# Recurrent Neural Networks with Stochastic Layers for Acoustic Novelty Detection

ICASSP 2019, Brighton, UK


## Authors

**Duong Nguyen**[(1)], **Oliver S. Kirsebom**[(2)], **Fábio Frazão**[(2)], **Ronan Fablet**[(1)], **Stan Matwin**[(2)(3)]

(1) IMT Atlantique, Lab-STICC, UBL, 29238 Brest, France
(2) Institute for Big Data Analytics, Dalhousie University, Halifax, Canada
(3) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland



## Acknowledgments

UNIVERSITE BRETAGNE LOIRE / MATHSTIC
A3Data
Lab-STICC
NSERC CRSNG
ACENET accelerate discovery
CominLabs



## References

[1] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller,"A novel approach for automatic acoustic novelty detection us-ing a denoising autoencoder with bidirectional LSTM neural networks," in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2015, pp. 1996–2000.

[2] E. Principi, F. Vesperini, S. Squartini, and F. Piazza, "Acoustic novelty detection with adversarial autoencoders," in 2017 International Joint Conference on Neural Networks (IJCNN), May 2017, pp. 3324–3330.

[3] J. Chung, K. Kastner, L. Dinh, K. Goel, A. Courville, and Y. Bengio, "A Recurrent Latent Variable Model for Sequential Data," in Advances in neural information processing systems, Jun. 2015, pp. 2980–2988.

[4] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, "A Multi-task Deep Learning Architecture for Maritime Surveillance using AIS Data Streams," in 2018 IEEE DSAA, Oct. 2018.


## Associated code

Python code available at:
https://github.com/dnguyengithub/AudioNovelty

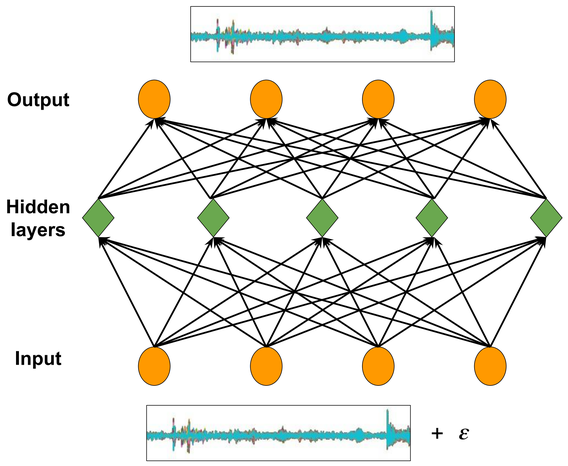

## Abstract

We adapt Recurrent Neural Networks with Stochastic Layers, which are the state-of-the-art for generating text, music and speech, to the problem of acoustic novelty detection. By integrating uncertainty into the hidden states, this type of network is able to learn the distribution of complex sequences. Because the learned distribution can be calculated explicitly in terms of probability, we can evaluate how likely an observation is then detect low-probability events as novel. The model is robust, highly unsupervised, end-to-end and requires minimum preprocessing, feature engineering or hyperparameter tuning. An experiment on a benchmark dataset shows that our model outperforms the state-of-the-art acoustic novelty detectors.

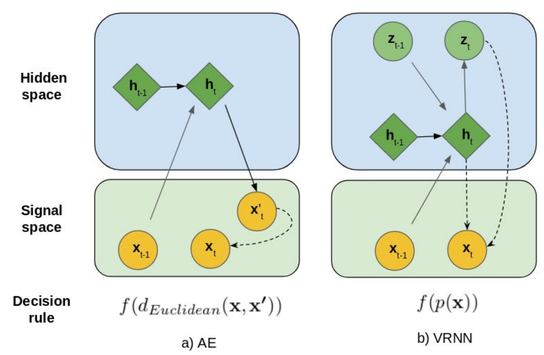

## Introduction

**Objective**: to detect novel/rare events in acoustic recording.

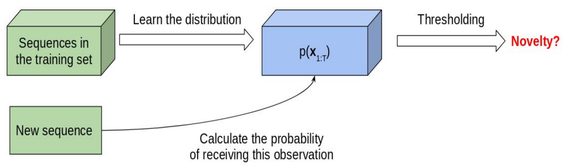
**State-of-the-art**:




## Proposed model




### Architecture




### Decision rule

## Experiments and Results

**Dataset:** We tested our model on the same dataset used in [1] and [2], which is part of the PASCAL CHiME speech separation and recognition challenge dataset. The original dataset contains in-home environment recordings with two children and two adults performing common activities, such as talking, eating, playing and watching television. Abnormal sounds (alarms, falls, breakages of objects, screams) were added digitally.

### Detection result

| Method | Online Processing | Precision | Recall | F1 score |
|---|---|---|---|---|
| GMM | Yes | 99.1 | 87.8 | 89.4 |
| HMM | Yes | 94.1 | 88.9 | 91.1 |
| LSTM-CAE | Yes | 91.7 | 86.6 | 89.1 |
| BLSTM-CAE | No | 93.6 | 89.2 | 91.3 |
| LSTM-DAE | Yes | 94.2 | 90.6 | 92.4 |
| BLSTM-DAE | No | 94.7 | 92.0 | 93.4 |
| Adversarial AE | ? | ? | ? | 93.3 |
| **VRNN** | Yes | 95.4 | 91.8 | **93.6** |
| **VRNN*** | Yes | 95.4 | 92.8 | **94.1** |

**TABLE 1**: Detection result, in comparison with state-of-the-art methods.

### Robustness test

| SNR | Precision | Recall | F1 score |
|---|---|---|---|
| 5dB | 96.0 | 91.2 | 93.6 |
| 10dB | 96.1 | 91.9 | 94.0 |
| 15dB | 96.1 | 92.1 | 94.0 |

**TABLE 2**: Robustness test.

### Insights

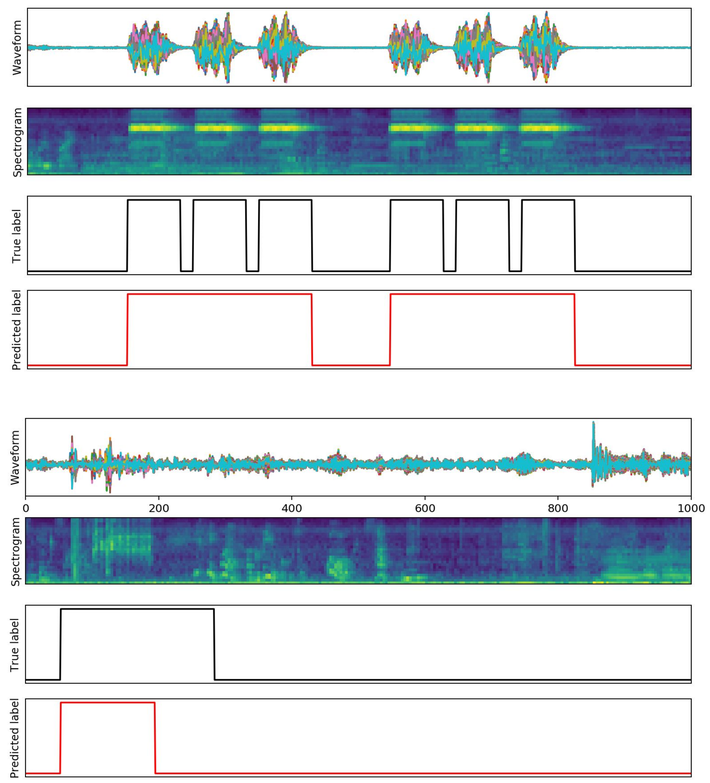



**FIGURE 1**: False detections because adjacent events are merged together (top) and miss detection because the tail of the sound is gradually submerged in the background (bottom).

## Conclusion and Future work

We have presented a novel approach for acoustic novelty detection:

- End-to-end and highly unsupervised
- Highly automatic
- Robust to noise

Future work could involve applying this model to more complex signals, *e.g.* underwater acoustic signals which depict even greater variabilities. The impact of the threshold is also being studied to obtain better threshold selection rule.

Contact: van.nguyen1@imt-atlantique.fr