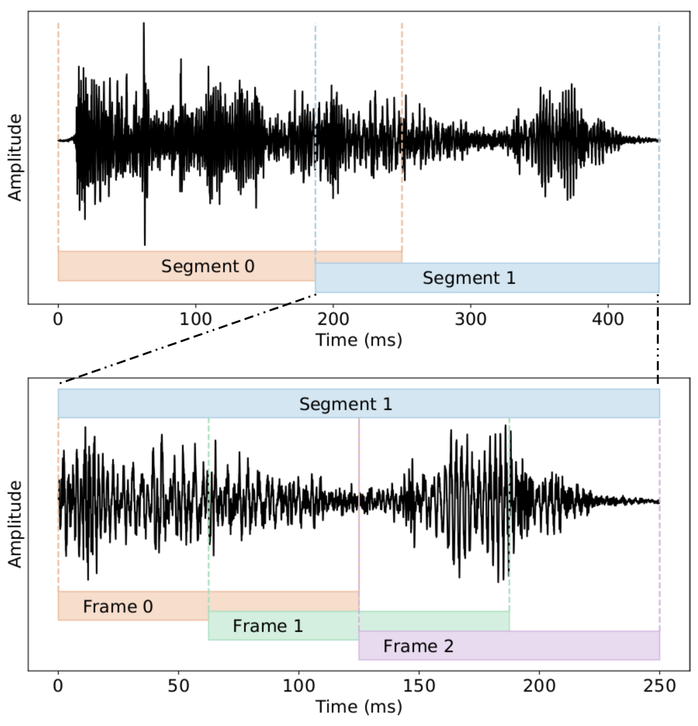
A COMPARATIVE STUDY OF FEATURES FOR ACOUSTIC COUGH DETECTION USING DEEP ARCHITECTURES

Igor Miranda[1], Andreas Diacon[2] and Thomas Niesler[1]

[1]Stellenbosch University, South Africa [2]TASK Applied Science, South Africa

Introduction

- Cough monitoring is key to tracking TB patients.
- The accuracy of automatic cough detectors may be improved with deep learning methods.

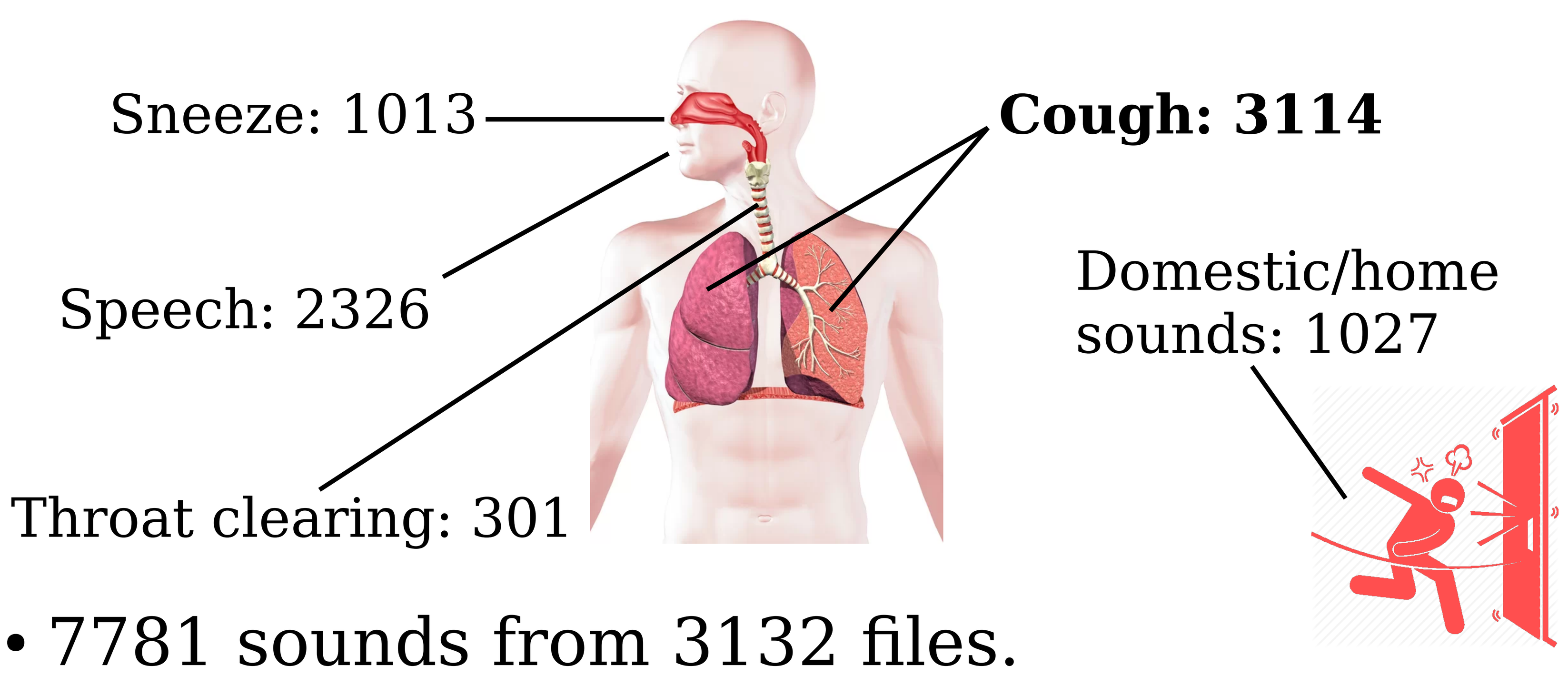


Research Questions

- Which acoustic features are most effective?
- Over which timeframe should they be calculated?
- How best to apply them?

Data Acquisition

- To conduct this evaluation, a meaningful data diversity is needed.
- Our dataset was compiled from:
  - Audio Set (provided by Google)
  - Freesound.org database
- 5 sound categories were considered.



- 7781 sounds from 3132 files.
- 1151 files were used to collect coughs, which suggests a high number of individuals for this particular category.
- Other studies considering coughing generally contain recordings from fewer than 20 people.

Experimental Setup

- Evaluated features: STFT, MFB and MFCC.
- Classifiers evaluated:
  - DNN with three 128-unit hidden layers
  - CNN with one convolutional layer and two 128-unit fully connected layers
  - LSTM with two 832-unit layers
- Two-class softmax output layer for all models.
- Stratified cross-validation applied.

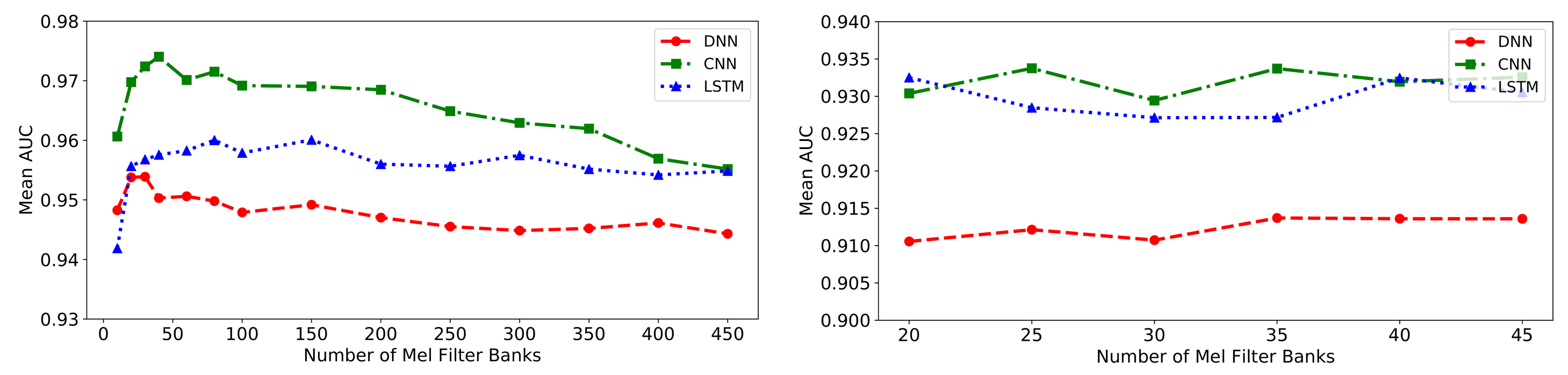
Evaluation of Features and Segment/Frame Durations

Top two cross-validated results for each feature type and segment/frame length (ms).

Classifier	STFT		MFB		MFCC		L-MFCC	
	Seg. / Frame	AUC	Seg. / Frame	AUC	Seg. / Frame	AUC	Seg. / Frame	AUC
DNN	640 / 32	0.943	640 / 32	0.955	640 / 128	0.917	640 / 128	0.935
	640 / 24	0.942	640 / 128	0.953	800 / 128	0.915	800 / 128	0.933
CNN	640 / 24	0.959	800 / 64	0.973	640 / 64	0.933	640 / 64	0.955
	640 / 64	0.958	640 / 64	0.973	640 / 24	0.931	640 / 128	0.951
LSTM	800 / 64	0.950	640 / 64	0.956	640 / 64	0.938	480 / 64	0.934
	640 / 64	0.949	800 / 64	0.955	480 / 64	0.936	640 / 128	0.928

Evaluation of Mel Filter Bank Dimension

Cough classification performance as a function of the filter bank dimension for (a) MFB features and (b) MFCC features.



(a) MFB (b) MFCC

Evaluation of MFCC Derivatives

Cough classification performance in terms of AUC when using MFCCs with and without derivatives.

Feature	DNN	CNN	LSTM
MFCC	0.914	0.933	0.932
MFCC+Δ+2Δ	0.904	0.920	0.905
L-MFCC	0.934	0.954	0.931
L-MFCC+Δ+2Δ	0.931	0.945	0.940

Test Set Results

The baseline system used MFCCs with derivatives and a 25 ms frame shifted by 10 ms.

	DNN		CNN		LSTM	
Feature	Acc.	AUC	Acc.	AUC	Acc.	AUC
Baseline	0.792	0.865	0.843	0.915	0.813	0.863
MFCC	0.805	0.881	0.853	0.925	0.847	0.919
L-MFCC	0.857	0.927	0.876	0.944	0.845	0.918
STFT	0.869	0.932	0.877	0.946	0.873	0.938
MFB	0.883	0.940	0.912	0.965	0.866	0.912

Conclusion

- STFT and MFB features perform best.
- 640 ms segments and 64 ms frames perform well across all classifiers and features.
- As for speech, 40-dimensional mel filter banks provided good results.
- For MFCCs, liftering helps. Derivatives do not.

Acknowledgements

The study was funded by the Lungenliga St. Gallen-Appenzell, Switzerland. We also thank the NVIDIA corporation for the donation of GPU equipment used for this research.