Abstract

The task of separating multiple streams of human speech from one mixed audio stream is a difficult one but is State of the art Models for Sound Source Separation have been making steady improvements to the quality of the metrics over the last 5 years. For a while, recurrent network architectures had been pushing the performance in the field, but recently these recurrent approaches were overtaken by a transformer based model. This thesis explores the application of parameter efficient Attention mechanisms within the framework to study the possible advantages of such mechanisms over vanilla transformers.