

# Speech Enhancement with Cycle-Consistent Neural networks

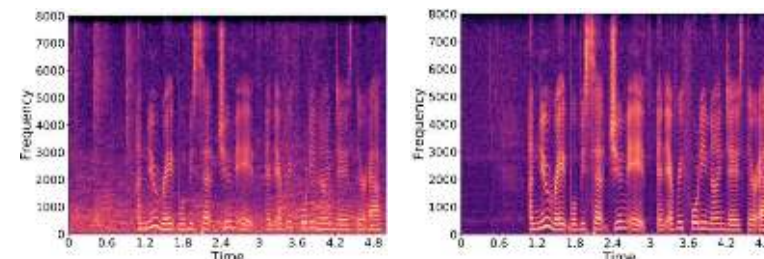
Author: Pavol Karlík

Supervisor: Kateřina Žmolíková

Department of Computer Graphics and Multimedia, Faculty of Information Technology, BUT

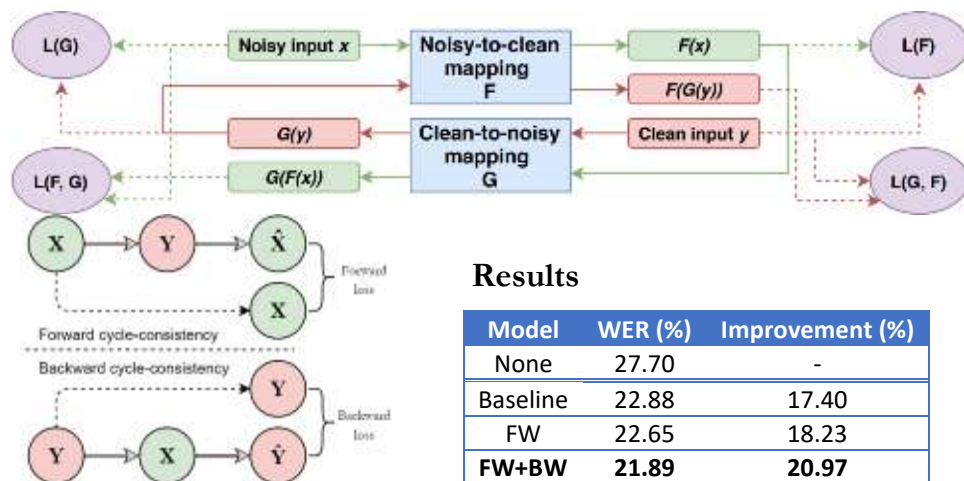
## Motivation

- Noise in speech recordings reduces the effectiveness of Automatic Speech Recognition (ASR) systems
- State-of-the-art speech enhancement systems use neural networks to remove noise
- Neural network models can further be strengthened by employing **cycle-consistency constraint**



## Cycle-Consistent Neural Network

- Uses a **second** neural network with the **opposite** goal during training
- First, the noisy speech signal is enhanced using a neural network
- Then, then noise is inserted back to that enhanced speech signal using the second network (**forward cycle-consistency**) or vice versa (**backward cycle-consistency**)
- The networks are pre-trained separately
- After initialization, they are trained simultaneously with the use of cycle-consistency losses

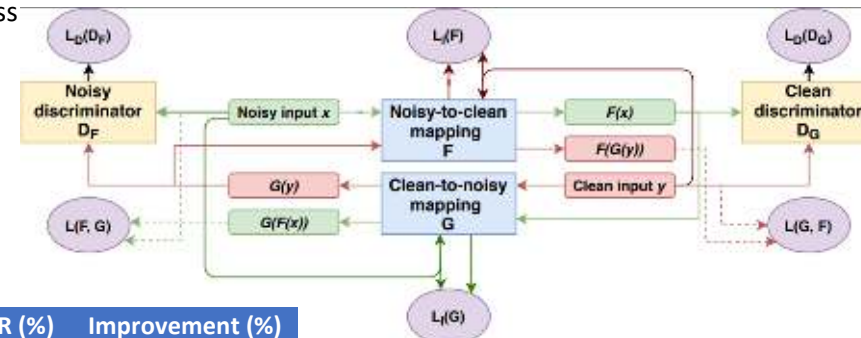


### Results

Model	WER (%)	Improvement (%)
None	27.70	-
Baseline	22.88	17.40
FW	22.65	18.23
<b>FW+BW</b>	<b>21.89</b>	<b>20.97</b>

## Cycle-Consistent GAN

- Sometimes it can be difficult to have a specific dataset of input-label pairs
- Generative Adversarial Networks (GANs) can be used to train model with unpaired data
- However, training GANs can be rather challenging, as it involves multiple DNNs
- We coupled GAN with cycle-consistency and identity-mapping constraints for adversarial speech enhancement in order to improve the unsupervised training process



### Results

Model	WER (%)	Improvement (%)
None	27.70	-
<b>GAN</b>	<b>24.18</b>	<b>12.10</b>

## Acoustic model re-training

- We further re-trained acoustic models with data enhanced using trained models to obtain even better results
- When the acoustic model is re-trained with noisier data, the ASR system performance improves significantly

### Results

Model	Acoustic model	WER (%)	Improvement (%)
None	Clean	27.70	-
FW+BW	Clean	21.89	20.97
<b>FW+BW</b>	<b>FW+BW</b>	<b>18.42</b>	<b>33.50</b>
GAN	Noisy	24.18	12.10
<b>GAN</b>	<b>GAN</b>	<b>14.72</b>	<b>48.86</b>