THE ANDREW & ERNA VITERBI FACULTY OF ELECTRICAL ENGINEERING

SIPL
Signal and Image Processing Lab

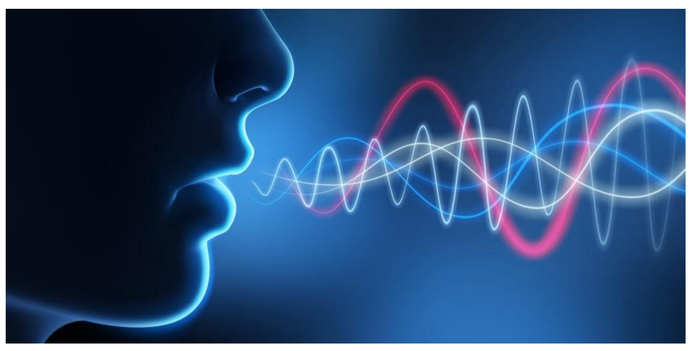
TECHNION Israel Institute of Technology

Speech Enhancement Evaluation Using Speech Recognition algorithm

Yotam Elia , Supervised by Hadas Ofir, Baruch Berdugo

Introduction

- Speech enhancement algorithms aims to improve the quality and intelligibility of noisy speech signals.
- There exist several objective measures for evaluating SE algorithm performance, but none of them gives a faithful representation of MOS.
- In this project we developed an objective measure for SE performance evaluation based on speech recognition.



A USV at sea

Goals

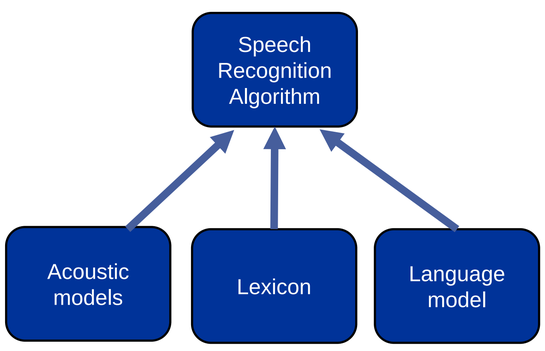
- Develop a tool for objective evaluation of Speech Enhancement algorithm's performance.
- The evaluation is based on using an Automatic Speech recognition algorithm.
- Comparison of spectral-subtruction based algorithms and to deep-learning based algorithms.

Challenges

- The performance of SE algorithm can be evaluated based on several criterions: speech clarity, speech intelligibility and more.
- Different people prioritize those criterions differently.
- Some ASR algorithms deal with noisy environment better than others.

Speech Recognition

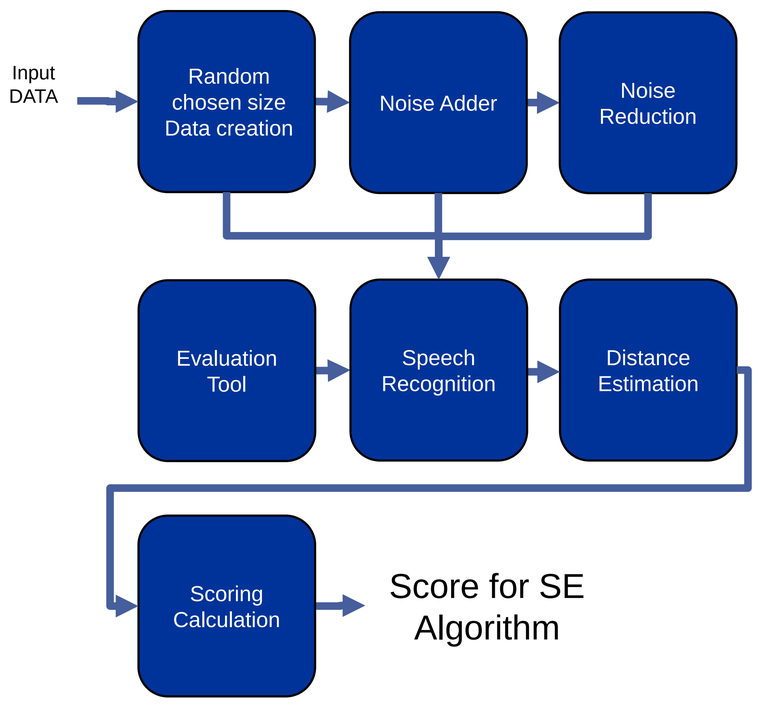
- Speech recognition is based on 3 models.



- In our project we used either Sphinx or Baidu ASR algorithms.
  - The acoustic model of Baidu is based on deep-learning. Uses RNN architecture.
  - The acoustic model of Sphinx is based on hidden markov model.
- Note that Sphinx give more reliable results since it is more sensitive to the presence of noise.

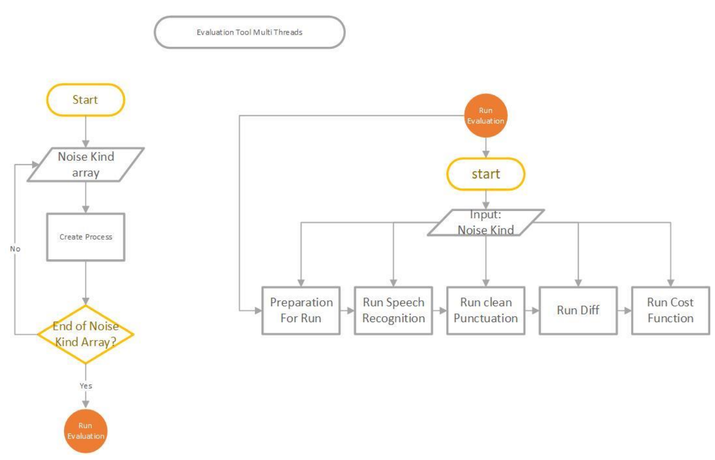
Evaluation SE Algorithm

- Multi-process base algorithms give faster and more accurate solution.
- Results are compared to DNS-MOS for sanity check.



Evaluation Tool

- Consist of speech recognition algorithm, distance estimation and scoring calculation
  - To every noise kind the program create a new process to create fast solution calculation to any process.
  - Give a reliable result and print it to the screen.

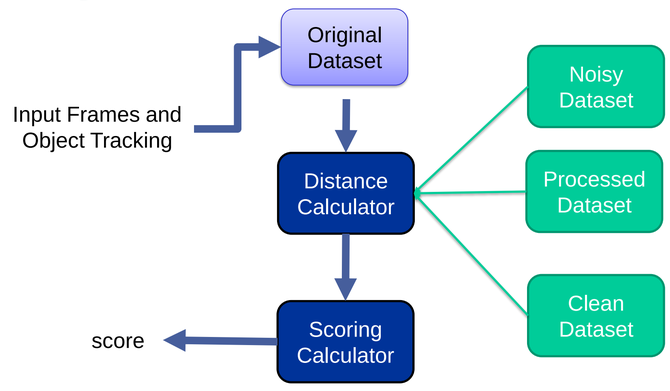


Block diagram showing the process creation and the evaluation procedure.

- The cost of each output is calculated by comparing the text output of the file under examination to the text output of the clean speech.

Distance Estimation

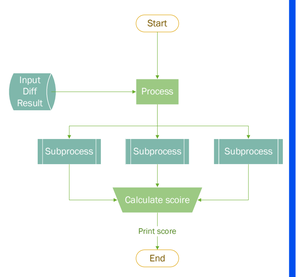
- Write the difference into files, for WER calculation.
- Calculate distance between output text files and original files.



The distance of two text files

Score Calculation

- $WRR = 1 - \frac{Nr - H}{Nr}$
- $Wsuccess = \frac{Wp\%n}{WRRnoise/WRRclean}$



- ***Limits Explained:***
- $\lim_{WRRclean \to 0} \frac{(WRRprocessed - WRRnoise)}{WRRnoise} \to \infty$
- $\lim_{WRRprocessed \to 1} \frac{(WRRprocessed - WRRnoise)}{WRRnoise} \in (0, -1)$

Results

- With Sphinx Algorithm our results is inversely proportional to the DNS-MOS results.
- The results are correlated to the success of the ASR recognition results.

5DB –OMLSA algorithm 5DB – DTLN algorithm

	babble	Mos
Clean Signal:	20.500%	3.70
Noised Signal:	88.100%	2.61
Processed Signal:	85.100%	2.71
Algorithm Score:	3.77	

	babble	Mos
Clean Signal:	20.500%	3.70
Noised Signal:	93.100%	2.13
Processed Signal:	87.500%	2.71
Algorithm Score:	7.04	

Estimated distance of OMLSA vs Estimated distance of DTLN

Sphinx algorithm

Conclusions

- Successful estimation of distance with suitable accuracy for evaluation application.
- Compare between spectral-subtruction algorithms and deep learning algorithms.
- Baidu algorithm have good immunity to noisy environment.
- Baidu algorithm works better when the speech is undistorted by the SE algorithm.

May 2021