**SEGNET-BASED DEEP REPRESENTATION LEARNING FOR DYSPHAGIA CLASSIFICATION**

Siddharth Subramani, Achuth Rao MV, Anwesha Roy, Prasanna Suresh Hegde, Prasanta Kumar Ghosh

Department of Electrical Engineering, Indian Institute of Science, Bangalore-560 012, India, *Dept. of Head and Neck Surgery, Health Care Global Enterprises Ltd Bangalore-560 002, India

---

**Introduction**

- Dysphagia - difficulty in swallowing food
- Dysphagia can potentially cause aspiration, dehydration & malnutrition
- Swallowing can be studied non-invasively through Cervical Auscultation (CA)
  - CA captures vibrations from the surface of the throat when swallowing food bolus
  - ... and gives reliable representations just as golden standards such as VideoFluoroScope
- Objective:
  - Learn acoustic representations to classify swallows as healthy or dysphagic

**Dataset**

- 24 subjects: 14 healthy controls, 10 patients

- CA device setup: Paediatric-AI Stethoscope connected to microphone (30Hz - 15000Hz) was patched on throat surface posterior to cricoid cartilage, around the trachea

- Swallow tasks: Dry, 5ml, 10ml and 15ml of water, 3-4 times per subject

- 172 healthy & 118 dysphagic swallow recordings in total - Dataset InD

**Swallow features and process outline**

- Mel-spectrogram
  - Healthy or Dysphagic
- Binary mask
- MSpecs classifier

- Input feature: Mel-Spectrogram (MSpec) for every swallow signal was computed with 21ms long hamming window and 2ms long hop
- Pre-processing - MSpecs (not signals) padded with zeros to equalize length of all MSpecs

---

**Classification pipeline**

- 1D-CNN (13 layers) SegNet-based feature learner
- Linear-SVM classifier
- MSpecs multiplied with binary 1D-mask prior to convolution layers in SegNet

**Proposed approaches & Experimental Setup**

- Three investigations

**Approach 1**

- (CNN1)

**Approach 2**

- (Jnt)

**Approach 3**

- (TsT-F) - Training
- (TsT-G) - Testing

- BF: Bottleneck Features
- MSpec reconstruction
- H/D: Healthy or Dysphagic

**Set up:**

- 5 groups of 24 subjects with two patients and three controls in every group - 3 groups as training set, 1 group each for validation and test sets (with no common subjects); 5-fold CV
- SegNet reconstruction loss: MSE, validation accuracy based selection of SVM regularization parameter
- Baseline scheme (BLS) comparison, fourier transform and spectrogram features to train an RBF-SVM classifier

**Results & Discussion**

- Depth analysis: 5 “RCF-trials” with bottleneck features from each of the 5 progressing maxpooling layers in SegNet encoder
  - Learning at different depths identifies time-scale resolution that maximizes performance
  - Each maxpooling layer has Receptive Field Size (RCF) of 3, 10, 20, 50, 120 respectively

- Best: TsT-F in RCF4-trial with a mean across folds test F1-score of 80.13%

**Baseline F1-score = 71.95%, Sensitivity = 76.67%, Specificity = 42.92%**

- BLS suffered on InD (larger and imbalanced data)
- Jnt performance only competitive with BLS but still showed low F1-scores
- TsT-F outperformed all approaches, including TsT-G, with highest mean F1-score and lowest mean SD across all RCF-trials

**Conclusion**

- Unsupervised feature learning followed by supervised training performed better than fully supervised training
- Performance variation at different depths hints the influence of swallow phase transition patterns

**Future directions**

- Adapt proposed method to identify severity of dysphagia
- Explore time-scale analysis to outline spectral signatures pertaining to levels of severity

**References**


---

**Acknowledgement:** The authors thank the Pratiksha Trust and the Department of Science and Technology, Govt. of India for their support.

---

[http://spire.ee.iisc.ac.in/spire/](http://spire.ee.iisc.ac.in/spire/)

[ssd9709@gmail.com, achuthr@iisc.ac.in, anwesha97@yahoo.in, prasanna.h@hcgei.com, prasantig@iisc.ac.in](mailto:ssd9709@gmail.com, achuthr@iisc.ac.in, anwesha97@yahoo.in, prasanna.h@hcgei.com, prasantig@iisc.ac.in)