UWI
ST. AUGUSTINE CAMPUS
FACULTY OF SCIENCE & TECHNOLOGY

DEPARTMENT OF PHYSICS

SEMINAR SERIES: 2025/2026

On the Introduction of *MaU-Net*: A Series of Foundational Experiments for Longer-mixture Speech Separation

Presenter: Mr. Bryan Hibbert

Date: Wednesday 11th February, 2026

Time: 11:00 am

Room: FST 414

ZOOM LINK: https://sta-uwi-edu.zoom.us/j/98473300664?pwd=y3GnSzcEacaiv1X49HX4WT9S5G5SUr.1

On the Introduction of *MaU-Net*: A Series of Foundational Experiments for Longer-mixture Speech Separation

The use of *Spectral* (*Time-Frequency*) *Masking* has been established for audio-related *Source Separation* and through experimentation, we have verified that *Complex* Spectral Masking is the most promising of the Time-Frequency masking techniques. Furthermore, it has been shown that *Deep Learning* algorithms can transform audio data into *Spectral-like* (having higher dimensionality) representations that are even better suited (than Time-Frequency spectra) for the audio separation process. Additionally, the use of *Transformer* and *Transformer-like* deep learning algorithms for *Speech Separation* tasks have been steadily increasing due to their ability to model dependencies in sequential data for *Sequence Transformation*. With our experimentation thus far, we have shown that our proposed *entropy minimization*-based *nested optimization* training method generally outperforms the more standard (*k-means clustering*-based) *nested optimization*. With the *pre-benchmarking* assessment from our 6^{th} run of experimentation, we noted that our proposed method gave *three* strong indicators of *proper model functionality*; however, there was *one* strong indicator that the *complexity* of our base experimental model was *insufficient*. With the *post-benchmarking* assessment from our 7^{th} and final run of experimentation, we noted that our proposed training method *matched* the benchmark on the metric of *separation performance*, and *outperformed* the benchmark on the metrics of *training stability* and *data usage efficiency*, but *was outperformed* on the arguably least important metric of *compute time*—we noted that, for both training methods, the model's actual *separation latency* was the same. We conclude that the experiments conducted thus far provide a *sturdy foundation* for subsequent experimentation with *longer* mixture sequences, involving novel combinations of Complex Spectral-like Masking and Sequence Transformation-based deep learning, to produce competitive results for various applications of *Binaural* Speech Separation.