Analysis of acoustic features for speech sound based classification of asthmatic and healthy subjects

Shivani Yadav¹, Merugu Keerthana², Dipanjan Gope³, Uma Maheswari K.⁴, Prasanta Kumar Ghosh⁵

¹ BioSystems Science and Engineering, Indian Institute of Science (IISc), Bangalore, India
² Rajiv Gandhi University of Knowledge Technologies, Kadapa, Andhra Pradesh, India
³ Electrical Communication Engineering, Indian Institute of Science (IISc), Bangalore, India
⁴ Pulmonary Medicine, St. Johns National Academy of Health Sciences, Bangalore, India
⁵ Electrical Engineering, Indian Institute of Science (IISc), Bangalore, India

Poster Session: TU2.PB.12: Speech Analysis and Coding
Tuesday, 5 May, 16:30 - 18:30
Introduction

1. Introduction
2. Motivation
3. Dataset
4. Experimental Setup
5. Results
6. Conclusion and Future work
7. References
8. Acknowledgement
Introduction

Objective

- Performance analysis of speech sounds for asthmatic and healthy subject classification by INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) features.
Objective

- Performance analysis of speech sounds for asthmatic and healthy subject classification by INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) features.

Why ISCB features?
Objective

- Performance analysis of speech sounds for asthmatic and healthy subject classification by  INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) features.

Why ISCB features?

1. Iwarsson et al. 1, Dogan et al. 2, Hamdan et al. 3 has showed acoustic characteristics of speech sounds differ between asthmatic and healthy subjects.
Objective

- Performance analysis of speech sounds for asthmatic and healthy subject classification by INTERSPEECH 2013 Computational Paralinguistics Challenge baseline (ISCB) features.

Why ISCB features?

1. Iwarsson et al. 1, Dogan et al. 2, Hamdan et al. 3 has showed acoustic characteristics of speech sounds differ between asthmatic and healthy subjects.

Spectral, cepstral, energy, and excitation/source information all are present in ISCB feature set 4.
What is Asthma??

1. Asthma is an **inflammatory disease** of the airways resulting in a number of symptoms including **obstruction of the airways**, chest discomfort or pain, cough, and wheezes or other peculiar sounds during breathing.

2. 334 million is global burden of asthma\(^2\).

---

Healthy and Asthmatic Airway \(^1\)

---


# Table of Contents

1. Introduction
2. Motivation
3. Dataset
4. Experimental Setup
5. Results
6. Conclusion and Future work
7. References
8. Acknowledgement
Motivation

- **Spirometry** is a lung function test that measures how much and how fast a person can exhale air.
- Helps in **diagnosis** and **monitoring** asthma in hospital.
- Drawbacks of spirometry
  1. Very strenuous
  2. More training required
  3. More time required to screen large population

---

1. [http://www.virtualimaging.org/pft.html](http://www.virtualimaging.org/pft.html)
Motivation

- **Spirometry** is a lung function test that measures how much and how fast a person can exhale air.
- Helps in **diagnosis** and **monitoring** asthma in hospital.
- Drawbacks of spirometry
  1. Very strenuous
  2. More training required
  3. More time required to screen large population

**Is there a simpler approach for asthma diagnosis?**

Motivation

- **Spirometry** is a lung function test that measures how much and how fast a person can exhale air.
- Helps in **diagnosis** and **monitoring** asthma in hospital.
- Drawbacks of spirometry
  1. Very strenuous
  2. More training required
  3. More time required to screen large population

Is there a simpler approach for asthma diagnosis?
Yes, **Sound based analysis** can be.

1. [http://www.virtualimaging.org/pft.html](http://www.virtualimaging.org/pft.html)
Table of Contents

1. Introduction
2. Motivation
3. Dataset
4. Experimental Setup
5. Results
6. Conclusion and Future work
7. References
8. Acknowledgement
Dataset Description

- 95 subjects
  - Patients: 47 (28M, 19F)
  - Control: 48 (24M, 24F)
- Age range: Patients 15-71 years, Controls 19-60 years.
- Stimulus used
  - Speech sounds: /ɑː/ (as in 'After'), /iː/ (as in 'Eat'), /uː/ (as in 'Cute'), /eɪ/ (as in 'Pay'), /ʌ/ (as in 'Only'), /s/ (as in 'Same'), and /z/ (as in 'Zoom')
  - Non speech sounds: cough, wheeze and inhale and exhale.
- Each stimulus was recorded on an average five times per subject.
- Sampling rate: 44.1 kHz.
Table of Contents

1 Introduction
2 Motivation
3 Dataset
4 Experimental Setup
5 Results
6 Conclusion and Future work
7 References
8 Acknowledgement
Experiments & Questions addressed

**Exp1.** *Comparison of ISCB and baseline Mel-frequency cepstral coefficients (MFCC) features performance for Asthmatic and Healthy subjects classification task.*

1. **How** does ISCB features perform **compared** to the baseline MFCC features classification?
2. **Which stimuli** (cough, wheeze and speech sounds) is better for classification?

**Exp2.** *Forward feature group selection for speech sounds.*

1. **Which feature group** in ISCB feature sets selected after forward feature group selection in each speech sound?
2. Is there any **feature group common** across all the speech sounds after forward feature group selection?

**Exp3.** *Forward feature selection in best selected group of Exp2 in all speech sounds.*

1. **Which features** are selected after forward feature selection in best feature group?
Experimental Setup

Experimental setup

- ISCB features computation: analysis window: 20ms, shift: 10ms except for $F_0$ which uses 60ms window with 10ms shift.
- ISCB feature groups into 2 sets: SetA has 5900 features, SetB has 473 features and total 6373 features. Total 21 groups.

Table: List of ISCB groups (number of features).

<table>
<thead>
<tr>
<th>SetA</th>
<th>SetB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G R O U P S</td>
<td></td>
</tr>
<tr>
<td>Loudness (100), Modulated loudness (100),</td>
<td>Fundamental frequency ($F_0$) (83),</td>
</tr>
<tr>
<td>Root mean square (RMS) Energy (100),</td>
<td>Probability of voicing (78),</td>
</tr>
<tr>
<td>Zero crossing rate (ZCR) (100),</td>
<td>Jitter (78), Jitter of Jitter (JJ) (78)</td>
</tr>
<tr>
<td>RASTA auditory bands (2600), MFCC (1400),</td>
<td>Shimmer (78),</td>
</tr>
<tr>
<td>Band energy (200), Spectral Roll Off (400),</td>
<td>logarithmic harmonic to noise ratio (LHNR) (78)</td>
</tr>
<tr>
<td>Spectral centroid (100), Spectral entropy (100),</td>
<td></td>
</tr>
<tr>
<td>Spectral moments (300), Spectral slope (100),</td>
<td></td>
</tr>
<tr>
<td>Harmonicity (100), Spectral Sharpness (100)</td>
<td></td>
</tr>
</tbody>
</table>
Experimental setup

- Support vector machine (SVM) with RBF kernel is used in 5 fold cross-validation setup.
- **Baseline:** MFCC statistics features with SVM as classifier\(^1\).
- Evaluation metric
  - Total classification accuracy

\[
TCA = \frac{TP + TN}{TP + TN + FP + FN}
\]

where, TP (True Positive), TN (True Negatives), FP (False Positives) and FN (False Negatives).

---

# Table of Contents

1. Introduction  
2. Motivation  
3. Dataset  
4. Experimental Setup  
5. Results  
6. Conclusion and Future work  
7. References  
8. Acknowledgement
Exp1. Comparison of ISCB and baseline features

1. How does ISCB features perform compared to the baseline MFCC features for the asthma vs healthy classification task?

- **ISCB features performed better** than the baseline in all stimuli except wheeze.
- All **vowels** sounds showed a **significant improvement** in TCA of at least 10% and by fricatives at least 6%.
Exp1. Comparison of ISCB and baseline features

Which stimuli among speech and non-speech sounds performed best for the classification?

- **Exhale** performed the best among all sounds.
- Among speech sounds /ou/ performed best which is comparable to Exhale.
- /z/ performed poor among all speech and non-speech sounds.
Exp2. Forward feature group selection

Table: Best selected feature groups for speech stimuli.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Best selected feature groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a:/</td>
<td>MFCC</td>
</tr>
<tr>
<td>/i:/</td>
<td>Jitter of the Jitter, MFCC</td>
</tr>
<tr>
<td>/ou/</td>
<td>Loudness, MFCC</td>
</tr>
<tr>
<td>/s/</td>
<td>Loudness, Spectral Entropy, Spectral Flux</td>
</tr>
<tr>
<td>/u:/</td>
<td>MFCC</td>
</tr>
<tr>
<td>/ei/</td>
<td>MFCC</td>
</tr>
<tr>
<td>/z/</td>
<td>Logarithmic Harmonic to Noise ratio</td>
</tr>
</tbody>
</table>

Which feature group in ISCB feature sets selected after forward feature group selection in each speech sounds?

- **Stimuli dependent.**
- **Loudness** and **MFCC** are selected for the best performing stimuli /ou/.

Is there any feature group common across all the speech sounds after forward feature group selection?

- **MFCC** is common across all the **vowels** stimuli.
Exp3. Forward feature selection in best group

Table: Best selected feature in best performing groups for speech stimuli.

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Best selected feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>/ɑː/</td>
<td>Interquartile range between 2\textsuperscript{nd} and 3\textsuperscript{rd} quartile of MFCCs 2\textsuperscript{nd} coefficient</td>
</tr>
<tr>
<td>/ɪː/</td>
<td>1\textsuperscript{st} quartile of MFCCs 2\textsuperscript{nd} coefficient</td>
</tr>
<tr>
<td>/ɔʊ/</td>
<td>Interquartile range between 2\textsuperscript{nd} and 3\textsuperscript{rd} quartile of mel-scale loudness</td>
</tr>
<tr>
<td>/s/</td>
<td>Linear regression 2nd coefficients of mel-scale loudness</td>
</tr>
<tr>
<td>/ʊː/</td>
<td>Linear prediction gain of MFCCs delta 8\textsuperscript{th} coefficient</td>
</tr>
<tr>
<td>/eɪ/</td>
<td>1\textsuperscript{st} percentile of MFCCs 3\textsuperscript{rd} coefficient</td>
</tr>
<tr>
<td>/z/</td>
<td>Skewness of derivative of logarithmic harmonic to noise ratio</td>
</tr>
</tbody>
</table>

Which features are selected after forward feature selection in best feature group?

- **MFCC group** features are the best performing in /ɑː/, /ɪː/, /ʊː/ and /eɪ/ speech sounds.
- For best performing /ɔʊ/ sound, mel-scale loudness group feature is selected at the top.
Conclusion and Future work

1. **ISCB** features showed **improvement over baseline** features in all sounds except wheeze.
Conclusion and Future work

1. **ISCB** features showed **improvement over baseline** features in all sounds except wheeze.

2. **Exhale** performed best among all stimuli for asthma and healthy classification and /ou/ performed best among all speech sounds.
Conclusion and Future work

1. **ISCB** features showed **improvement over baseline** features in all sounds except wheeze.

2. **Exhale** performed best among all stimuli for asthma and healthy classification and /ou/ performed best among all speech sounds.

3. Feature group selected across speech sounds is **stimuli dependent**, although **vowels** have **MFCC** as common group.
Conclusion and Future work

1. **ISCB** features showed **improvement over baseline** features in all sounds except wheeze.

2. **Exhale** performed best among all stimuli for asthma and healthy classification and /ou/ performed best among all speech sounds.

3. Feature group selected across speech sounds is **stimuli dependent**, although **vowels** have **MFCC** as common group.

4. **MFCC** and **loudness** is selected as best performing group for /ou/ stimuli.
Conclusion and Future work

1. **ISCB** features showed *improvement over baseline* features in all sounds except wheeze.

2. **Exhale** performed best among all stimuli for asthma and healthy classification and */ou/* performed best among all speech sounds.

3. Feature group selected across speech sounds is *stimuli dependent*, although **vowels** have **MFCC** as common group.

4. **MFCC** and **loudness** is selected as best performing group for */ou/* stimuli.

5. **MFCC group features** are best selected among all **vowels** sounds except */ou/* which has **Interquartile range between 2\(^{nd}\) and 3\(^{rd}\) quartile of mel-scale loudness** as the best feature.
Conclusion and Future work

1. ISCB features showed improvement over baseline features in all sounds except wheeze.

2. Exhale performed best among all stimuli for asthma and healthy classification and /ou/ performed best among all speech sounds.

3. Feature group selected across speech sounds is stimuli dependent, although vowels have MFCC as common group.

4. MFCC and loudness is selected as best performing group for /ou/ stimuli.

5. MFCC group features are best selected among all vowels sounds except /ou/ which has Interquartile range between 2nd and 3rd quartile of mel-scale loudness as the best feature.

6. We want to further explore the use of ISCB features for running speech.
# Table of Contents

1. Introduction
2. Motivation
3. Dataset
4. Experimental Setup
5. Results
6. Conclusion and Future work
7. References
8. Acknowledgement


# Table of Contents

1. Introduction
2. Motivation
3. Dataset
4. Experimental Setup
5. Results
6. Conclusion and Future work
7. References
8. Acknowledgement
Authors thank the Achuth Rao MV for his valuable inputs and the Department of Biotechnology, Govt. of India for their support in this work.
THANK YOU
Questions

Have Questions/suggestions? Write to us at spirelab.ee@iisc.ac.in