Analysis of Sustained Phonation of Person with Voice Disorder

Hitesh A H

TRV20ECSP14, Signal Processing
Department of Electronics and Communication Engineering

March 3, 2022

Overview

Research Problem

Literature Review

Methodology

Works & Planning Ahead

Introduction

- The **human voice** is a wonderfully, perhaps uniquely, expressive instrument, exhibiting a bewildering number of expressive variations beyond those of pitch and loudness, including trill, effort level etc.
- The speech represents an intrinsic characteristic of human behaviour. Any disturbances in the normal speech of a human being are called **speech disorder**.
- Automatic detection and assessment of voice disorders is important in diagnosis and treatment planning of voice disorders.

Introduction

1. Speech production requires airflow from the lungs to be phonated through vocal folds of the larynx and resonated in the vocal cavities shaped by the tongue, jaw, soft palate, lips, and other articulators.
2. Phonation is a process by which the vocal folds produce certain sounds through quasi-periodic vibration(voicing).
3. Any abnormality in the larynx that affect voicing in speech production: **Voice disorder**.

Significance

- According to American speech-language-hearing association, voice disorders: **organic** & **neurological**.
- Affects communication and social integration: patients will also have psychological and emotional issues. Deterioration of the quality of life.
- Expensive tests like video or laryngoscopy/ stroboscopy.
- Hence, early detection of speech disorder is of utmost importance.

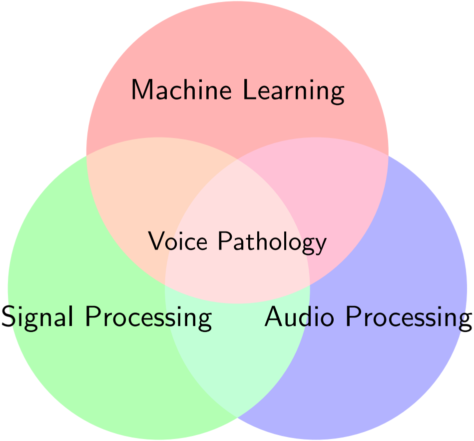
From Literature

Article: *Prevalence of Voice Disorders in School Teachers in a District in South India*

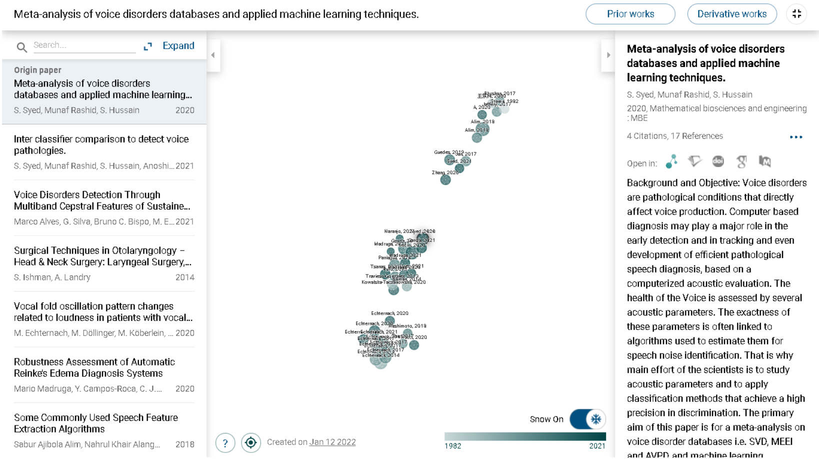
Results

Teachers reporting a current voice complaint at the time of study were administered the validated vernacular version of the voice handicap index questionnaire (VHI 30). 702 teachers, The reported prevalence was 45.4% for present difficulty related to their voice, 52.8% for some voice problem in the last 1 year[2019], and 70.1% for problems experienced during the duration of their teaching career.

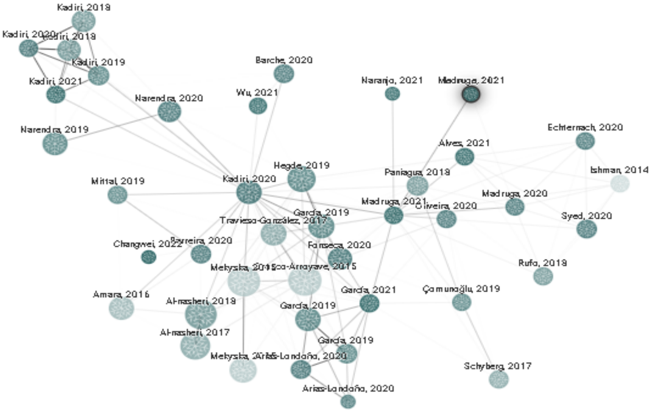
Field Overview



Citation Graph



Research so far..



Source: *Connected Papers*

Comparison Summary

Article: *] Pavol Harar, Jesus B. Alonso-Hernandez, Jiri Mekyska, Zoltan Galaz, Radim Burget and Zdenek Smekal: Voice Pathology Detection Using Deep Learning: a Preliminary Study*
Ext. features: RMS energy & PCM loudness
Classifier: SVM
Accuracy: 71.3%

Comparison Summary

Article: *] Alireza Bayestehtashk, Meysam Asgaria, Izhak Shafran, James McName: Fully automated assessment of the severity of Parkinson's disease from speech*
Ext. features: pitch frequency, jitter, shimmer & MFCC
Classifier: Open SMILE
Accuracy: 91.8%

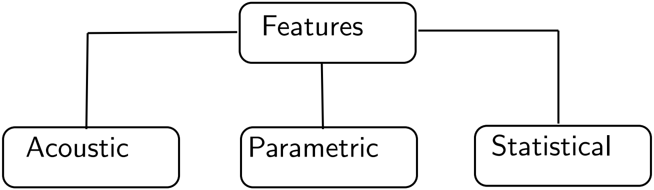
Comparison Summary

Article: *J. R. Orozco-Arroyave, Florian Hönig, J. D. Arias-Londoño, J. F. Vargas-Bonilla1 and Elmar Nöth : Spectral and cepstral analyses for Parkinson'sdisease detection in Spanish vowels and words. Expert Systems*
Ext. features: LPC, MFCC
Classifier: SVM with Gaussian Kernal
Accuracy: 80-92%

Key Points

- In literature, widely used system features are: Mel Frequency Cepstral Co-efficient (MFCC), perceptual linear prediction (PLP), and Linear Prediction Cepstral Coefficient (LPCC) reliable acoustic measures of voice impairment.
- Adversarial Instance have Robust Model with minimum error.
- Closed domain adaptation with fine tuning.

Feature extraction approaches in voice disorder detection



Performance of voice disorder detection and assessment systems in terms of classification accuracy

Feature type	Exp. 1	Exp. 2	Exp. 3	Exp. 4
ComParE	**82.8**	71.7	**74.3**	**65.3**
eGeMAPS	76.0	70.1	67.3	57.5
MFCC	74.4	72.4	67.8	63.4
PLP	74.2	**72.7**	70.5	64.1
Glottal	67.4	64.8	59.9	58.3
Intonation	69.3	66.0	60.2	52.8
MFCC-Residual	67.4	70.8	64.3	61.0
MFCC-ZFF	68.5	69.2	66.4	64.2

Source: *Towards automatic assessment of voice disorders: A clinical approach (INTERSPEECH 2020)*

Performance of voice disorder detection and assessment system

Feature type	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Glottal + ComParE	**85.2**	72.7	73.1	59.2
Glottal + eGeMAPS	79.0	70.8	65.5	60.1
Glottal + MFCC	74.4	71.2	66.7	64.1
Glottal + PLP	78.0	71.5	67.8	63.0
Intonation + ComParE	84.9	72.8	74.9	60.3
Intonation + eGeMAPS	81.5	68.5	68.1	60.1
Intonation + MFCC	77.5	**75.0**	65.2	64.4
Intonation + PLP	77.6	72.7	69.3	62.4
MFCC-Residual + ComParE	84.1	73.0	**76.0**	65.0
MFCC-Residual + eGeMAPS	84.3	70.9	62.6	63.3
MFCC-Residual + MFCC	73.1	74.6	69.6	66.2
MFCC-residual + PLP	74.2	73.0	68.4	65.3
MFCC-ZFF + ComParE	84.5	72.3	74.0	67.3
MFCC-ZFF + eGeMAPS	84.3	71.8	67.5	62.1
MFCC-ZFF + MFCC	71.7	72.3	68.7	63.6
MFCC-ZFF + PLP	74.4	70.1	70.5	65.9
Glottal + Intonation + MFCC-Residual + MFCC-ZFF	75.6	72.4	67.0	**70.0**

Source: *Towards automatic assessment of voice disorders: A clinical approach* (INTERSPEECH 2020)

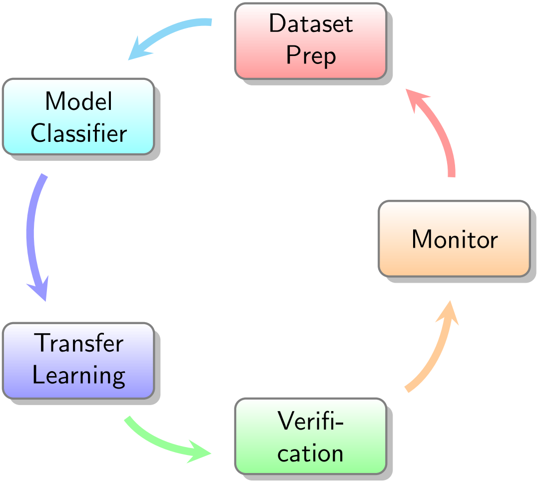
Inferences

- The classification of functional and **psychogenic** voice disorders is more challenging compared to classification of structural and neurogenic voice disorders.
- Detection of voice disorder has a higher classification accuracy compared with the assessment of voice disorders

Machine Learning Life Cycle

- **Define Objectives** — Voice Pathology, Sample Size, and Audio Features
- **Data** — Preprocesing, labelling, sorting or frame conversion
- **Training** — Feauture Extraction
- **Evaluation** — Accuracy

ML Concept


Dataset Prep
Model Classifier
Transfer Learning
Verifi- cation
Monitor

Database

- Saarbruecken Voice Disorder dataset which is freely available on `http://www.stimmdatenbank.coli.uni-saarland.de/`
- 2000 voice recordings sampled at 50 kHz: [Healthy 428F+259M, Subject 727F+629M]
- 71 different voice disorders

Disorder Type		Disorder name	#Samples
Organic	Structural	Laryngitis	30
		Leukoplakia	41
		Polyp	45
	Neurogenic	SD	30
		RLNP	188
Non-organic	Functional	FD	254
	Psychogenic	PD	91

Feature: Mel frequency cepstral coefficients of LP-residual, and ZFF signal

- 39 dimensional cepstral coefficients consisting in 13 static coefficients, and their first and second order derivatives
- Modelling about 120+ dimensional vector. (156)

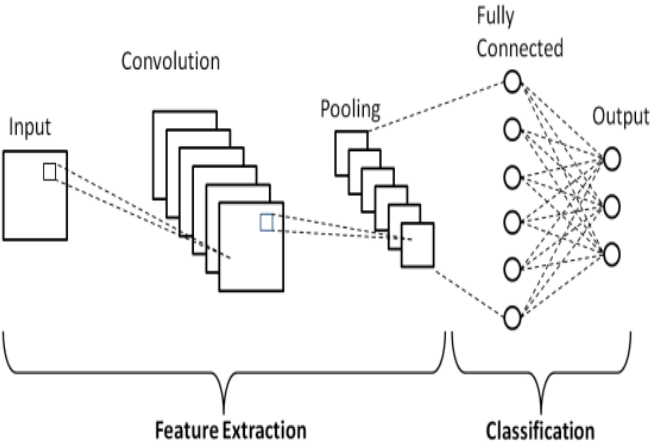
Classifier: Support vector machine (SVM) classifier is the most widely used classifier in voice disorder detection as it gives consistence performance even on small dataset.

Project Model: CNN-SVM

1. To develop a hybrid model of a powerful Convolutional Neural Networks (CNN) and Support Vector Machine (SVM) for detection.
2. CNN works as an automatic feature extractor and SVM works as a binary classifier.

SVM replaces the softmax layer of CNN

CNN Model


Fully
Connected
Convolution
Pooling
Output
Input
Feature Extraction
Classification

Works Done so far

- **Literature Survey:**
  1. Project Area's SWOT Analysis
  2. Scope for improvement [Literature or Algorithm]

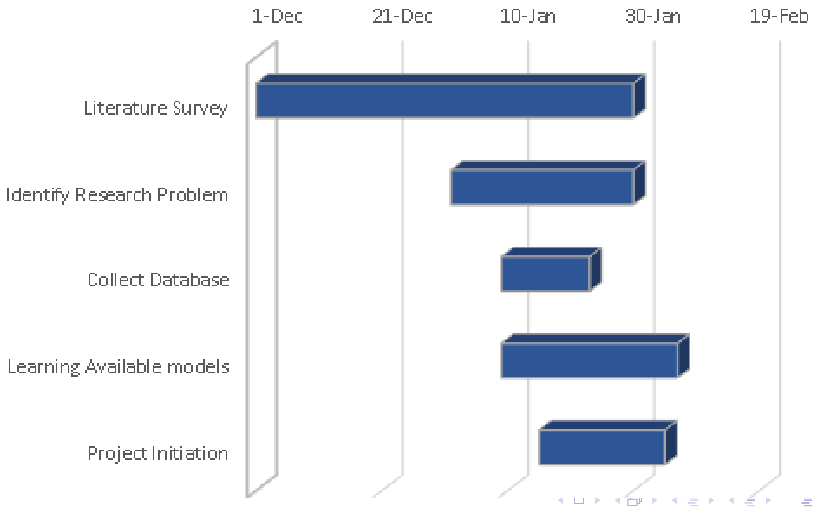
Works Done so far

- **Gathering Resources:**
  1. Revise: Statistics, Calculus, Linear Algebra & Probability
  2. Programming: Preprocessing data, feature extraction
  3. Model: Exploring other databases & Classifiers

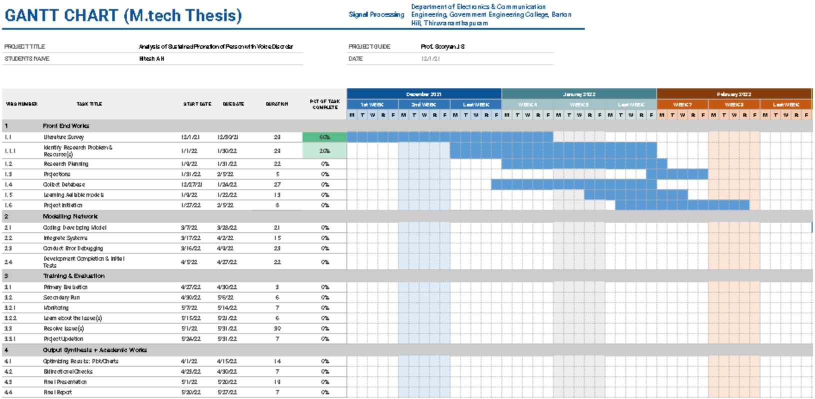
Works Done so far

- **Preparing Project Execution Steps:**
  1. Contacting resource persons.
  2. Delegating task on a Gantt Chart.

Current Status on Work



Gantt Chart



GANTT CHART (M.tech Thesis)

Signal Processing — Department of Electronics & Communication Engineering, Government Engineering College, Barton Hill, Thiruvananthapuram

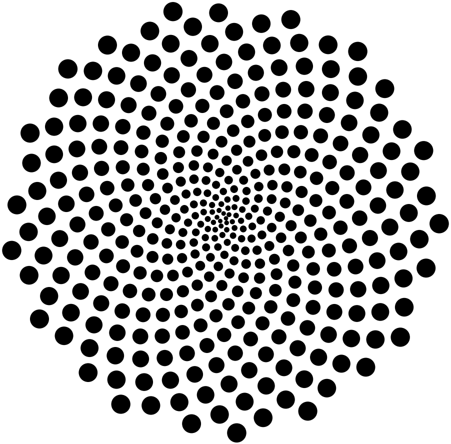
PROJECT TITLE	Analysis of Sustained Phonation of Person with Voice Disorder	PROJECT GUIDE	Prof. Sooryan J S
STUDENTS NAME	Hitesh A H	DATE	12/1/21

WBS NUMBER	TASK TITLE	START DATE	DUE DATE	DURATION	PCT OF TASK COMPLETE
1	Front End Works				
1.1	Literature Survey	12/1/21	12/30/21	29	60%
1.1.1	Identify Research Problem & Resource(s)	1/1/22	1/30/22	29	20%
1.2	Research Planning	1/9/22	1/31/22	22	0%
1.3	Projections	1/31/22	2/5/22	5	0%
1.4	Collect Database	12/27/21	1/24/22	27	0%
1.5	Learning Available models	1/9/22	1/22/22	13	0%
1.6	Project Initiation	1/27/22	2/5/22	8	0%
2	Modelling Network				
2.1	Coding: Developing Model	3/7/22	3/28/22	21	0%
2.2	Integrate Systems	3/17/22	4/2/22	15	0%
2.3	Conduct Error Debugging	3/16/22	4/9/22	23	0%
2.4	Development Completion & Initial Tests	4/5/22	4/27/22	22	0%
3	Training & Evaluation				
3.1	Primary Evaluation	4/27/22	4/30/22	3	0%
3.2	Secondary Run	4/30/22	5/6/22	6	0%
3.2.1	Monitoring	5/7/22	5/14/22	7	0%
3.2.2	Learn about the Issue(s)	5/15/22	5/21/22	6	0%
3.3	Resolve Issue(s)	5/1/22	5/31/22	30	0%
3.3.1	Project Updation	5/24/22	5/31/22	7	0%
4	Output Synthesis + Academic Works				
4.1	Optimizing Results: Plot/Charts	4/1/22	4/15/22	14	0%
4.2	Bidirectional Checks	4/23/22	4/30/22	7	0%
4.3	Final Presentation	5/1/22	5/20/22	19	0%
4.4	Final Report	5/20/22	5/27/22	7	0%

Click to View full Chart

Retrospective

- Existing approaches are reactive post fixes to an outdated method.(Monolithic Models)
- Modular Components have clear specification interface for individual feature.
- Building a reusable structure of learning is not a solution.
- Manual efforts are required in certain task(s). (Models don't communicate in between)



Thank You