Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

ACCEPTED MANUSCRIPT

Detection of tuberculosis by automatic cough sound analysis

To cite this article before publication: G H Renier Botha *et al* 2018 *Physiol. Meas.* in press https://doi.org/10.1088/1361-6579/aab6d0

Detection of tuberculosis by automatic cough sound analysis

GHR Botha[1], G Theron[2], R Warren[2], M Klopper[2], K Dheda[3], P van Helden[2] and TR Niesler[1]

[1]Department of Electrical and Electronic Engineering
Stellenbosch University, Stellenbosch, South Africa
[2]DST/NRF Centre of Excellence for Biomedical Tuberculosis Research
SU/SAMRC Centre for Tuberculosis Research
Division of Molecular Biology and Human Genetics
Faculty of Medicine and Health Sciences
Stellenbosch University, Stellenbosch, South Africa
[3]Lung Infection and Immunity Unit
Division of Pulmonology and University of Cape Town (UCT) Lung Institute
Department of Medicine
University of Cape Town, Cape Town, South Africa

E-mail: trn@sun.ac.za

November 2016

Abstract. Globally, tuberculosis (TB) remains one of the most deadly diseases. Although several effective diagnosis methods exist, in lower income countries clinics may not be in a position to afford expensive equipment and employ the trained experts needed to interpret results. In these situations, symptoms including cough are commonly used to identify patients for testing. However, self-reported cough has suboptimal sensitivity and specificity, which may be improved by digital detection. This study investigates a simple and easily applied method for TB screening based on the automatic analysis of coughing sounds. A database of cough audio recordings was collected and used to develop statistical classifiers. These classifiers use short-term spectral information to automatically distinguish between the coughs of TB positive patients and healthy controls with an accuracy of 78% and an AUC of 0.95. When a set of five clinical measurements is available in addition to the audio, this accuracy improves to 82%. By choosing an appropriate decision threshold, the system can achieve a sensitivity of 95% at a specificity of approximately 72%. The experiments suggest that the classifiers are using some spectral information that is not perceivable by the human auditory system, and that certain frequencies are more useful for classification than others. We conclude that automatic classification of coughing sounds may represent a viable low-cost and low-complexity screening method for TB.

Keywords: cough audio analysis, tuberculosis, classification

1. Introduction

The World Health Organization reports that, in 2015, 10.4 million people were newly infected with tuberculosis (TB) and that 1.8 million people died of the disease. (World Health Organization 2016). Most new infections occurred in the developing world, with China, India, Indonesia, Nigeria, Pakistan and South Africa accounting for 60%.

The burden of TB is exacerbated by late detection, which is also especially frequent in lower-income countries. For example, a South African citizen from a poor socio-economic background on average experiences a 33-day delay between the first TB symptoms and the first TB test. This delay increases to 90 days for even poorer citizens (Foster et al. 2015). A single infectious TB patient who remains untreated can infect between ten and fifteen people every year (Rossato Silva et al. 2012). Therefore, the consequences of delayed diagnosis and treatment are serious. The recent public health crisis surrounding drug-resistant TB, which is now mostly spread by person-to-person transmission, makes the prevention of transmission even more important (Kendall et al. 2015, Dheda et al. 2017).

Currently there are several methods for diagnosing TB (Knechel 2009). These include smear microscopy, sputum culture and chest radiography. Recently, TB diagnosis has also become possible using rapid molecular tests such as Xpert MTB/RIF, which has been recommended by the WHO since 2011. All of these diagnosis methods are costly in the context of developing countries, and most require specialist skills and in some cases laboratory facilities. Therefore there is a need for an inexpensive and simple method of screening patients for further diagnostic testing (Kik et al. 2014). Globally, a major TB diagnostic research priority is a rapid, non-invasive triage test for screening populations for TB (World Health Organization 2014). Such rule-out tests, which should have high sensitivity, can be used to select patients who test positive and require confirmatory microbiological testing, using for example Xpert MTB/RIF, which is expensive and is of limited availability in high burden settings (Dheda et al. 2013). Cough audio analysis may prove a useful rule-out test.

Tuberculosis is a disease caused by the bacillus *Mycobacterium tuberculosis* (MTB) and usually affects the lungs. The most distinguishing symptoms of TB used to screen patients are chronic coughing (more than 2 weeks), night sweats, fever, weight loss and coughing up blood. The latter is caused by inflammation and destruction of the airways. The bacilli are mainly transmitted by aerosol droplets expelled by the coughing of a person with active TB. A new infection is established when the bacilli enter the macrophages (immune cells) in the alveoli of the lungs.

Since coughing is a dominant symptom of TB, one may hypothesise that the coughing sound of an individual with pulmonary TB can contain information indicative of the disease. This hypothesis is central to the research we present, which considers the automatic analysis of the acoustic quality of cough sounds.

When regarding past work on the automatic processing of coughing sounds, one may distinguish between cough detection and cough classification. Cough detection

deals with the identification and localisation of coughing sounds in general audio (Woolf & Rosenberg 1964, Birring et al. 2008, Larson et al. 2012, Liu et al. 2015, Proaño et al. 2016). After coughs have been detected, their frequency can be calculated and correlated to disease diagnosis or treatment. Cough classification, on the other hand, aims to diagnose specific illnesses or conditions through analysis of cough sounds. The classification by audio analysis of coughing sounds has been an active area of research for some time. Studies have for example investigated the acoustic differences in coughs due to pulmonary diseases such as asthma and acute and chronic bronchitis (Piirilä & Sovijärvi 1989). Other work has considered the automatic identification of wet and dry coughs (Swarnkar et al. 2012). Spectral analysis has revealed differences in the timbre and tonal quality of different types of cough, which in turn can assist a diagnosis (Martinek et al. 2013). A cough with a brassy timbre, for example, has been found to be such a strong characteristic of lymphoid gland tuberculosis that it may suffice as a diagnosis tool in itself (Korpáš et al. 1996). More recently, researchers successfully applied statistical classifiers to distinguish between asthma and pneumonia in children (Abeyratne et al. 2013, Amrulloh et al. 2015) and to detect the distinct coughing sounds associated with pertussis (Pramono et al. 2016).

Our contribution focusses on cough classification, and assumes that cough detection has been reliably achieved. We acknowledge that, by sidestepping the detection step, difficult practical challenges, such as processing of cough spasms, have been left for future work. Using manually segmented cough sounds as ground truth, we present a system that automatically distinguishes between patients with and without tuberculosis using acoustic analysis, logistic regression and classifier fusion. We investigate the performance achieved with different spectral features, and are able to draw some interesting conclusions.

2. Methods

2.1. Audio data acquisition

Our study includes the compilation of a data corpus of coughing sounds recorded in a controlled environment. Recordings of both TB positive patients and healthy controls were made in a specially designed facility.

The clinical diagnosis (ground truth) was established using sputum culturing. Our TB positive patients had culture-confirmed *Mycobacterium tuberculosis* sputum. The healthy controls were contacts of the TB postive patients that had no or few symptoms and were confirmed sputum culture-negative.

All audio was recorded using a Tascam DR-44WL hand-held audio recorder and a Rhode M3 microphone at a sampling rate of 44.1kHz and a resolution of 16 bits per sample. Since patients were prompted, the coughs can be regarded as voluntary. In particular, the healthy controls generally did not have any inherent need to cough. Patients were seated during recording and faced the microphone, which was supported

by a shock mount at distance of between 15 and 40cm.

Coughs were extracted from the recordings by detecting bursts of audio energy delimited by silence, and then manually validating and adjusting the start and end times of the detected region. This resulted in an average of 20 segments of audio per patient, with each segment corresponding to one complete cough.

Table 1. Composition of the audio dataset.

	TB Positive	Healthy controls	Total
Patients	17	21	38
Total coughs	501	245	746
Mean coughs per patient	29.5	11.7	19.6
Std dev coughs per patient	23.3	11.0	19.5
Min coughs per patient	7	2	2
Max coughs per patient	85	45	85
Total time (s)	278.0	99.3	377.3
Mean time per patient (s)	16.4	4.7	9.9
Std dev time per patient (s)	13.0	3.9	10.7
Min time per patient (s)	3.5	0.5	0.5
Max time per patient (s)	52.2	14.1	52.2

The final dataset consists of recordings from 38 subjects of whom 17 were known to be TB positive and 21 were healthy controls, as summarised in Table 1. In total, 746 individual coughs were extracted in this way, resulting in 377 seconds of audio data. All recordings were normalized to account for amplitude differences resulting from variations in the precise distance between the subject and the microphone, as well as naturally differing loudness of the coughs. This was achieved by scaling the amplitudes to ensure that each cough segment had the same sample variance.

2.2. Clinical data acquisition

In addition to the audio recordings, the five objective clinical measurements described in Table 2 were collected for each patient. This choice of features was motivated by the work in (Wejse et al. 2008), which compiled a set of objective clinical measurements based on TB symptoms identified by the WHO (Harries et al. 2004).

Table 2. Objective measurements included as features in this clinical dataset.

Clinical indicator	Description	Units
MUAC	Mid upper arm circumference, measured with a measuring tape.	mm
Temperature	Body temperature, measured with a thermometer.	°C
BMI	Body mass index calculated as $height/weight^2$.	kg/m^2
Anaemic conjunctivae	Paleness of the conjunctiva determined by eye examination.	yes/no
Heart rate	Frequency of heartbeat.	beats/min

The collection of the clinical data was the subject of a larger study for which we had access to a final group of patients for audio recording. Therefore, as indicated in Table 3, the clinical measurements are available for a larger group of patients than are present in the audio dataset described in Table 1.

Table 3. Composition of the clinical dataset. All patients in the audio dataset (Table 1) are also in clinical dataset.

TB Positive	Healthy controls	Total
396	233	518

2.3. Feature Extraction

Cough audio data was characterised by its short-time spectral content for subsequent statistical modelling and classification. First, the recorded audio corresponding to each cough was divided into non-overlapping frames, each consisting of N samples. The value of N was regarded as a parameter to be optimised during training, as described in Section 2.4. Each frame was subsequently parameterised as both a vector of log spectral energies and a vector of mel frequency cepstral coefficients (MFCCs). In both cases the total log frame energy was included as an additional feature.

Some authors have identified consecutive phases within a coughing sound, such as inhalation, forced exhalation and glottal closure (Murata et al. 1998, Thorpe et al. 2001). While the segmentation of the coughs into distinct phases using a hidden Markov model (HMM) topology was considered, this always led to deteriorated performance. We therefore report only on classifiers that use all frames extracted from a cough, and do not distinguish between different phases of the sound.

2.3.1. Log spectral energies Log spectral energies are calculated by multiplicative application of a set of linearly-spaced overlapping triangular filters to the log power spectrum of each frame. Figure 1 shows an example of a filterbank with 10 filters, while Figure 2 illustrates the calculation of the energy for a single filter.

The log spectral energy feature vector therefore represents the log power spectrum of each frame with a reduced frequency resolution determined by the number of triangular filters. This number of filters will be varied during experimental evaluation.

2.3.2. MFCCs Mel frequency cepstral coefficients (MFCCs) also represent the short-term spectral content of a signal (Davis & Mermelstein 1980). However MFCCs do not employ a linearly spaced filterbank as was shown in Figure 1. Instead, the filters are spaced nonlinearly in a way which approximates the frequency resolution of the human auditory system. Figure 3 shows a set of 10 filters spaced according to the mel scale, and indicates how the frequency resolution is higher at lower frequencies. We have considered MFCCs as a parametrisation due to their widespread success in the field of

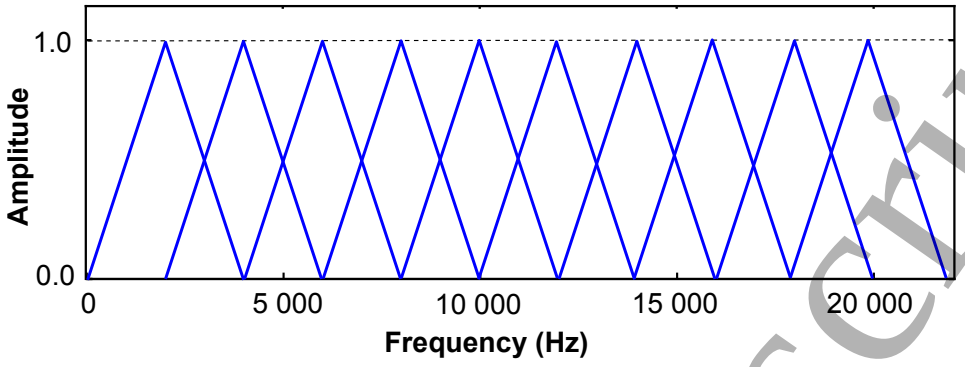

Figure 1. Example of a triangular filterbank with 10 filters as used in the calculation of log spectral energy features.

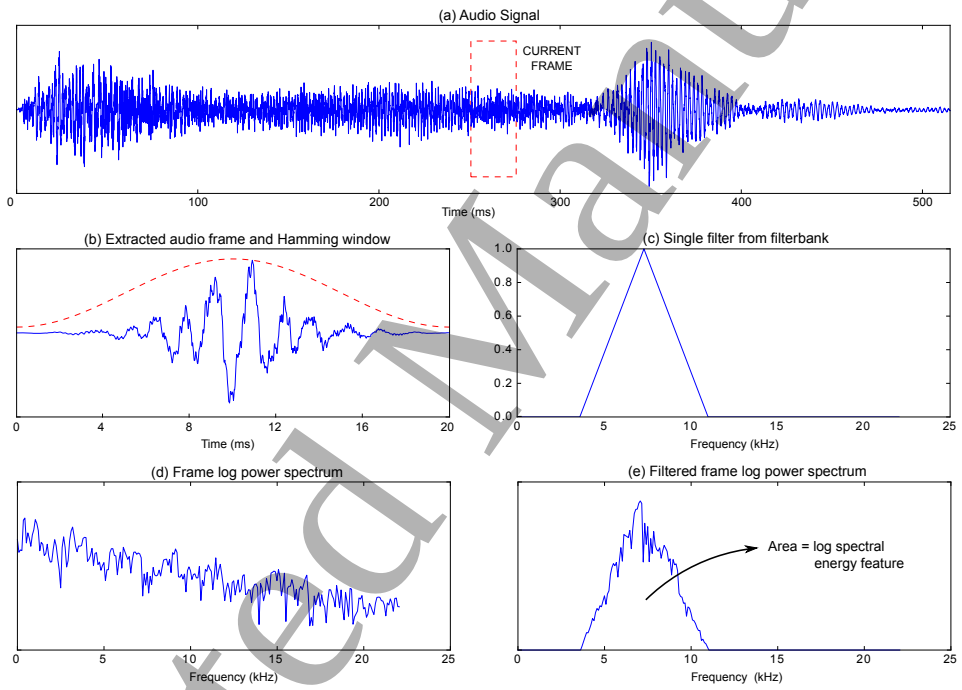

Figure 2. Illustration of the process followed to calculate the log spectral energy for a single filter. (a) The audio signal of a cough indicating the frame to be analysed. (b) The extracted frame after application of the Hamming data window. (c) A single filter from the triangular filterbank. (d) The log power spectrum of the frame in (b). (e) The log power spectrum after multiplication with the triangular filter in (c).

automatic speech processing (Young et al. 2002). Contributing to this success is the ability to easily normalise for convolutive channel effects when using MFCCs by the simple process of cepstral mean subtraction (Moreno et al. 1995). This normalisation step was applied in all our experiments.

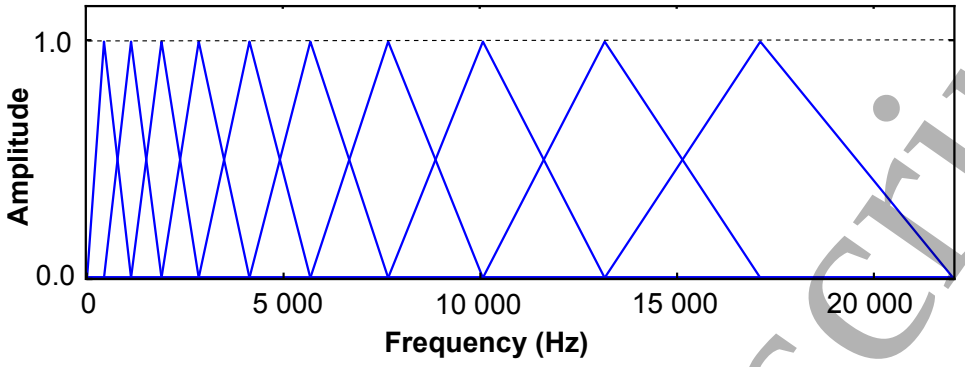

Figure 3. Example of a logarithmically distributed triangular filterbank with 10 filters as used in the calculation of mel-frequency cepstral coefficients.

2.4. Statistical classification

Classification was achieved by logistic regression. Hidden Markov models (HMMs) and decision trees were also considered, but led to consistently and substantially inferior performance, and will therefore not be reported on.

The statistical classifier must determine the probability that an audio frame was produced by a TB positive patient. Assuming that the frame is represented by a d-dimensional feature vector $\mathbf{x} = x_1, x_2, \ldots, x_d$, a two-class logistic classifier would estimate this probability $P_{TB}(\mathbf{x})$ according to the following equation.

$$P_{TB}(\mathbf{x}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \mathbf{x}}}$$

where $\boldsymbol{\theta} = \theta_0, \theta_1, \ldots, \theta_d$ are the $d + 1$ parameters of the classifier. The parameters of the classifier are obtained by maximum likelihood estimation, including an L2 regularisation term λ (Park & Hastie 2008).

Since our dataset is limited, all logistic regression classifiers were trained and evaluated using nested cross validation, as illustrated in Figure 4. An outer 5-fold cross validation loop splits the dataset into a training (80%) and a testing (20%) partition, with no overlap of subjects. This configuration ensures that there are between 3 and 4 TB-positive patients and between 4 and 5 healthy controls in each of the 5 outer testing partitions. Nested within this outer loop, hyper-parameter optimization was achieved by 4-fold cross validation on the training partition, again with no overlap of subjects (inner loop A). These hyperparameters included N, the number of samples in a frame of audio, F, the number of filters applied to the power spectrum during log spectral energy and MFCC computation, M, the number of MFCCs included in the feature vector, and the logistic regression regularisation parameter λ. In order to determine the decision threshold required for an equal error rate, it is necessary to compute a receiver operating characteristic (ROC) within the inner cross-validation loop. Since the partitions used for hyperparameter optimisation are too small for ROC determination, this is achieved separately by 2-fold cross validation on the same outer training partition (inner loop

B). From this ROC, the equal error rate γ_{EE} was determined. The equal error rate is the decision threshold at which the classifier's true positive rate (TPR) equals its false positive rate (FPR), and was used later to determine the classifier performance. This process was repeated for all 5 testing partitions in the outer loop. In this way all data is used for both training and testing, while maintaining complete independence between test and training sets at all times.

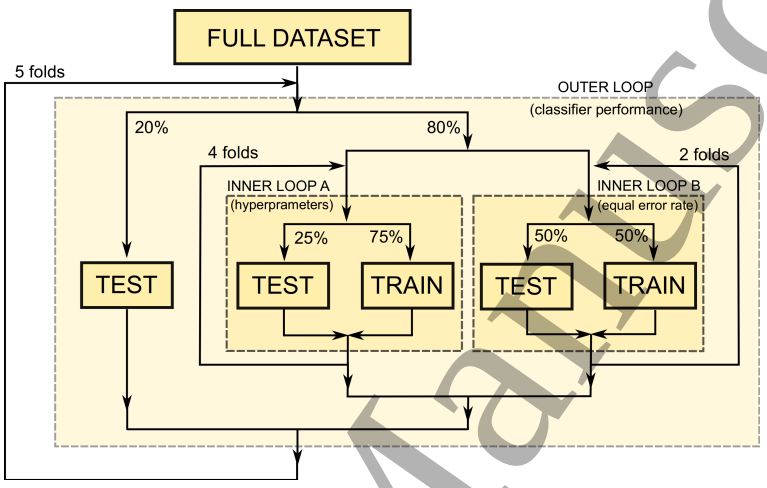

Figure 4. The nested cross validation process used to train and evaluate all statistical classifiers.

In order to apply the classifier to a sequence of frames drawn from a test cough, we calculate the average probability of a TB positive classification S as indicated in Equation 1.

$$S = \frac{1}{L}\sum_{i=1}^{L} P_{TB}(\mathbf{x}_i) \tag{1}$$

In Equation 1 the average is taken over all frames $\mathbf{x}_1 \ldots \mathbf{x}_L$ from all coughs by a specific patient. Hence a patient is diagnosed as TB positive when $S \geq \beta$, with $\beta = \gamma_{EE}$ if the decision is to be made at the equal error rate. This classification strategy is applied to the the 20% testing partition that is reserved for independent testing in the outer loop of cross validation.

2.5. Classifier performance evaluation

The performance of our classifier was evaluated using the well established measures sensitivity, specificity and accuracy, each measured at the equal error rate γ_{EE}. In addition, we present the average area under the ROC curves (AUC) calculated over the five test sets of the outer cross validation loop. The AUC has the advantage of insensitivity to a class imbalance in the dataset (Fawcett 2006). It also gives an indication of classifier performance over a range of decision thresholds, and not just at the equal error rate. Table 1 shows that our data is skewed towards TB positive patients.

2.6. Feature Selection

To investigate the importance of different frequency bands when using log spectral energies as features, we applied the sequential forward search (SFS) algorithm (Devijver & Kittler 1982). This algorithm first identifies the single feature that leads to best classification performance. It then successively adds features to the classifier input that lead to the greatest improvement in performance.

2.7. Classifier Fusion

The decisions made by different classifiers can sometimes be combined to make improved decisions (Moreno-Seco et al. 2006). We make use of such classifier fusion to combine the decisions made by classifiers trained on the audio and the clinical datasets respectively. Weighted voting, the sum and product rules, decision trees and logistic regression were considered as fusion methods. Fusion by logistic regression was found to lead to best performance and will be reported on.

3. Results

Several classification systems were trained and evaluated according to the process described in Section 2.4. In each case, all hyperparameters were optimised within the inner loop of the cross-validation process, while evaluation took place independently in the outer loop. First, a system using a feature vector comprising MFCCs and frame energy was considered. MFCC features mimic the frequency resolution of the human auditory system, and have been very successful in the field of automatic speech recognition. A classifier using this established feature representation can therefore be regarded as a baseline against which other configurations can be measured. The optimal system, which was found to use $M = 26$ mel-scale filterbanks to compute 13 MFCCs and $N = 1024$ samples per frame, exhibited the performance indicated as System 1 in Table 4. The inclusion of a larger number of cepstral coefficients by omitting the liftering step during MFCC computation, and the appending of velocity and acceleration coefficients as is common practice in automatic speech recognition systems, did not lead to further improvements (Young et al. 2002).

Table 4. Classification performance achieved by various systems.

System	Experiment	Sensitivity	Specificity	Accuracy	AUC
1	Audio dataset (MFCC)	0.40	0.82	0.63	0.71
2	Audio dataset (log spectral energy)	0.62	0.95	0.80	0.81
3	Clinical dataset	0.72	0.69	0.71	0.80
4	SFS on log spectral energies (5 features)	0.60	0.78	0.78	0.95
5	Fusion	0.82	0.81	0.82	0.95

Next, a corresponding system using linearly-spaced log spectral energies as features was considered. The optimum configuration used $F = 140$ filters and $N = 2018$ samples per frame, and its performance indicated as System 2 in Table 4. We note that log spectral energy features are superior to MFCCs in terms of all performance measures listed.

Both systems 1 and 2 base their classification decisions exclusively on audio data. To determine the relative usefulness of the clinical measurements, logistic regression classifiers were trained on the clinical data described in Table 3 using the features listed in Table 2. The performance achieved by this approach is given by System 3 in Table 4. We note that the classifier trained on the clinical dataset achieves a similar AUC but slightly lower accuracy than the audio classifier using log spectral energies.

Above we found that a system using log spectral energies as features achieved better performance than one using MFCCs. This may be due to the low frequency resolution of MFCCs in the upper part of the spectrum. We were interested to determine which part of the frequency spectrum provides the most useful information for classification. To do this, sequential forward search, as described in Section 2.6, was applied to train classifiers using log spectral energy features. Figure 5 shows how the performance of the classifier changes in terms of AUC as the number of features increases from 1 to the full set of 140.

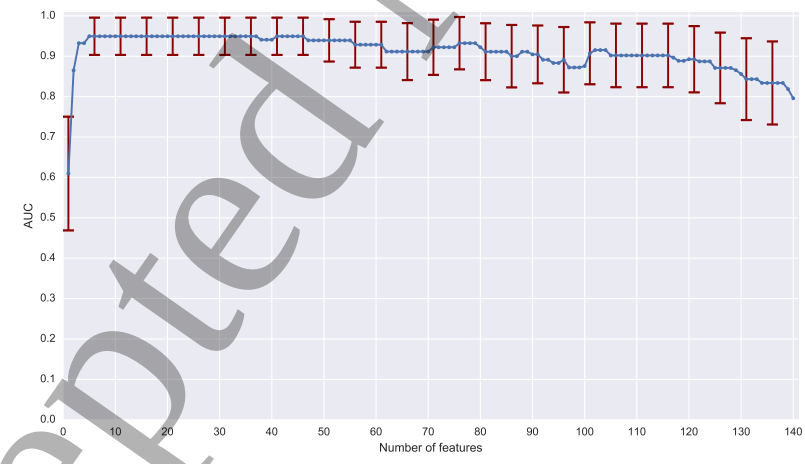

Figure 5. Classifier performance in terms of AUC during sequential forward search (SFS). The horizontal axis indicates the number of log spectral energy features included in the feature vector by SFS. The vertical lines indicate the standard deviation of the AUC, calculated over 100 bootstrap samples (Bisani & Ney 2004).

We note that, starting from a single feature, performance increases steeply and quickly reaches a plateau. The curve in Figure 5 is not monotonic because SFS occurs on a development set while AUC computation occurs on independent test partitions. Best performance is achieved using just 5 features. These consist of the energies in the filters centred around frequencies 236, 550, 10418, 79 and 4894 Hz, listed in the order

determined by SFS.

As a final experiment, the clinical classifier (System 3) was combined with the best audio classifier (System 4), again by means of logistic regression. The performance of the fused system is given as System 5 in Table 4. We observe that classifier fusion leads to further improvements in terms of accuracy but not AUC.

Figure 6 shows the receiver operating characteristic (ROC) for System 5 in Table 4. The ROC shows that, by varying the decision threshold used during classification, there is a tradeoff between sensitivity and specificity. Assuming a requirement of 95% sensitivity, we see that the system exhibits a specificity of approximately 72%. In this configuration the system would be expected to detect 95% of true cases, while falsely labelling 28% of true negatives as positive. Such performance may be very useful for early-stage screening purposes, particularly in view of the method's simple application and low cost.

3.1. Discussion

A comparison of the performance of Systems 1 and 2 in Table 4 shows that better performance is achieved with a linearly spaced log spectral energies than with MFCCs. This suggests that the reduction in spectral resolution based on perceptual models of the human auditory system used by MFCCs discards information that is useful for classifying cough sounds. One can therefore speculate that the system is basing its decisions on audio phenomena in the recorded sound that are not perceivable by a human listener.

Since System 2, which uses exclusively audio information, outperforms System 3, which uses exclusively clinical measurements and has a larger dataset, it appears that the audio data is more information rich than the objective clinical measurements. This indicates that the use of audio measurements as a diagnosis tool for TB screening is a viable method in resource-poor settings.

When testing the usefulness of different frequency bands for classification using SFS in Figure 5, we found that the addition of some features actually leads to a degradation in classification performance. In addition, the standard deviation of the AUC for the model trained on the five best features is lower than that of models trained on a larger number of features, indicating that this model is more consistent in its classification accuracy. Hence we can conclude that some frequency bands add uncertainty to the classification task, and should best be omitted from the feature vector.

Finally, we see that the fusion of System 3 and System 4 by means of logistic regression led to improvements in terms of sensitivity, specificity and accuracy. Hence there seems to be scope for augmenting the audio measurements with objective clinical measurements for the purpose of TB detection.

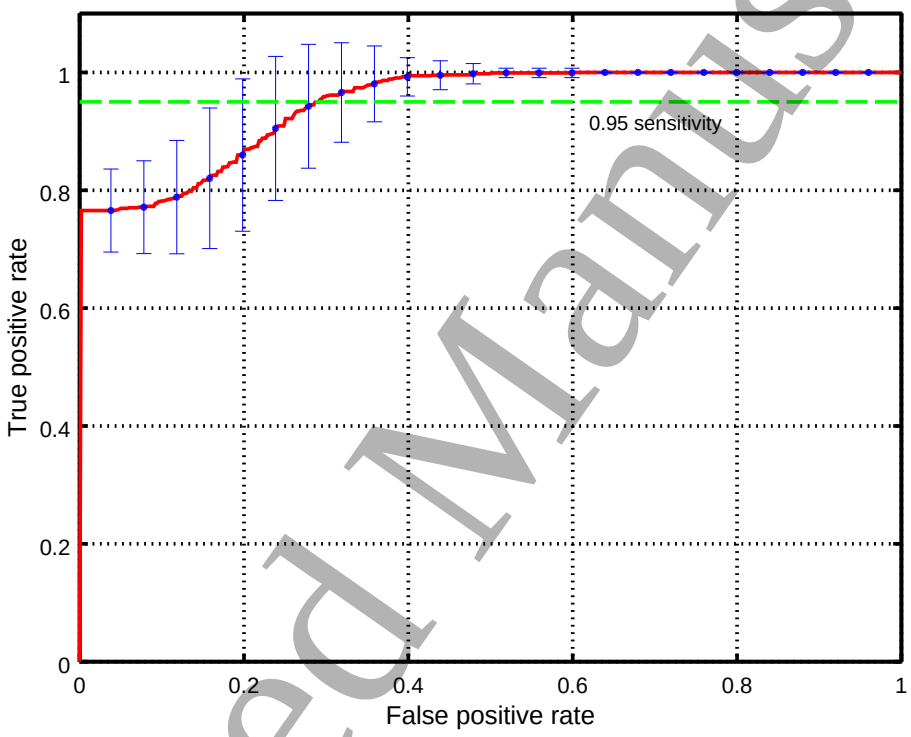

Figure 6. Reciever operating characteristic for System 5 in Table 4. The vertical bars indicate standard deviation as calculated by bootstrap estimation. Note that True Positive Rate = Sensitivity and False Positive Rate = 1− Specificity.

4. Conclusion

We have considered the automatic classification of audio recordings of coughing sounds for the purpose of detecting tuberculosis (TB). Our experiments indicate that

classification of short term spectral features using logistic regression can distinguish between coughing sounds by TB positive patients and healthy controls with 78% accuracy and an AUC of 0.95. Our experiments also indicate that the classifier is using some spectral distinctions that are beyond the limit of human auditory perception. When the spectral features are augmented with five objective clinical measurements, the accuracy improves to 82%. This performance is better than that achieved with the clinical measurements by themselves, indicating that the information contained in both data sources is to some degree complimentary.

We acknowledge the limitations of cohort size and the absence of coughs produced by patients with lung ailments other than TB. However this study was intended as proof of concept, and future studies will include larger numbers of TB patients as well as patients with other lung diseases

Despite these limitations, we could demonstrate a sensitivity of 95% at a specificity of approximately 72%. This performance is already useful for early-stage screening purposes. Furthermore, the technique is simple to apply, and requires neither highly skilled staff for its administration nor expensive laboratory facilities. It is potentially also very low cost, since it can be deployed on readily-available mobile computing hardware such as smartphones or tablet computers. This simplicity and economy makes it attractive in less well resourced settings.

Taken together, these factors imply significant cost saving if automatic cough classification can be utilized to identify non-TB cases that would not need an Xpert test. Additionally, the potential benefits speak to some of the current shortcomings of TB control efforts. For example, the technique has the advantage of obviating the need for sputum samples, which is often problematic in HIV positive patients and children. Furthermore, it is conveniently non-invasive, opening up possibilities for intensified case finding, for example by including all clinic attendees as well as household contacts. Improved early case detection is vital to addressing the global TB epidemic.

References

Abeyratne, U. R., Swarnkar, V., Setyati, A. & Triasih, R. (2013). Cough sound analysis can rapidly diagnose childhood pneumonia, *Annals of Biomedical Engineering* **41**(11): 2448–2462.

Amrulloh, Y., Abeyratne, U., Swarnkar, V. & Triasih, R. (2015). Cough sound analysis for pneumonia and asthma classification in pediatric population, *6th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, IEEE, pp. 127–131.

Birring, S., Fleming, T., Matos, S., Raj, A., Evans, D. & Pavord, I. (2008). The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough, *European Respiratory Journal* **31**(5): 1013–1018.

Bisani, M. & Ney, H. (2004). Bootstrap estimates for confidence intervals in asr performance evaluation, *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, IEEE, pp. I–409.

Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **28**(4): 357–366.

Devijver, P. A. & Kittler, J. (1982). *Pattern recognition: A statistical approach*, Vol. 761, Prentice-Hall London.

Dheda, K., Gumbo, T., Maartens, G., Dooley, K. E., McNerney, R., Murray, M., Furin, J., Nardell, E. A., London, L., Lessem, E. et al. (2017). The epidemiology, pathogenesis, transmission, diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable tuberculosis, *The lancet Respiratory medicine* **5**(4): 291–360.

Dheda, K., Ruhwald, M., Theron, G., Peter, J. & Yam, W. C. (2013). Point-of-care diagnosis of tuberculosis: Past, present and future, *Respirology* **18**(2): 217–232.

Fawcett, T. (2006). An introduction to ROC analysis, *Pattern Recognition Letters* **27**(8): 861–874.

Foster, N., Vassall, A., Cleary, S., Cunnama, L., Churchyard, G. & Sinanovic, E. (2015). The economic burden of TB diagnosis and treatment in South Africa, *Social Science & Medicine* **130**: 42–50.

Harries, A. D., Maher, D. & Graham, S. (2004). *TB/HIV: a clinical manual*, World Health Organization.

Kendall, E. A., Fofana, M. O. & Dowdy, D. W. (2015). Burden of transmitted multidrug resistance in epidemics of tuberculosis: a transmission modelling analysis, *The Lancet Respiratory Medicine* **3**(12): 963–972.

Kik, S. V., Denkinger, C. M., Casenghi, M., Vadnais, C. & Pai, M. (2014). Tuberculosis diagnostics: which target product profiles should be prioritised?, *European Respiratory Journal* **44**(2): 537–540.

Knechel, N. A. (2009). Tuberculosis: pathophysiology, clinical features, and diagnosis, *Critical Care Nurse* **29**(2): 34–43.

Korpáš, J., Sadloňová, J. & Vrabec, M. (1996). Analysis of the cough sound: an overview, *Pulmonary Pharmacology* **9**(5): 261–268.

Larson, S., Comina, G., Gilman, R. H., Tracey, B. H., Bravard, M. & López, J. W. (2012). Validation of an automated cough detection algorithm for tracking recovery of pulmonary tuberculosis patients, *PloS one* **7**(10): e46229.

Liu, J.-M., You, M., Wang, Z., Li, G.-Z., Xu, X. & Qiu, Z. (2015). Cough event classification by pretrained deep neural network, *BMC Medical Informatics and Decision Making* **15**(4): S2.

Martinek, J., Klco, P., Vrabec, M., Zatko, T., Tatar, M. & Javorka, M. (2013). Cough sound analysis, *Acta Medica Martiniana* **13**(Supplement 1): 15–20.

Moreno, P. J., Raj, B., Gouvea, E. & Stern, R. M. (1995). Multivariate-gaussian-based cepstral normalization for robust speech recognition, *The IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, IEEE, pp. 137–140.

Moreno-Seco, F., Iñesta, J., de León, P. & Micó, L. (2006). Comparison of classifier fusion methods for classification in pattern recognition tasks, *Structural, Syntactic, and Statistical Pattern Recognition* pp. 705–713.

Murata, A., Taniguchi, Y., Hashimoto, Y., Kaneko, Y., Takasaki, Y. & Kudoh, S. (1998). Discrimination of productive and non-productive cough by sound analysis, *Internal Medicine* **37**(9): 732–735.

Park, M. Y. & Hastie, T. (2008). Penalized logistic regression for detecting gene interactions, *Biostatistics* **9**(1): 30–50.

Piirilä, P. & Sovijärvi, A. R. (1989). Differences in acoustic and dynamic characteristics of spontaneous cough in pulmonary diseases, *Chest* **96**(1): 46–53.

Pramono, R. X. A., Imtiaz, S. A. & Rodriguez-Villegas, E. (2016). A cough-based algorithm for automatic diagnosis of pertussis, *PloS one* **11**(9): e0162128.

Proaño, A., Bravard, M. A., Tracey, B. H., López, J. W., Comina, G., Zimic, M., Coronel, J., Lee, G. O., Caviedes, L., Cabrera, J. L. et al. (2016). Protocol for studying cough frequency in people with pulmonary tuberculosis, *BMJ Open* **6**(4): e010365.

Rossato Silva, D., Müller, A. M. & de Tarso Roth Dalcin, P. (2012). Factors associated with delayed diagnosis of tuberculosis in hospitalized patients in a high tb and hiv burden setting: a cross-sectional study, *BMC Infectious Diseases* **12**(1): 1–6.

Swarnkar, V., Abeyratne, U., Amrulloh, Y. A. & Chang, A. (2012). Automated algorithm for wet/dry cough sounds classification, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, pp. 3147–3150.

Thorpe, W., Kurver, M., King, G. & Salome, C. (2001). Acoustic analysis of cough, *The Seventh Australian and New Zealand Intelligent Information Systems Conference*, IEEE, pp. 391–394.

Wejse, C., Gustafson, P., Nielsen, J., Gomes, V. F., Aaby, P., Andersen, P. L. & Sodemann, M. (2008). TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course, *Scandinavian Journal of Infectious Diseases* **40**(2): 111–120.

Woolf, C. & Rosenberg, A. (1964). Objective assessment of cough suppressants under clinical conditions using a tape recorder system, *Thorax* **19**(2): 125–130.

World Health Organization (2014). High-priority target product profiles for new tuberculosis diagnostics: report of consensus meeting, Report.
URL: *http://www.who.int/tb/publications/tpp_report*

World Health Organization (2016). Tuberculosis Report 2016, Report.
URL: *http://www.who.int/tb/publications/global_report*

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D. et al. (2002). The htk book, *Cambridge University Engineering Department* p. 175.