Automatic bird sound detection in long range field recordings using Wavelets & Mel filter bank features

Suhas

April 2020

PennState
College of Engineering

ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE

The Plan

- Motivation
- Previous works
- Dataset considered
- Features, methods and evaluation
- Some challenges involved
- Results - Where does Wavelets stand against the currently “*popular*” methods?
- Demo
- Q&A

Why are we looking at this topic?

- Bird sound detection from far-field recordings is an important means for understanding the effect of urbanization on our environment [1].

  1. spatial heterogeneity, 2. habitat fragmentation & 3. intermediate disturbance [1,2]

- Can we substitute manual sightings?

[1] M. L. McKinney, “Effects of urbanization on species richness: A review of plants and animals,” Urban Ecosystems, vol. 11, no. 2, pp. 161–176, 2008.
[2] S. Zhang, M. Suo, S. Liu, and W. Liang, “Do major roads reduce gene flow in urban bird populations?,” PloS one, vol. 8, no. 10, 2013.

The story so far

- Earliest works such as Tzanetakis [3] looked at applications : Speech vs music, Voices, Classical music recognition,etc. using DWTC, MFCC, STFTC

- M. Daniels [4] proposed a wavelet based method for percussion sound analysis. (In-house data) with db4, db5, and sym5 wavelets v/s MFCC features + SVM

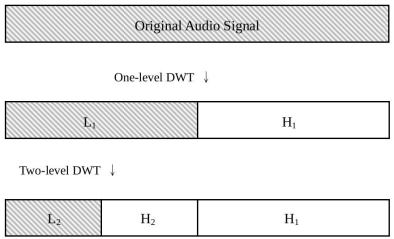
[3] Tzanetakis, G., Essl, G., & Cook, P. (2001, September). Audio analysis using the discrete wavelet transform. In Proc. Conf. in Acoustics and Music Theory Applications (Vol. 66).

[4] Daniels, M. (2010). Classification of Percussive Sounds Using Wavelet-Based. CCRMA, Stanford University thesis.

The story so far

- Muscle Fish dataset, which consists of 410 sounds in 16 classes, is used to evaluate the performance. (Wavelets · FCC + SVM) [5]
- Hsieh [6] proposed a method of extracting wavelet features from audio - reduced the size of recordings. Followed by suprasegmental features.

Feature		Type of transforms	Number of features
Perceptual feature	Subband power P_j	Wavelet	3
	Pitch frequency f_p	Wavelet	1
	Brightness ω_c	Fourier	1
	Bandwidth B	Fourier	1
Frequency cepstral coefficient (FCC) c_n		Fourier	L



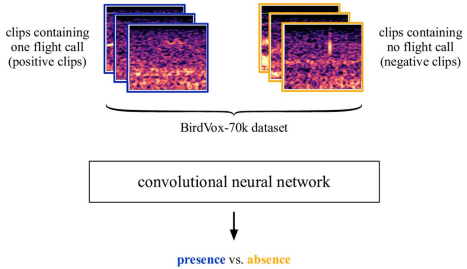
[5] Lin, C. C., Chen, S. H., Truong, T. K., & Chang, Y. (2005). Audio classification and categorization based on wavelets and support vector machine. IEEE Transactions on Speech and Audio Processing, 13(5), 644-651.

[6] Hsieh, S. L., & Wang, H. C. (2005). Feature Extraction for Audio Fingerprinting Using Wavelet Transform. In National Computer Conference.

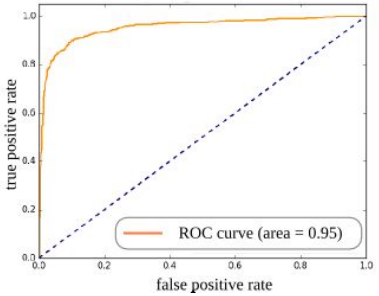
The story so far

- Mel Spectrogram + CNN using Urban-8K dataset and species classification from clips of avian flight calls (CLO-43SD dataset) [7]

  Looked at short term & long term recordings



- SoA : MFBE + CNN using Freefield1010 + Warblr (FF W 1) [8]

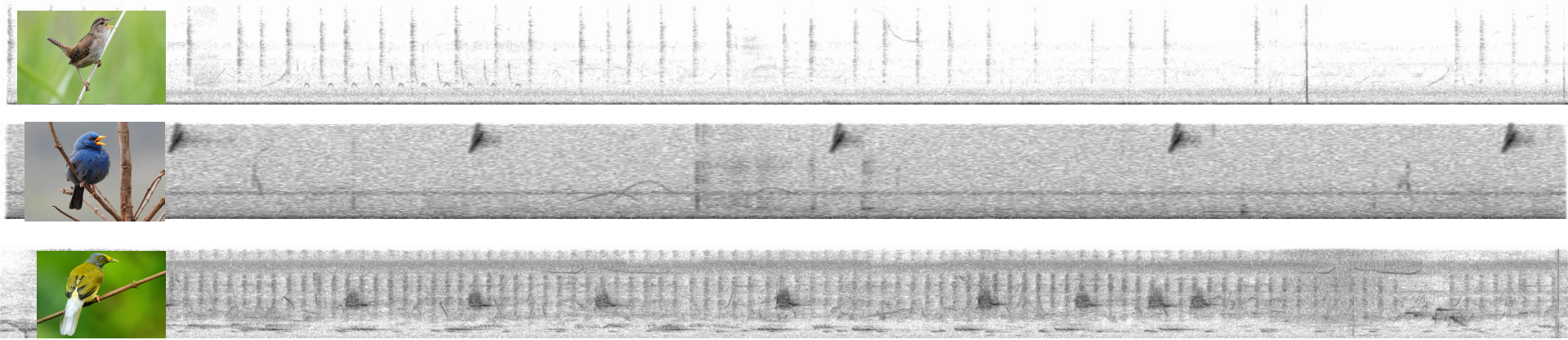


[7] Lostanlen, V., Salamon, J., Farnsworth, A., Kelling, S., & Bello, J. P. (2019). Robust sound event detection in bioacoustic sensor networks. PloS one, 14(10).

[8] Pellegrini, T. (2017, August). Densely connected CNNs for bird audio detection. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1734-1738). IEEE.

Dataset

- Field recordings (freefield1010)
- A collection of over 7,000 excerpts from field recordings around the world, gathered by the FreeSound project, and then standardised for research.
- This collection is very diverse in location and environment.

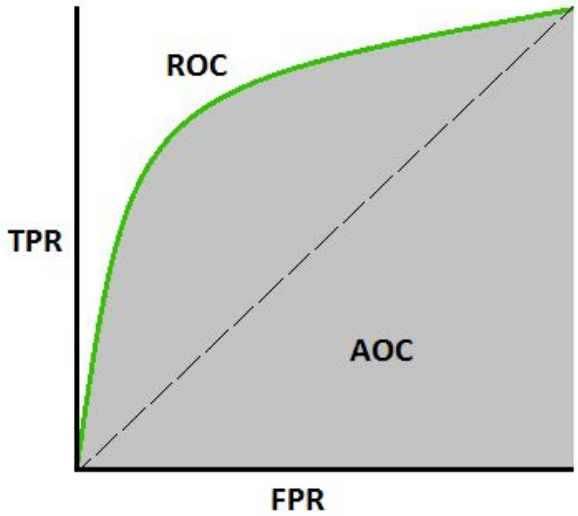


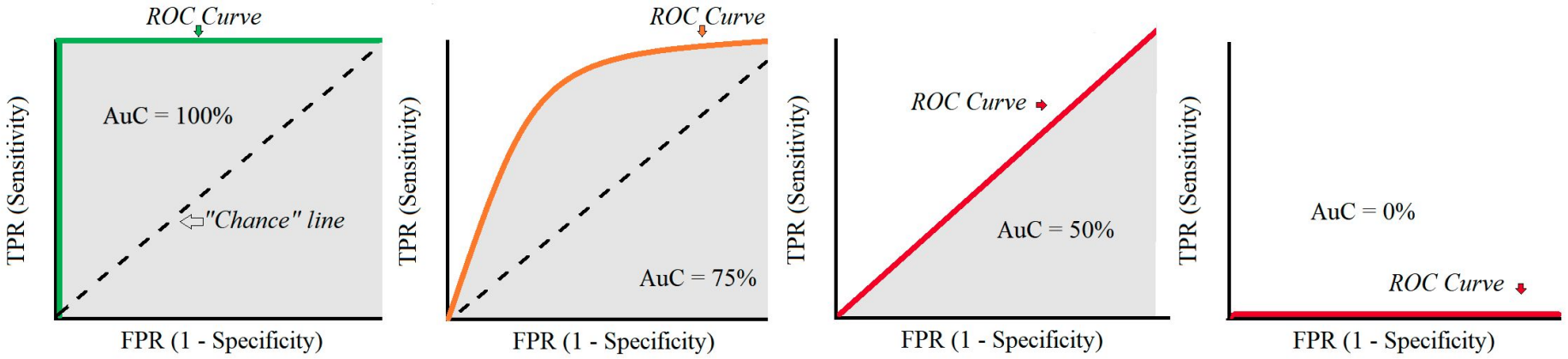
Features, methods and evaluation

- Mel frequency cepstral coefficients
- Mel filter bank energies
- Two - level DWT (sym4, db5)
- Spectrogram features
- Other wavelet based features (db3,db5, sym4, sym6)
- Other baseline papers (p1, p2, p3 & p4)

- SVM
- CNN (SoA)

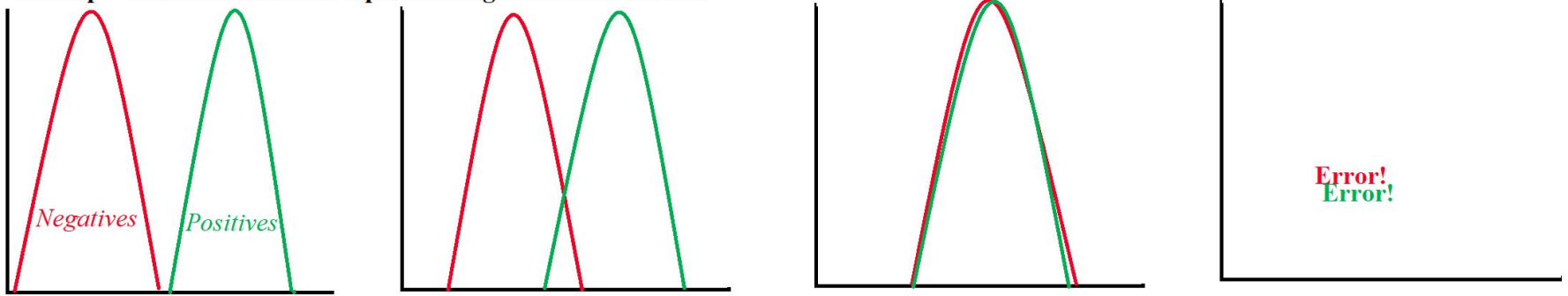
- AUC-ROC curves





Excellent → Good → No Separability → Problematic

Overlap = How well the model separates Negatives and Positives

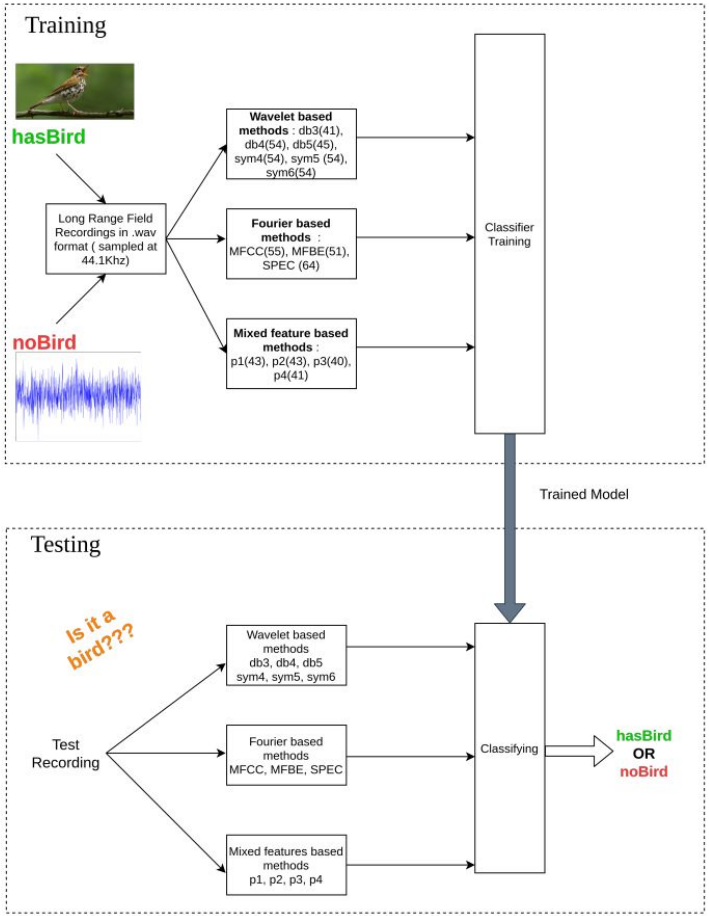


Features and dimensions :

Wavelet based : db3 (41), db4 (54), db5 (45), sym4 (54), sym5 (54), sym6 (54)

Mixed features : p1 (43), p2 (43), p3 (40), p4 (41)

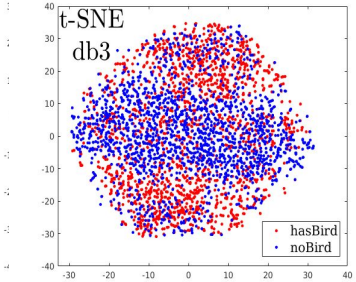
Fourier based : SPEC (64), MFBE (51), MFCC (55)

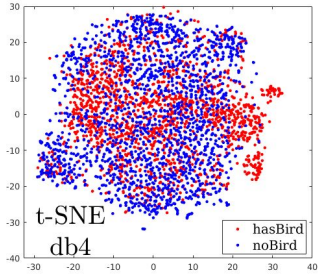


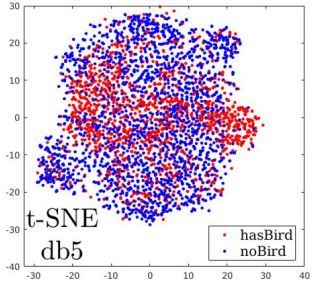
PennState College of Engineering

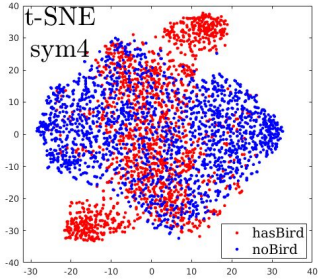
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

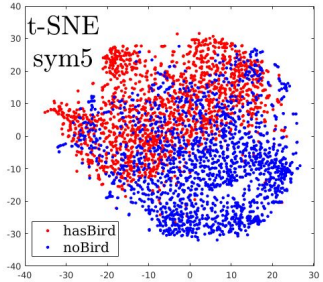
t-SNE for Wavelet based features

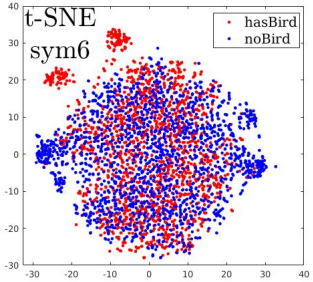




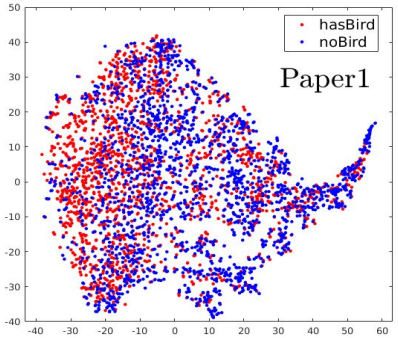


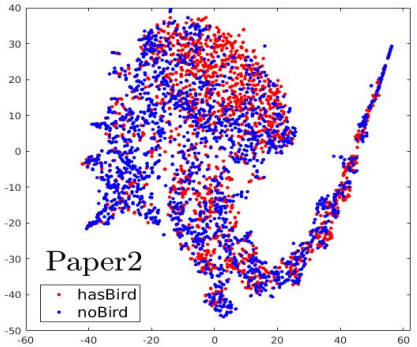


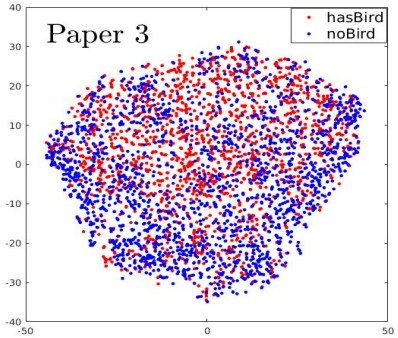


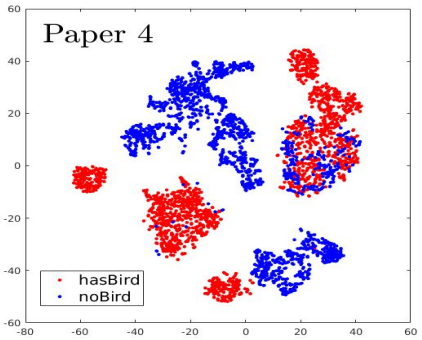


t-SNE for Combined features

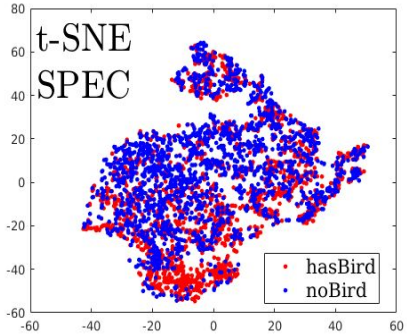


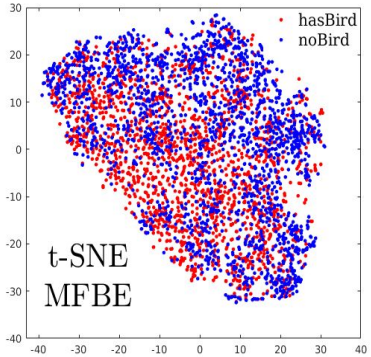


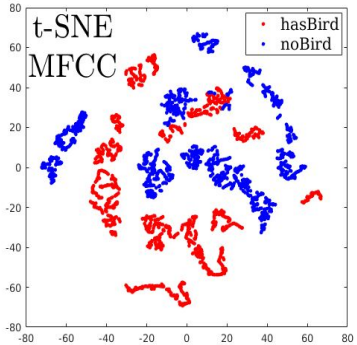




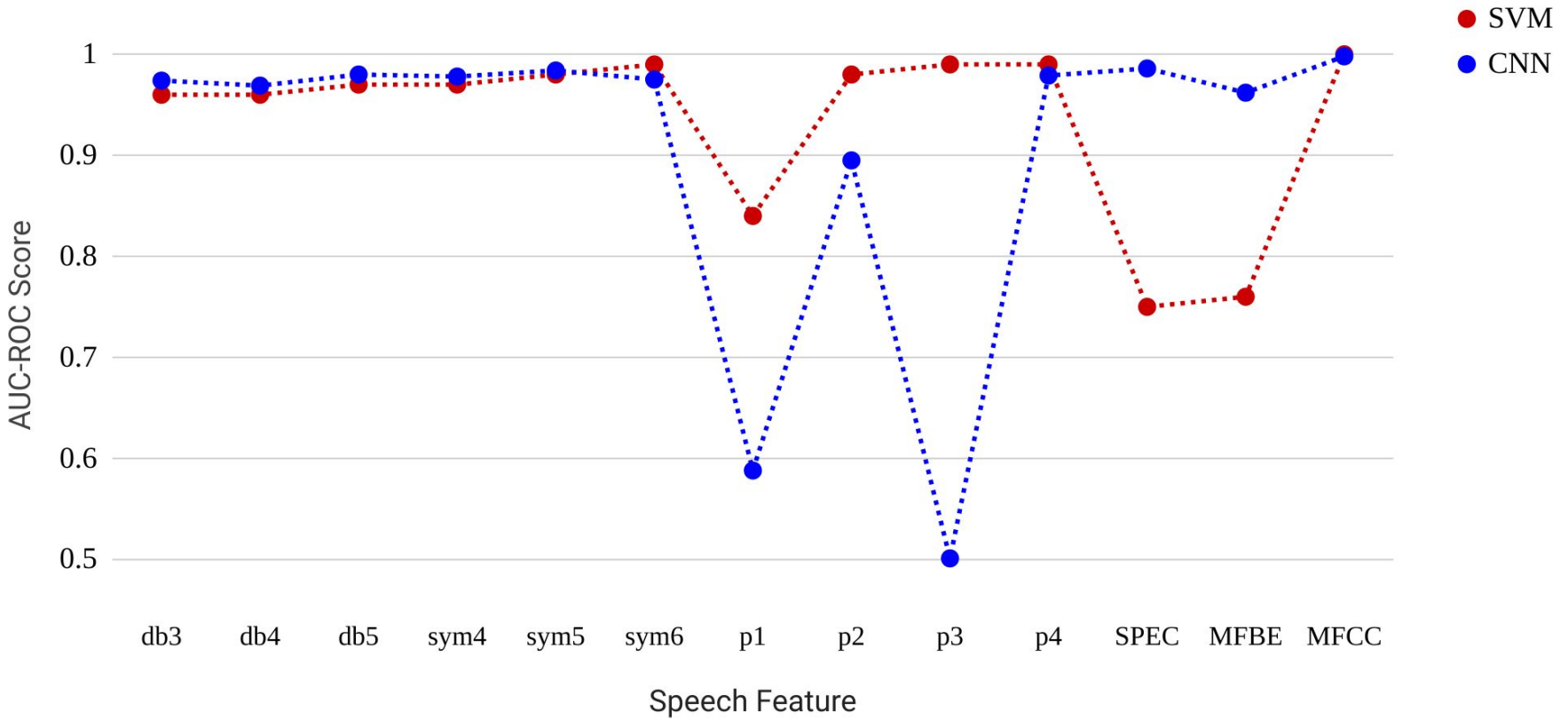
t-SNE for Fourier based features





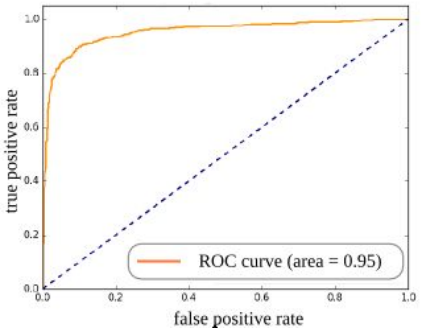


AUC-ROC for different features



How it performs against the baseline (SoA)

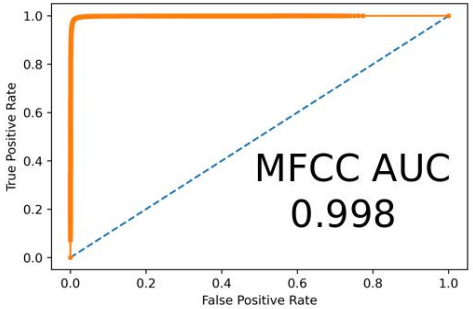
Image taken from paper



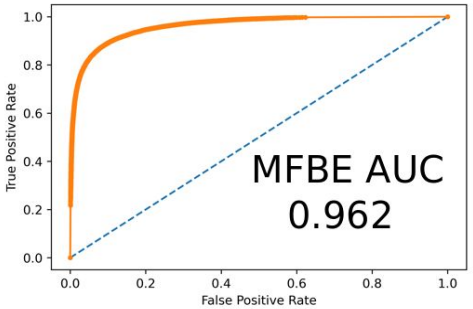
SoA : MFBE + CNN using
Freefield1010 + Warblr (FF W 1) [6]

[6] Pellegrini, T. (2017, August). Densely connected CNNs for bird audio detection. In 2017 25th European Signal Processing Conference (EUSIPCO) (pp. 1734-1738). IEEE.

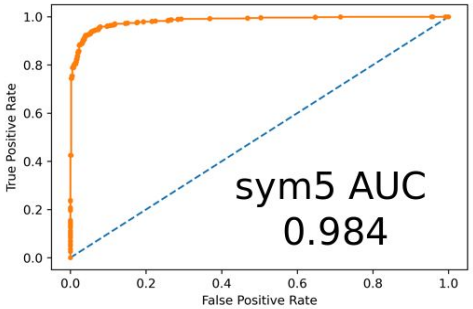
Best Fourier and overall



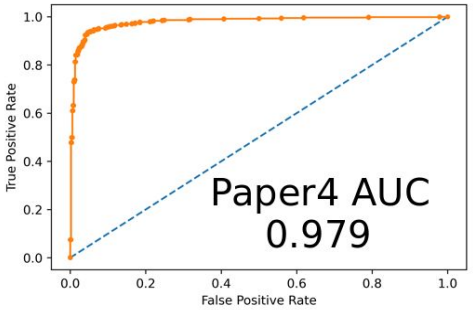
Improvement over baseline



Best wavelet based feature



Best mixed feature



PennState
College of Engineering

ELECTRICAL ENGINEERING
AND COMPUTER SCIENCE

Challenges involved

- freefield1010 : Annotated yet contains a lot of noise!

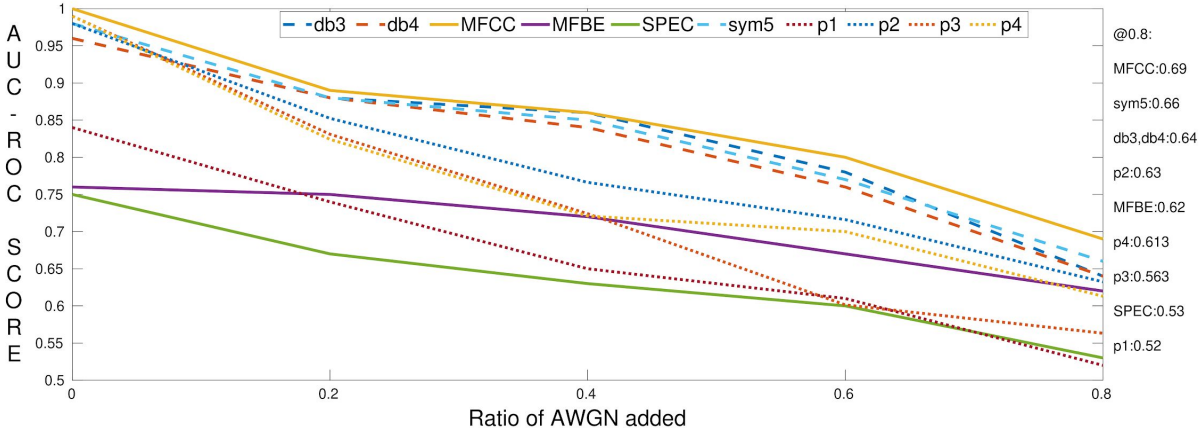
  1. Noise is **Broadband** : overlap sounds that may provide vital information.

  2. **Tedious**

- Simulating such noise?

AWGN

1. How robust are features to AWGN?
2. Does any feature perform better than its counterparts with AWGN than before?




That's all Folks!