

IMPLEMENTATION OF REAL-TIME SPEECH SEPARATION MODEL USING TASNET AND DPRNN



PURPOSE :

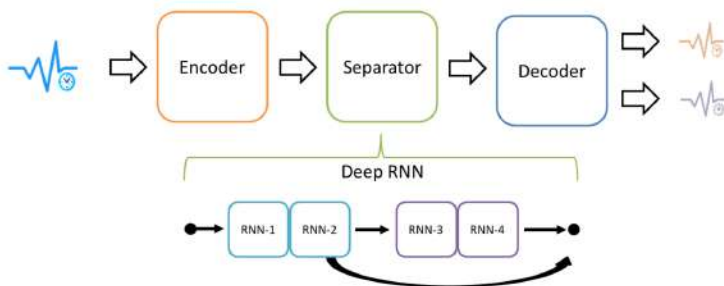
- Build a reliable independent multi-talker speech separation speaker model to do real-time prediction based on the TasNet and DPRNN reference models
- Comparing the implementation of GRU with LSTM and batch size values on TasNet and DPRNN
- Comparing the effect of Adam and Radam's optimizers on TasNet and DPRNN

BENEFIT :

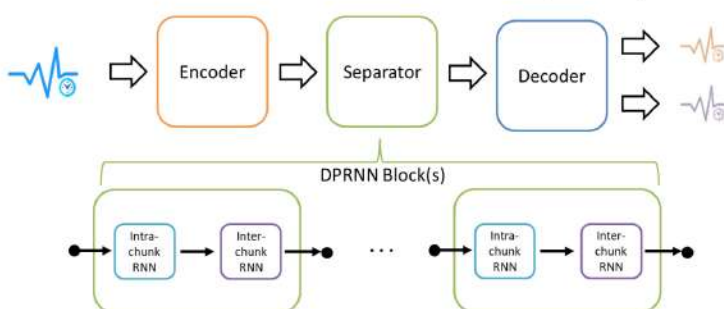
- Knowing the performance and reliability of the TasNet and DPRNN neural network models for conducting online / real-time prediction.
- Determine the effect of the implementation of the GRU and LSTM architecture and the value of batch size parameters on TasNet and DPRNN neural networks.
- Knowing the effect of the implementation of the Adam and RAdam optimizer algorithms on the TasNet and DPRNN neural networks.

NEURAL NETWORK PROCESS

• TIME-DOMAIN AUDIO SEPARATION NETWORK (TASNET)



• DUAL PATH RECURRENT NEURAL NETWORK (DPRNN)



EXPERIMENTAL VARIABLE

BATCH SIZE

OPTIMIZER

RNN

TOTAL EXPERIMENT


TASNET
8 Experiments


DPRNN
2 Experiments

RESULT PREVIEW


TASNET


DPRNN

RNN	Inference Time (ms)	RNN	Inference Time (ms)
LSTM	37.98038292	LSTM	10.43117383
GRU	28.26099630	GRU	10.37165984
	25.59%		0.57%

SUMMARY PREVIEW

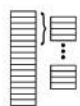
REAL-TIME PREDICTION

  
TASNET Qualified DPRNN


INFERENCE TIME

  
GRU better than LSTM

BATCH SIZE

 Smaller batch sizes, do not guarantee solving the generalization problem, because the batch size doesn't match with the sweet spot of the experiment model configurations.

OPTIMIZER

 The RAdam Optimizer does not meet performance improvements during training, because the RAdam optimizer does not match with the Learning Rate value during experiment.