Robust Classification of Parkinson's Speech: an Approximation to a Scenario with Non-Controlled Acoustic Conditions

D. A. López-Santander, C. D. Rios-Urrego, C. Bergler, E. Nöth, J. R. Orozco-Arroyave
GITA research group, University of Antioquia
Ostbayerische Technische Hochschule Amberg-Weiden
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
diego.lopez9@udea.edu.co
December 9, 2024

Outline

Context and motivation

Parkinson's Disease

- Neurological disorder prevalent in old age.
- It affects around 1% of people aged 60 or older
- Caused by the loss of dopaminergic neurons.
- Symptoms include loss of muscle control, tremor, behavioural changes.
- Produces speech impairments.

Objectives

Objectives

- To imitate a scenario of pathological speech under noisy acoustic conditions.
- To test a denoising method with varying parameters for reducing the added noise.
- To classify the different signals obtained from the process of denoising in order to determine its effectiveness.

Methods

Spectrogram Representation

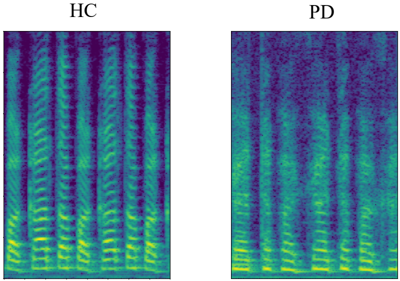


Figure: Comparison between Healthy Control Subject and Parkinson's disease Patient - Pataka

- Segment length: 1.28 s
- Linear frequency scale: 0-10 kHz
- STFS window size: 100 ms
- Hop length: 10 ms
- Spectrogram size: 256×128

Adding Noise

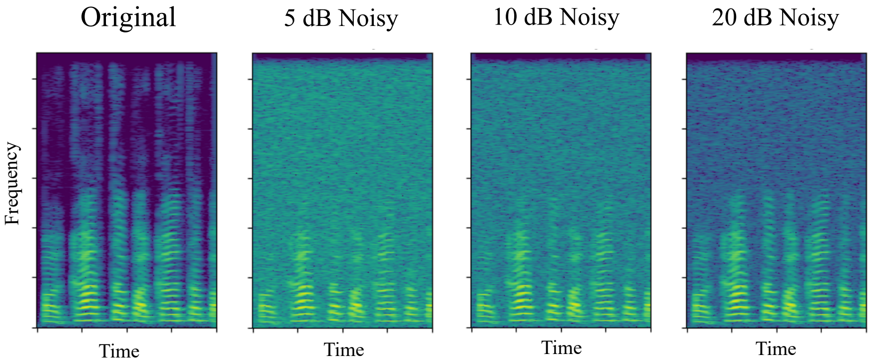


Figure: Different SNR levels spectrogram representation.

Denoising ORCA-CLEAN[1]

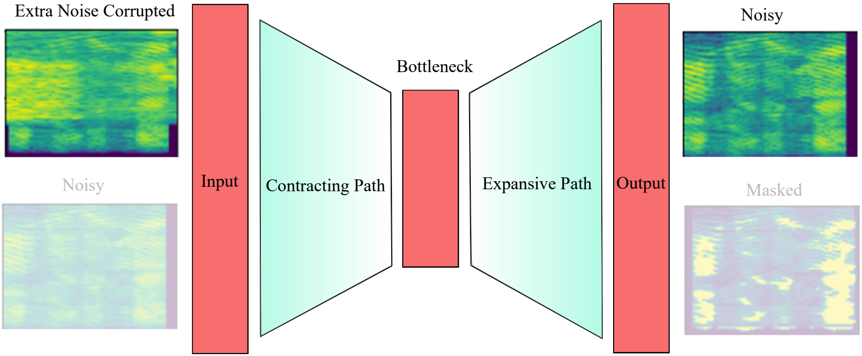


Figure: ORCA-CLEAN Architecture

[1] Bergler, C., Schmitt, M., Maier, A., Smeele, S., Barth, V., & Noeth, E. (2020, October). ORCA-CLEAN: A Deep Denoising Toolkit for Killer Whale Communication. In INTERSPEECH (pp. 1136-1140).

Denoising ORCA-CLEAN[1]

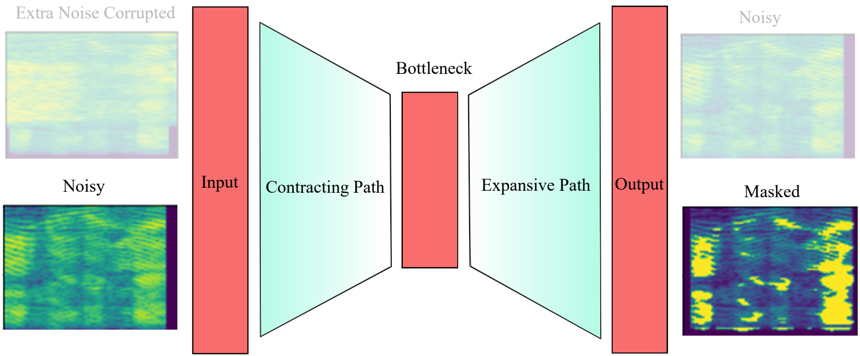


Figure: ORCA-CLEAN Architecture

[1] Bergler, C., Schmitt, M., Maier, A., Smeele, S., Barth, V., & Noeth, E. (2020, October). ORCA-CLEAN: A Deep Denoising Toolkit for Killer Whale Communication. In INTERSPEECH (pp. 1136-1140).

ORCA-CLEAN Binary Masking

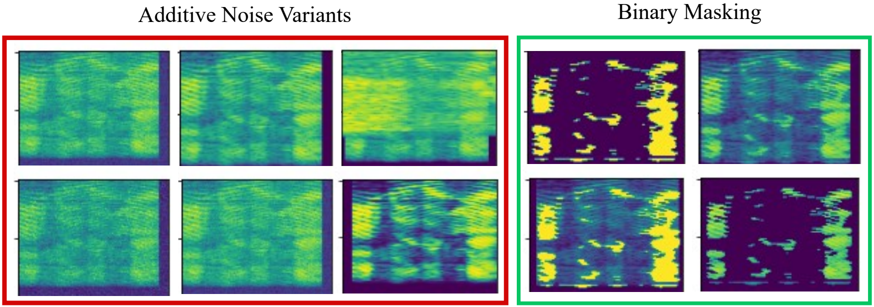
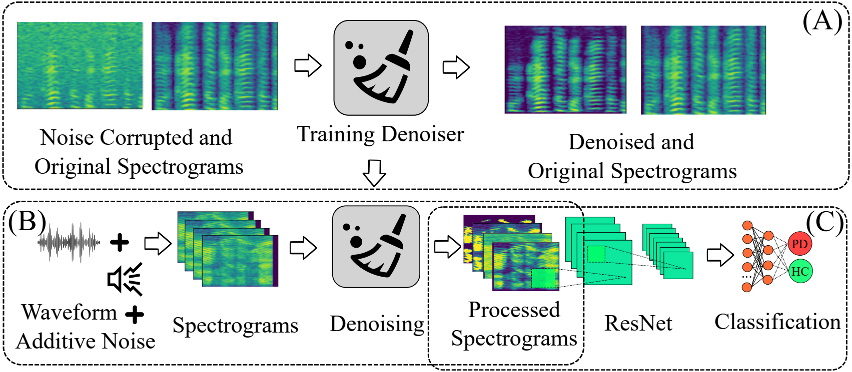


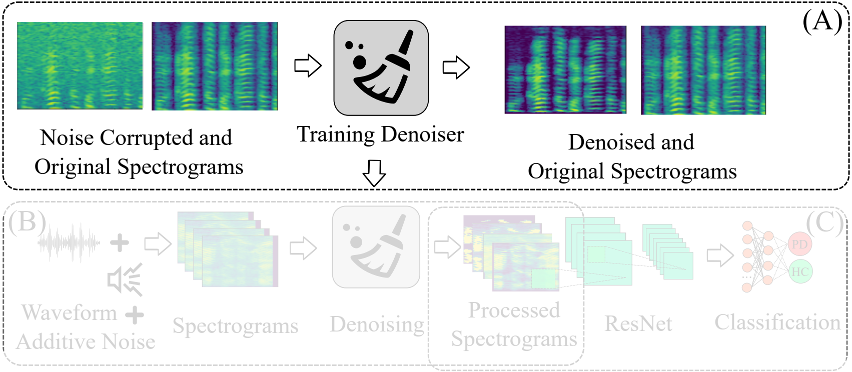
Figure: Binary masking and additive noise examples.

Binary Masking Percentage: 40 %

Methodology



Methodology - Training Denoiser



Datasets for Training Denoiser

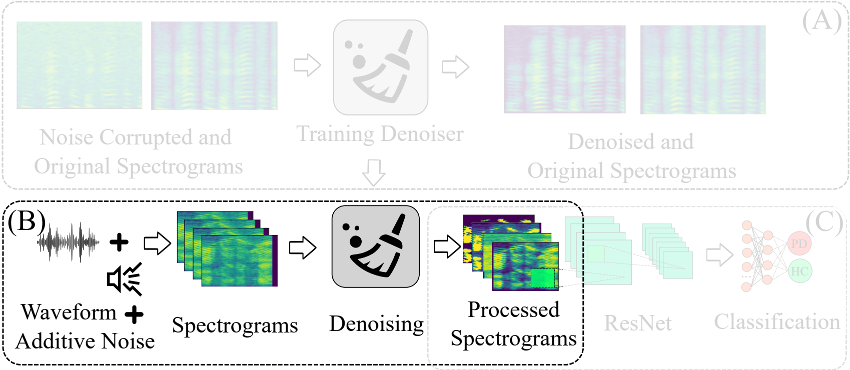
CIEMPIESS

- Spanish dataset
- 16717 audio recordings
- 45 Female, 96 Male
- Avg. duration 25 s.

Microsoft Scalable Noisy Speech Dataset (MS-SNSD)

- Clean Speech
- Environmental Noise

Methodology - Denoising PD Speech



Database - PC-GITA

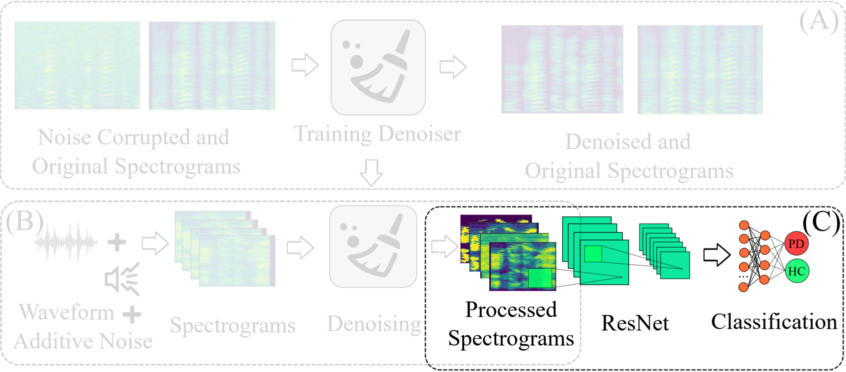
Table: General information of the subjects in PC-GITA

	PD Patients F/M	HC Subjects F/M
Number of Subjects	25/25	25/25
Age [years]	60.7(7)/61.3(11)	61.4(7)/60.5(12)
Time Since Diagnosis [years]	12.6(12)/8.7(6)	
MDS-UPDRS-III	37.6(14)/37.8(22)	

PD patients: Parkinson's patients. HC subjects: Healthy Controls.Values are expressed as mean (standard deviation). F: female. M: male. The total MDS-UPDRS-III ranges from 0 to 132.

Diadochokinetic tasks: "pataka", "petaka", "pakata" .

Methodology - Classification



PD Classification

ResNet Architecture

- 256x128 spectrogram input
- 3 Residual Blocks
- Main blocks with 16, 32, and 64 feature maps
- Fully connected layers: 32 and 2 neurons
- Trained from scratch

Signal Input

- Original
- Noisy
- Denoised
- Residual

Experiments and results

Experiments

- SNR levels:
  a) 5 dB
  b) 10 dB
  c) 20 dB
- Binary Masking Percentages:
  10%, 30%, 40%, 50%, 70%, 90%
- Input Signal:
  a) Baseline - Clean Signals
  b) Noisy Signals
  c) Denoised Signals
  d) Residual Signals
- Each experiment is trained and evaluated following a 10-fold speaker-independent stratified cross-validation strategy.

Binary Masking Percentage Variation

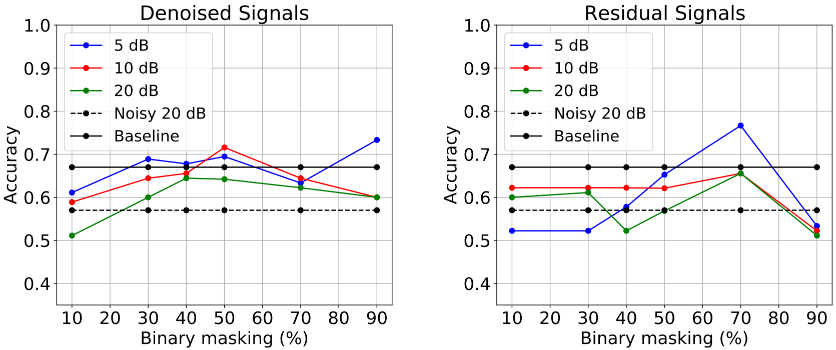


Figure: Classification accuracy of PD patients vs HC subjects for different levels of SNR and binary masking percentages.(A) Denoised Signal. (B) Residual Signals.

Best Results

Table: Classification results for 40% binary masking.

SNR	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-Score(%)
Baseline				
N/A	67.8 ± 16	62.0 ± 36	75.0 ± 30	64.6 ± 24
Noise Corrupted Signals				
5 dB	61.1 ± 11	54.0 ± 35	70.0 ± 33	57.1 ± 23
10 dB	53.3 ± 10	38.0 ± 37	73.0 ± 33	54.7 ± 13
20 dB	57.8 ± 16	56.0 ± 44	60.0 ± 34	51.7 ± 22
Denoised Signals				
5 dB	70.0 ± 8	62.0 ± 12	77.8 ± 10	70.7 ± 7
10 dB	72.6 ± 5	64.0 ± 8	82.2 ± 15	73.4 ± 7
20 dB	63.1 ± 7	42.0 ± 19	84.4 ± 17	67.4 ± 7
Residual Signals				
5 dB	68.5 ± 9	58.0 ± 26	82.2 ± 15	71.9 ± 5
10 dB	67.4 ± 9	48.0 ± 13	88.9 ± 7	72.3 ± 7
20 dB	57.9 ± 13	24.0 ± 26	95.6 ± 5	68.9 ± 7

ROC

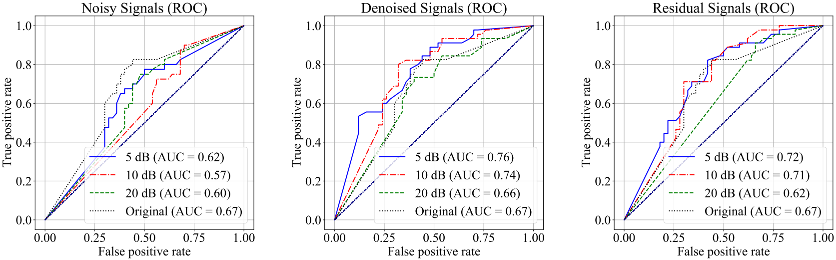


Figure: Receiver operating characteristics

Conclusions

Conclusions

- The proposed approach approximates real-world scenarios where controlling acoustic conditions is not possible.
- Results showed robustness in classification accuracy, even in cases with high levels of noise (5dB)
- The best results of the denoiser were obtained for a binary masking percentage of 40%, with an increase in classification accuracy of around 10% for the three SNR levels w.r.t. the noise corrupted signals.
- The results obtained from the residuals evidence that the denoiser also removes some pathology specific information.

Ongoing work

- Varying input segment length.
- Further adaptation of ORCA-CLEAN model.
- Evaluating other Denoising methods.

Robust Classification of Parkinson's Speech: an Approximation to a Scenario with Non-Controlled Acoustic Conditions

D. A. López-Santander, C. D. Rios-Urrego, C. Bergler, E. Nöth, J. R. Orozco-Arroyave
GITA research group, University of Antioquia
Ostbayerische Technische Hochschule Amberg-Weiden
Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
diego.lopez9@udea.edu.co
December 9, 2024