Classification of Healthy Subjects and Patients with Essential Vocal Tremor using Empirical Mode Decomposition of High Resolution Pitch Contour

Mekhala H S¹, Yamini B.K², Ketan J³, Pal P³, Shivashankar N²
Prasanta Kumar Ghosh⁴

¹Department of Electronics and Communications, RVCE, Bangalore-560060,
²Department of Speech Pathology and Audiology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore-560029,
³Department of Neurology, National Institute of Mental Health and Neurosciences (NIMHANS), Bangalore-560029 and
⁴Department of Electrical Engineering, Indian Institute of Science (IISc), Bangalore-560012

March 4, 2017
Section 1

1 Introduction

2 Pitch Oscillation Characteristics (POC)

3 Dataset

4 Experimentation

5 Evaluation of the classification performance

6 Results and discussion

7 Conclusion and Future Works
Motivation for the study

Figure: Illustration of phonation by a male patient with EVT and a male healthy control.
Introduction

Focus

- A new set of attributes for acoustic analysis of EVT called pitch oscillation characteristics (POC) is introduced.
- Automatic early detection of EVT symptoms in a non invasive manner from a person’s voice is attempted using the POC.
Introduction

Pitch as a characteristic of EVT

- Vocal cords vibrate at varying frequency (pitch) during speech.
- We hypothesise that the nature of variation of pitch at different time scales is a robust characteristic of EVT that could be different for EVT patients compared to that for healthy subjects and, thus, would aid in the classification of EVT patients and healthy subjects.
- Sustained phonation instead of spoken words is used to eliminate the effects of irrelevant factors such as speaking rate and temporal variation in acoustic characteristics.
Section 2

1. Introduction

2. Pitch Oscillation Characteristics (POC)

3. Dataset

4. Experimentation

5. Evaluation of the classification performance

6. Results and discussion

7. Conclusion and Future Works
Pitch Oscillation Characteristics (POC)

Figure: Block diagram summarizing the steps involved in computing the pitch oscillation characteristics (POC) features.
High resolution pitch extraction using Glottal Closure Instants (GCIs)

Figure: Comparison of pitch trajectory from block based and cycle-to-cycle pitch estimation methods.
High resolution pitch extraction using Glottal Closure Instants (GCIs)

The pitch at the $i$-th pitch cycle is computed as

$$\pi_i = \frac{F_s}{\eta_{i+1} - \eta_i}$$  \hspace{1cm} (1)

where, $F_s$ is the sampling frequency, $\eta_i$ is the GCI location [1].

---

1 A Matlab implementation of SEDREAMS [2] is used in this work.
Pitch oscillation characteristics (POC) using Empirical Mode Decomposition

- The pitch $p(n)$ and its derivative $p'(n)$ are represented as a linear combination of 6 level IMFs $c_i(n)$ and $c'_i(n)$, $i = 1, \ldots, 6$ respectively and the respective monotonic residue functions $r(n)$ and $r'(n)$ using EMD $^2[4][5]$ as follows:

\[
p(n) = \sum_{i=1}^{6} c_i(n) + r(n),
\]

\[
p'(n) = \sum_{i=1}^{6} c'_i(n) + r'(n), \quad n = 1, \cdots, L
\]

$^2$A Matlab implementation of bivariate EMD [3] is used in this work.
Pitch oscillation characteristics (POC) using Empirical Mode Decomposition

Figure: Illustration of the IMFs from the EMD of high resolution pitch
Feature Extraction

Given an analysis window of duration $L$ samples, the POC features are computed as:

$$f_1 = \frac{1}{L} \sum_{n=1}^{L} p(n),$$

$$f_2 = \log(V(p[n]))$$

$$\Delta = \log \left( \frac{1}{L} \sum_{i=1}^{i=L} \left( p(n) - \frac{1}{L} \sum_{n=1}^{L} p(n) \right)^2 \right),$$

$$f_3 = \log \left( \max_{1 \leq n \leq L} p(n) - \min_{1 \leq n \leq L} p(n) \right),$$

$$f_i = \log(V(c_{i-3}(n))), \quad i = 4, \cdots, 9 \quad (3)$$

Features computed:

- 9 POC_P : $f_1$, $f_2$, $f_3$ and $f_i$ using $p(n)$ and $c_i(n)$
- 8 POC_DP : $f_2$, $f_3$ and $f_i$ using $p'(n)$ and $c'_i(n)$
- 17 POC_PDP features combining both POC_P and POC_DP.
Section 3

1. Introduction
2. Pitch Oscillation Characteristics (POC)
3. Dataset
4. Experimentation
5. Evaluation of the classification performance
6. Results and discussion
7. Conclusion and Future Works
A group of 45 volunteers, consisting of 18 male (of age 48±10 years) and 27 female (of age 49±13 years) was used as subjects.

There were 20 subjects with EVT (11 female, 9 male) and 25 HC (16 female, 9 male).

A total of 103 recordings were obtained, with the average length of recording being 4s.
Figure: Voice recording details of 20 patients with EVT and 25 HC. For each subject, a stacked bar plot indicates the number of recordings and their respective durations.
Section 4

1. Introduction
2. Pitch Oscillation Characteristics (POC)
3. Dataset
4. Experimentation
5. Evaluation of the classification performance
6. Results and discussion
7. Conclusion and Future Works
Experimental Setup

- The analysis window duration was chosen to be 2 seconds and the shift between consecutive windows to be 400 milliseconds.
- All 103 recordings were divided into four groups each containing approximately equal number of male and female HC as well as EVT patients.

Table: The number of recordings of HC and patients with EVT in the 4 folds.

<table>
<thead>
<tr>
<th>Fold No</th>
<th>Male EVT</th>
<th>Female EVT</th>
<th>Male HC</th>
<th>Female HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>9</td>
</tr>
</tbody>
</table>
Experimental Setup

- Baseline features from MDVP (Kay Elemetrics Corp)[6]:
  1. Percent jitter
  2. Percent shimmer
  3. Frequency Tremor Intensity Index (FTRI)
  4. $F_0$-Tremor Frequency (Fftr)
  5. Amplitude Tremor Intensity Index (ATRI)
  6. Amplitude-Tremor Frequency (Fatr)
A non-linear SVM using Gaussian radial basis function (RBF) was used for classification.

RBF kernel parameter $\gamma$ and the cost of misclassification $c$ for training the SVM were separately optimized for each fold of cross-validation.

**Table: Optimized $\gamma$ and $c$ for training SVMs**

<table>
<thead>
<tr>
<th>Fold</th>
<th>$\gamma$</th>
<th></th>
<th></th>
<th></th>
<th>$c$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MDVP</td>
<td>POC_P</td>
<td>POC_DP</td>
<td>MDVP</td>
<td>POC_P</td>
<td>POC_DP</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.5</td>
<td>1</td>
<td>2</td>
<td>64</td>
<td>16</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.5</td>
<td>1</td>
<td>0.5</td>
<td>16</td>
<td>8</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.5</td>
<td>2</td>
<td>1</td>
<td>16</td>
<td>1024</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>
Section 5

1. Introduction
2. Pitch Oscillation Characteristics (POC)
3. Dataset
4. Experimentation
5. Evaluation of the classification performance
6. Results and discussion
7. Conclusion and Future Works
Aid of three speech pathologists from the National Institute of Mental Health and Neuroscience (NIMHANS), Bangalore.

Interactive Google form was used.

Their consistency (in percentage) was 70%, 85% and 100%.
F-score was calculated as the harmonic mean of precision and recall, with equal weight age provided to both as below:

$$F\text{-score} = \frac{P \times R}{P + R}$$  \hspace{1cm} (4)

where, \( P = \frac{tp}{tp + fp} \) and \( R = \frac{tp}{tp + fn} \). \( tp \) is the number of true positives of EVT, \( fp \) is the number of false positives of EVT, and \( fn \) is the number of false negatives of EVT.
Section 6

1. Introduction
2. Pitch Oscillation Characteristics (POC)
3. Dataset
4. Experimentation
5. Evaluation of the classification performance
6. Results and discussion
7. Conclusion and Future Works
Table: F-score of the classification using different features and human expert in all four folds.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Fold No</th>
<th>Avg F-score (SD)</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>MDVP</td>
<td>1.0000</td>
<td>0.8667</td>
<td>0.9333</td>
<td>1.000</td>
</tr>
<tr>
<td>POC_P</td>
<td>0.9677</td>
<td>0.9032</td>
<td>0.9286</td>
<td>0.9231</td>
</tr>
<tr>
<td>POC_DP</td>
<td>0.9091</td>
<td>0.9375</td>
<td>0.9333</td>
<td>0.8148</td>
</tr>
<tr>
<td>MDVP_POC_PDP</td>
<td>1.0000</td>
<td>0.9032</td>
<td>0.9655</td>
<td>0.9600</td>
</tr>
<tr>
<td>Human Expert</td>
<td>1.0000</td>
<td>0.9286</td>
<td>1.0000</td>
<td>0.9167</td>
</tr>
</tbody>
</table>
The F-score from human expert is:
- 1.175 (absolute) higher than that using MDVP features.
- 0.427 (absolute) higher than that using the combined features of MDVP and POC.

Thus, use of POC resulted in a 63.66% improvement of F-score with reference to the human experts.
Classification Result

Figure: Normalized histogram of each of the proposed POC and baseline MDVP features for EVT patients and HC. Title in each subplot indicate the feature for the respective.
Classification Result

Figure: Normalized histogram of each of the proposed POC and baseline MDVP features for EVT patients and HC. Title in each subplot indicate the feature for the respective.
Conclusion and Future Works

Section 7

1. Introduction
2. Pitch Oscillation Characteristics (POC)
3. Dataset
4. Experimentation
5. Evaluation of the classification performance
6. Results and discussion
7. Conclusion and Future Works
Conclusion

Through classification experiments using sustained phonation data from 25 HC and 20 patients with EVT, we show that the proposed features are complementary to the baseline features computed from MDVP.
Conclusion and Future Works

**Future Work**

- Understanding of relative contribution of each feature could help in selecting features or performing weighted combination of features for improving classification performance further.

- More varied data can be collected to build better SVM models.

- Resolution of pitch can be further improved using fractional sample interpolation of GCIs.
THANK YOU


Conclusion and Future Works
