# Speech Enhancement for Hearing-impaired Patients Using Deep Learning and Real-Time Deployment on Smartphones

Varsha Venkatapathy, Muhammad Patel, Arian Azarang, Virginie Papadopoulou
Lampe Joint Department of Biomedical Engineering, University of North Carolina & North Carolina State University, Raleigh, NC

## INTRODUCTION

- Hearing aids often struggle to suppress background noise, reducing speech intelligibility.
- Deep learning-based speech enhancement offers a promising solution by improving clarity in noisy environments.
- This project explores the real-time deployment of speech enhancement models on smartphones to assist hearing aid users.
- Four models from existing literature are selected for evaluation.
- Models are evaluated on accuracy using Perceptual Evaluation of Speech Quality (PESQ), Short Time Objective Intelligibility (STOI), and Mean Squared Error (MSE), and Mean Absolute Error metrics (MAE).
- The goal is to identify the most effective model for improving speech intelligibility while ensuring compatibility with real-time smartphone deployment.

## METHODS: SPEECH ENHANCEMENT PIPELINE

- Clean speech from a public domain dataset was mixed with Babble, Traffic, and Machinery noise at SNRs of -5, 0, and 5 dB to create realistic noisy datasets → split 70/20/10 into training, validation, and test sets.
- **Article 1 (CRNN):** STFT → Magnitude Spectrum → Normalization → Reshape to 2D matrix [1].
- **Article 2 (CNN):** Frame Segmentation → FFT → Log Power Spectrum → Add Noise → Mel-filterbank → Temporal Frame Concatenation → Normalization [2].
- **Article 3 (RNN):** Frame Segmentation (16 ms, 50% overlap) → STFT → Extract Real & Imaginary Components → Concatenate & Normalize → Format for RNN input [3].
- **Article 4 (FCNN):** STFT → Log-Power Spectrum → Normalization → Reshape into 2D matrix for IRM estimation [4].

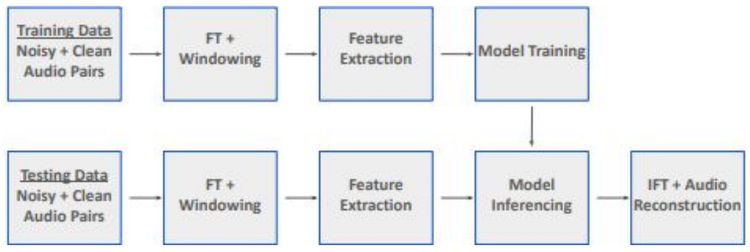


**Figure 1: General SE Pipeline**

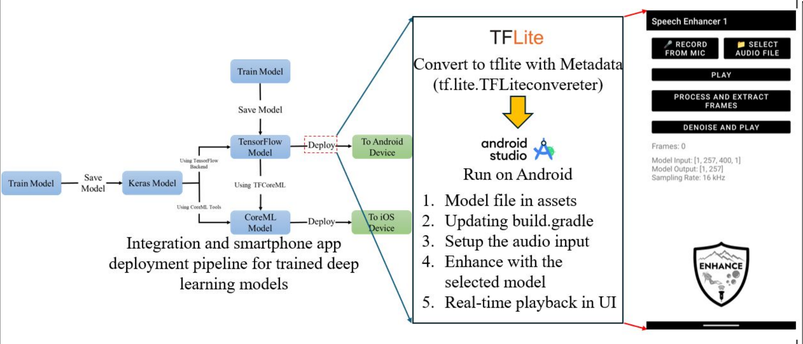


**Figure 2: Deployed deep learning model on Android app for real time SE**

## METHODS: MODEL ARCHITECTURE

- The CRNN model combines convolutional and recurrent layers to extract features and model temporal patterns for real-time speech enhancement (**Fig. 1**).
- The CNN model applies frame segmentation and temporal stacking with convolutional layers to suppress noise efficiently (**Fig. 2**).
- The RNN model uses sequential recurrent layers to process dual-mic inputs and capture temporal speech features (**Fig. 3**).
- The FCNN model estimates a 2D ideal ratio mask from LPS of noisy speech using fully convolutional layers (**Fig. 4**).
- All experiments were conducted in Jupyter Notebook on a Windows 10 system equipped with an NVIDIA GeForce RTX 4090 GPU (24 GB, 10,496 CUDA cores), an Intel i9-10850K CPU, and 128 GB of RAM. This high-performance computing environment supported efficient model training and reproducibility.

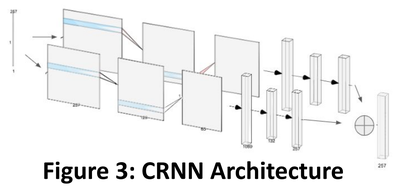

**Figure 3: CRNN Architecture**

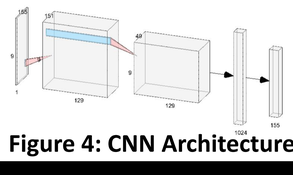

**Figure 4: CNN Architecture**

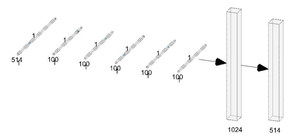

**Figure 5: RNN Architecture**

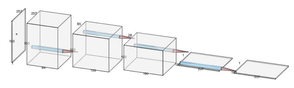

**Figure 6: FCNN Architecture**

## RESULTS

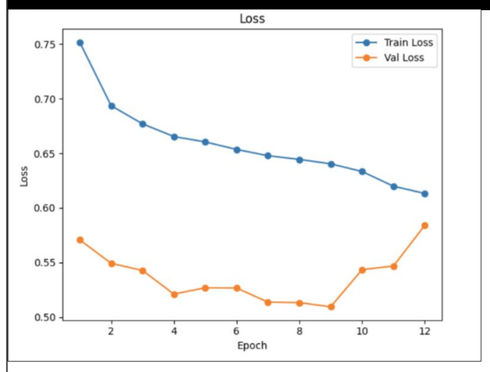


**Figure 7: MSE/MAE**

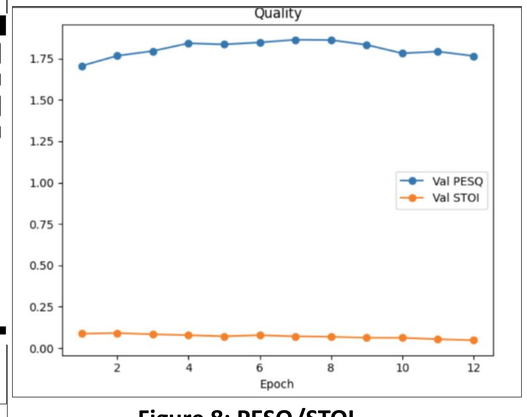


**Figure 8: PESQ/STOI**

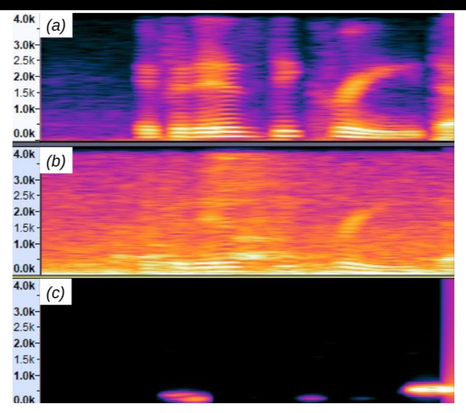


**Figure 9: Clean vs. Noisy Spectrogram**

Above all the results from running the CNN model to denoise speech. **Figure 7** shows Epoch Wise Trends for MSE(Mean Square Error) and MAE (Mean Absolute Error) for Training and Validation. **Figure 8** shows the Validation Speech Quality over Epochs for Perceptual Evaluation of Speech Quality(PESQ) and Short-Time Objective Intelligibility (STOI). **Figure 9** shows the Time-Frequency Spectrograms of Clean Speech, Noisy Input, and Learned Enhancement Mask.

## CONCLUSION

Based upon our results from the CNN model that used Log-Power Spectral and Log-Mel Filter Bank Energy Features:

- **Training:**
  - Adam at $1 \times 10^{-4}$ LR, reached early stopping at epoch 12/15
  - Batch size 4 on ~100k windows, 9-frame context
- **Loss (MSE/MAE):**
  - Fell from ~0.75 to ~0.61, indicating tighter match to clean log-power spectra
- **Perceptual Quality (PESQ):**
  - Improved to ~1.9/4.5 (up from ~1.5), meaning noticeably clearer, more natural speech
- **Intelligibility (STOI):**
  - Rose to ~0.08/1.0 (vs. ~0.05), a modest but consistent gain in speech intelligibility
- **Conclusion:**
  - Attention-CNN reliably reduces noise, yielding clearer, more natural speech with modest intelligibility gains. Further training needs to be done to boost STOI. Ongoing training currently be done for CRNN, RNN, and FCNN models

## FUTURE DEVELOPMENT

- Optimize the best-performing model for reduced memory and computational requirements to improve smartphone deployment efficiency.
- Expand testing to include a broader range of noise types and acoustic environments for improved generalizability.
- Conduct user testing with hearing-impaired individuals to evaluate perceived improvements in speech intelligibility.
- Explore integration with existing hearing aid hardware or smartphone-based hearing aid apps.
- Retrain models on higher-memory devices with increased batch sizes and more training epochs to improve performance.

## REFERENCES

1. N. Shankar, G. S. Bhat, and I. M. S. Panahi, "Real-Time Single-Channel Deep Neural Network-Based Speech Enhancement on Edge Devices," *PubMed Central*, Oct. 2020
2. G. S. Bhat, N. Shankar, C. K. A. Reddy, and I. M. S. Panahi, "A Real-Time Convolutional Neural Network Based Speech Enhancement for Hearing Impaired Listeners Using Smartphone," *IEEE Access*, vol. 7, pp. 78421–78433, 2019
3. N. Shankar, G. S. Bhat, and I. M. S. Panahi, "Efficient two-microphone speech enhancement using basic recurrent neural network cell for hearing and hearing aids," vol. 148, no. 1, pp. 389–400, Jul. 2020
4. Y.-H. Tu, J. Du, and C. Lee, "2D-to-2D Mask Estimation for Speech Enhancement Based on Fully Convolutional Neural Network," May 2020

## ACKNOWLEDGMENTS

We thank the Office of Undergraduate Research (OUR) at UNC Chapel Hill and the BME Abrams Scholars Program at UNC Chapel Hill and NC State University for funding.

**Lampe Joint Department of Biomedical Engineering**

UNC CHAPEL HILL | NC STATE UNIVERSITY