AUTOMATIC CLASSIFICATION OF VOLUMES OF WATER USING SWALLOW SOUNDS FROM CERVICAL AUSCULTATION

Siddharth Subramani\textsuperscript{*}, Achuth Rao M V\textsuperscript{*}, Divya Giridhar\textsuperscript{*}, Prasanna Suresh Hegde\textsuperscript{†}, Prasanta Kumar Ghosh\textsuperscript{*}

\textsuperscript{*} SPIRE LAB, Electrical Engineering, Indian Institute of Science (IISc), Bangalore, India
\textsuperscript{†} Dept. of Head and Neck Surgery, Health Care Global Enterprises Ltd., Bangalore, India

09 April 2020
Flow of presentation

- Introduction
- Data
- Proposed Study
- Results
- Conclusion
Plan

1 Introduction

2 Data

3 Proposed Study

4 Results

5 Conclusion
Swallow physiology and Dysphagia

- Swallow - Movement of food bolus from mouth through pharynx to the esophagus
- Process in sync with respiration
- Swallowing disorders (Dysphagia) can occur due to
  - Neurological disorders - Parkinson’s disease, stroke
  - Irregularities in esophageal and pharyngeal muscles
  - Amyotrophic lateral sclerosis
  - Head and Neck cancer, etc.

- Common clinical assessments - VFSS, FEES - invasive and expensive methods
- Cervical Auscultation (CA) characterizes swallow in terms of signal - non-invasive
Applications of Swallow Sound analysis

- Wearable MIB (Monitoring of Ingestive Behaviour) devices for food intake characterization
- Identification of dysphagia through Spontaneous Swallowing frequency analysis
- Volume specific models can help in studying the severity of dysphagia
Swallow signal characterization

- Swallow signals are characterized by their Swallow Components[1]

Red, black and magenta lines represent SC1, SC2 and SC3 regions, respectively

- Why volume-specific features?: Signatures of SCs found to vary with bolus volume

- Objective: Learn features for bolus volume characterization through classification of swallows of water against dry swallows

Plan

1. Introduction
2. Data
3. Proposed Study
4. Results
5. Conclusion
Swallow Sound Recording

- 56 subjects: 34 male, 23 female (20 - 30 years of age)
- Subjects made to swallow water of volumes - 2ml, 5ml and 10ml, and also perform dry swallow
- CA setup
  - Life-Line Paediatric-Al Stethoscope
  - Acoustic tube
  - Sorella’z portable 3.5mm microphone (frequency range of 30Hz - 15000Hz)
- Device placement site: Posterior inferior to the cricoid cartilage encircling the trachea

Plan

1. Introduction
2. Data
3. Proposed Study
4. Results
5. Conclusion
Proposed Study

Feature sets

- **Baseline Feature set (BLF):** 12 features - duration of SSW1, SSW2, SC1, SC2, SC3, intervals between swallow components (I1, I2), duration to peak intensity (DPI), peak intensities (PI(SSW1, PI(SSW2, PI) and total duration (TD) of the swallow signal.

- **ComPARE 2016 Feature set (OSF):** 6373 features - Voice source based (group A) and glottal excitation based (group B) features

<table>
<thead>
<tr>
<th>Classes</th>
<th>Feature name / Abbreviation / Dimension</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>Loudness (Ldns)(1), RASTA (26), MFCC(14), RMS energy (RMSe)(1), Modulated Loudness (MLdns)(1), ZCR(1),</td>
<td>59 X 54 functionals +</td>
</tr>
<tr>
<td></td>
<td>Band energy (BE)(2), Spectral: ROP (SR)(4), Flux (SF)(1), Centroid (SC)(1), Slope (SSI)(1), Entropy</td>
<td>59 delta X 46 functionals =</td>
</tr>
<tr>
<td></td>
<td>(SE)(1), Variance (SV)(3), Harmonicity (SH)(1), Sharpness (Shs)(1)</td>
<td>5900 features</td>
</tr>
<tr>
<td>Group B</td>
<td>F0(1), Prob. Voicing (PV)(1), Jitter (J)(2), Shimmer (Shr)(1), log(HNR)(1)</td>
<td>6 X 39 functionals +</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 delta X 39 functionals =</td>
</tr>
<tr>
<td></td>
<td></td>
<td>468 features</td>
</tr>
</tbody>
</table>
Proposed Study

Experimental setup

- Baseline features computed from swallow signals annotated by a clinical expert
- ComPARE 2016 features computed using OpenSMILE Audio Extraction and Analysis Tool - require no manual annotation
- 10-fold cross-validation step
- Three classification tasks: Dry-vs-2ml, Dry-vs-5ml, Dry-vs-10ml
- Classifier: Linear SVM
- Grid search for optimal C-parameter selection
- Evaluation metric: F-score
- Significance test: Wilcoxon signed rank test for zero median
Study Design 1 - BLF vs OSF

- Assessment of overall performance of BLF and OSF
  Features requiring manual annotation vs features not requiring manual annotations

![Diagram showing Design 1: BLF and OSF connected to SVM, with Volume class output]
Proposed Study

Study Design 2

- Assessment of performance of subgroups of OSF vocal tract related features vs glottal excitation related features

Design 2
OSF - Grp. A
OSF - Grp. B
SVM
Volume class
Study Design 3

- Feature ranking for each fold and each task can be different
- Joint features: FFS is computationally expensive!
  - LASSO feature selection algorithm used to rank order OSF features
  - Features selected in two ways
    - OS-ranked: 12 features for each of the three classification tasks
    - OS-common: 12 features common across all three classification tasks
Plan

1. Introduction
2. Data
3. Proposed Study
4. Results
5. Conclusion
Results of Study 1: **BLF vs OSF**

Table: Mean **F-scores (%)** of the three classification tasks

<table>
<thead>
<tr>
<th>Classification task</th>
<th>BLF</th>
<th>OSF</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry vs 2ml</td>
<td>37.26 (15.05)</td>
<td><strong>70.14</strong> (8.56)</td>
<td>0.03</td>
</tr>
<tr>
<td>Dry vs 5ml</td>
<td>57.77 (15.05)</td>
<td><strong>73.44</strong> (3.03)</td>
<td>0.0098</td>
</tr>
<tr>
<td>Dry vs 10ml</td>
<td>69.10 (9.67)</td>
<td><strong>77.45</strong> (6.65)</td>
<td>0.002</td>
</tr>
</tbody>
</table>

- OSF outperformed BLF by an average of around **18.9%**
- Reduced standard deviation in OSF indicates the robustness of OSF, across subjects and volumes
Results of Study 2: OSF - **Group A vs Group B**

- Features unrelated to voice source outperformed features related to voice source by **21.02%** and also OSF by **5.7%**
- Top performing features - MFCC, RASTA filtered audio spectrum and RMS energy

**Figure:** Mean F-scores(%) of different Group A features (ranked by F-scores of dry-vs-2ml classification); violet and blue horizontal lines are the baseline F-scores of dry-vs-5ml and dry-vs-10ml respectively
Results of Study 3

Mean F-scores (%) of OS-common and OS-ranked features with standard deviations in (.)

<table>
<thead>
<tr>
<th>Volume / Feature set</th>
<th>OS-common</th>
<th>OS-ranked</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry vs 2ml</td>
<td>73.55 (6.78)</td>
<td>74.84 (4.85)</td>
<td>0.6523</td>
</tr>
<tr>
<td>Dry vs 5ml</td>
<td>75.88 (9.05)</td>
<td>77.67 (8.17)</td>
<td>0.4316</td>
</tr>
<tr>
<td>Dry vs 10ml</td>
<td>80.68 (4.88)</td>
<td>80.36 (7.61)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

- *p*-values show that the performance of OS-common & OS-ranked features are not significantly different.

Top 12 OS-common features with Pearson correlation coefficient (PCC) between feature value and the volume of water in [.] $P_k$ indicates $k$ percentile

<table>
<thead>
<tr>
<th>Ldns</th>
<th>Derivative of Ldns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inter-quartile range [0.14] ($P_{75} - P_{50}$)</td>
<td>$P_1$ [0.39], Standard deviation[0.41], Kurtosis [0.03], Mean segment length [0.25], Minimum segment length [0.18], Up-level time 25,50,75,90 [0.03] (%), Risetime [0.023], Left curvature time [0.06]</td>
</tr>
</tbody>
</table>
Plan

1 Introduction
2 Data
3 Proposed Study
4 Results
5 Conclusion
Key Takeaways

- Eliminated need for expert manual annotation
  - Improved F-score of OSF over BLF
  - LASSO selected functionals provided cues & indirectly indicated the boundaries of SCs
- 12 functionals of Ldns outperformed (significantly different from) the 12 baseline features
- Lower standard deviations - proposed features more robust to variations in signal characteristics with volume
References


- Florian Eyben, Martin Wöllmer, and Björn Schuller, ”openSMILE - the Munich Versatile and Fast Open-Source Audio Feature Extractor,” 01 2010, pp. 1459-1462.
THANK YOU
Have questions/suggestions?

Write to us at spirelab.ee.@iisc.ac.in