness and independence from Cyclic Prefix.
- **Speech Emotion and Sentiment Analysis @ National University of Singapore**: Explored chirp wavelet-based features for classifying emotions in speech. Secured 2nd prize in Kaggle text sentiment analysis competition with a custom-developed LSTM model.

ACADEMIC PROJECTS & OPEN SOURCE CONTRIBUTIONS

Telluride Neuromorphic Workshop Summer' 23
Python, PyTorch, MNE-Python, EEGLab *Presentation Link*

- **Decoding Audio from MEG**: Investigated discrete representations from EncoDec Model for decoding audio from MEG signals of subjects watching the Forest Gump Movie.
- **Brain Computer Interfacing (BCI) through real-time Auditory Attention**: Explored the idea of using Auditory Attention to control a car in an obstacle avoidance game. Algorithm based on the ERP template match and mismatch.

Large Language Model Prompting and Reasoning Fall' 22
Python, PyTorch *ArXiv Report*

- **Stress Testing Chain-of-Thought (CoT) Prompting for Large Language Models**: Examined the effectiveness of CoT prompting in improving multi-step reasoning abilities of LLMs on logical reasoning tasks.

- **Speech, Breathing and Cough Sound analysis for COVID-19 Detection**   Fall' 21
  Python, PyTorch, Opensmile   *carankt/CovidDetection*
  - ○ **Audio Biomarkers for COVID-19 Detection**: Implemented a two-part feature selection process and performed in-depth analysis to identify effective features for COVID-19 classification on the COSWARA dataset, enhancing signal-level understanding and classification accuracy.

- **Inner Speech Decoding from EEG**   Fall' 21
  Python, PyTorch, SKLearn, MNE-Python   *carankt/synthetic_telepathy*
  - ○ **Multi-Class Classification of Inner Speech Using EEG**: Collaborated on an EEG data project for inner speech decoding (4-direction and 16-syllable datasets), assisting patients with speech impairments. Utilized preprocessing, clustering, SVMs, and various deep learning models for effective decoding.

- **Enabling Visual Perception for the Blind through Auditory Signals**   Fall' 18
  Python, PyTorch, Keras   *Report*
  - ○ **Advancing Inclusive Technology**: AI-powered smart glasses translating scenes into speech for visual assistance, and training programs for perceiving images through sound, aimed at enhancing accessibility for the specially abled.

- **Open Source Contributions**   Summer '20 - Present
  Python, PyTorch   *github.com/carankt*
  - ○ **Audio**: Contributed to cutting-edge audio processing and generation projects. Worked on Audiocraft for deep learning-based audio compression and music generation, implemented a PyTorch version of carankt/FastSpeech 2 for efficient text-to-speech conversion, and modified the VocGAN vocoder in rishikksh20/VocGAN for improved training efficiency and voice quality. Collaborated on re-implementing haidog-yaqub/Audio-Stable-Diffusion.
  - ○ **EEG Decoding**: Contributed an unofficial PyTorch implementation of the VLAAI network, carankt/vlaai-pytorch, for decoding speech envelopes from EEG signals, enhancing the accessibility of this deep learning model for EEG data analysis and interpretation.

PUBLICATIONS & PREPRINTS

- **Thakkar, K.**, Hai, J., Elhilali M. (2023). Investigating Self-Supervised Deep Representations for EEG-based Auditory Attention Decoding. arXiv:2311.00814.(Accepted at ICASSP '24)

- Hai, J., Wang, H., Yang, D., **Thakkar, K.**, Chong, D., Dehak, N., Elhilali, M. (2023). DPM-TSE: A Diffusion Probabilistic Model for Target Sound Extraction. arXiv:2310.04567(Accepted at ICASSP '24).

- Bellur, A., **Thakkar, K.** Elhilali, M. Explicit-memory multiresolution adaptive framework for speech and music separation (2023) https://doi.org/10.1186/s13636-023-00286-7.

- **K. Thakkar**, A. Goyal, and B. Bhattacharyya, (2020) Emergence of Deep Learning as a potential solution for Detection, Recovery, and De-noising of Signals in Communication Systems https://doi.org/10.1016/j.ijin.2020.12.001..

- **K. Thakkar**, A. Goyal, and B. Bhattacharyya, (2020) Deep Learning and Channel Estimation doi:10.1109/ICACCS48705.2020.9074414..