STATISTICAL ANALYSIS OF SPEECH DISORDER SPECIFIC FEATURES TO CHARACTERISE DYSARTHRIA SEVERITY LEVEL

Amlu Anna Joshy[1,3],
P. N. Parameswaran[1,3], Siddharth R. Nair[1,3],
Dr. Rajeev Rajan[2,3]

[1] **College of Engineering Trivandrum, Thiruvananthapuram,** [2] **Government Engineering College, Barton Hill,**
[3] **APJ Abdul Kalam Technological University,**
Kerala, India.

Contents

Introduction

Dysarthria :

- Neuro-motor speech disorder
- Occurs due to any neurological injury/neuro-degenerative disease
- Any of the speech production subsystems (respiration, phonation, resonance, prosody, and articulation) can be affected.
- Characterised by utterances having prolonged pause intervals, slow speaking rates, poor articulation of phonemes, syllable deletions, etc.
- Leads to poor intelligibility/ low audibility/ unnaturalness/ hyper-nasality/ weak facial reflexes/ harsh voice quality/ increased fatigue on speaking. [1].

Assessment of Dysarthria Severity

- Identify severity level for proper medication and speech therapy during rehabilitation
- Speech examinations by a speech language pathologist (SLP) : biased, time-consuming, and expensive.
- Automation => "Mimic the human perception system"
- A classifier that establishes a mapping between the speech features and the severity labels (very low/low/medium/high) as determined during perceptual evaluation of speech intelligibility by an SLP.

Objectives and Motivation

To comparatively study the prosodic, glottal, phonetic and articulatory features for ranking their efficacy in recognizing the dysarthria severity level.

- Our initial experiments using these speech disorder specific features on deep learning classifiers [2] suggested that a detailed statistical analysis is required to understand the potential correlation within each class.
- Enables a choice of the optimum feature descriptor that could be used by a simple predictor for aiding SLPs.

Objectives and Motivation

To use the paraconsistent feature engineering (PFE) technique [3] for the analysis

- Can picture the intra-class similarities and the inter-class distinctions exhibited by the features.
- Has been shown to be efficient in feature ranking for applications such as replay attack detection [4] and speaker verification [5].

Paraconsistent Feature Engineering (PFE) [3]

- Feature ranking method
- All the available X number of feature vectors are L2-normalized.
- Intra-class similarities analysed using α, the level of faith.
- Define $A = max - min$ within each class for each feature.
- Calculate $Y = 1 - A$, for feature vectors of dimension D and an N-class problem.

$$\bar{Y}_N = \frac{1}{D}\sum_{i=1}^{D} y_N(i) \tag{1}$$

$$\alpha = min\{\bar{Y}_1, \bar{Y}_2, ..., \bar{Y}_N\} \tag{2}$$

PFE Approach [3]

- Inter-class dissimilarities analysed using β, the level of discredit.
- Inter-class distinction analysed using range vectors, the degree of overlaps

$$\beta = \frac{R}{F} \tag{3}$$

where, R = count of features in one class lying within the range vector of all the other classes and F = maximum possible number of overlaps = $N.(N-1).X.D$.

PFE Approach [3]

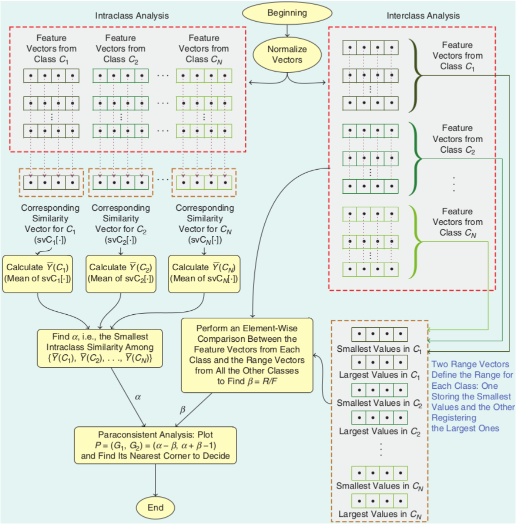

FIGURE – Flowchart of the PFE approach reproduced from [3]

PFE Approach [3]

- Degree of certainty $G1$ and degree of contradiction $G2$ define the paraconsistent plane
- Ideal case : linearly separable features : gives (1, 0).
- Distance D from the point $P = (G1, G2)$ to the ideal point (1, 0) is calculated.

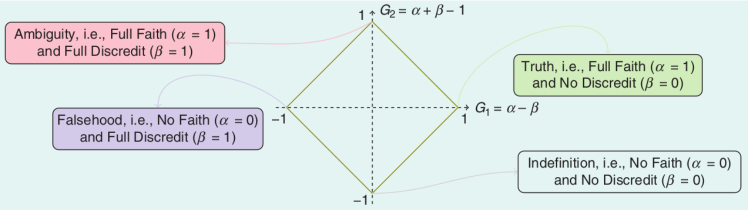

FIGURE – The paraconsistent plane reproduced from [3]

Feature Description

Prosody Features

- Abnormal changes in pitch, loudness and time duration
- Prevent conveying the right emotion and rhythm to the speech
- Estimated in terms of mean pitch, jitter, shimmer, the proportion of the vocalic duration and the degree of voiced breaks
- 103 prosody features based on duration, fundamental frequency and energy computed [6]

Glottal Features

- Irregular glottal closure pattern and related breathy voice
- Glottal flow patterns characterised by nine different time-frequency parameters as in [7].
- Statistical measures applied => 36 features per utterance

Feature Description

Articulatory Features

- Retardation in the lip, tongue, and jaw movements => imprecise articulations
- 122 descriptors including the bark band energies, formants and mel-frequency cepstral coefficients(MFCCs) during the onset and offset transitions

Phonation Features

- Deteriorated voice quality in terms of stability and periodicity
- 7 phonation measures corresponding to the jitter and shimmer, amplitude and pitch perturbation quotients, glottal-to-noise excitation ratio, and harmonics-to-noise ratio, cepstral harmonics to noise ratio, and the normalized noise energy [8]

Experimental Setup

- The speech disorder specific features are extracted using the DisVoice python library [1] and the Kaldi toolkit.
- Static/utterance level features are computed
- Four statistical functions applied : mean, standard deviation, skewness, and kurtosis
- Concatenated vector : 655 dimension

1. https://github.com/jcvasquezc/DisVoice

Databases Used

- **UA-Speech [9]**
  - Data of 15 cerebral palsy(CP) patients available.
  - Utterances correspond to three repetitions of the 10 digits, 19 computer commands, 26 international radio alphabets, 100 common words and 300 distinct uncommon words.
  - 465 common words and 300 uncommon words per speaker.
- **TORGO [10]**
  - Data from 8 dysarthric speakers with CP or amyotrophic lateral sclerosis (ALS) used.
  - Dysarthric word utterances only used => 2227 available in total
  - 80% for training and 20% for testing.

TABLE – Class-wise patient description

Severity	UA-Speech	TORGO
VERY LOW	F05, M08, M09, M10, M14	F03, F04, M03
LOW	F04, M05, M11	F01,M05
MEDIUM	F02, M07, M16	M01, M02, M04
HIGH	F03, M01, M04, M12	-

Results using PFE Ranking

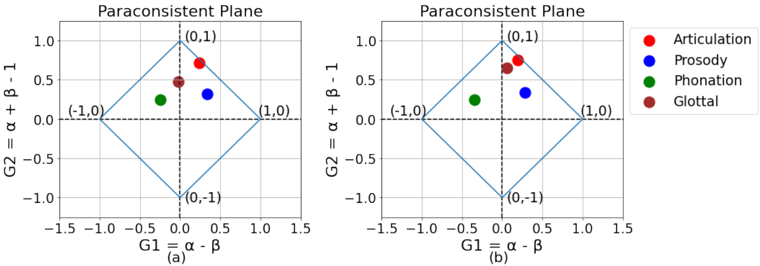

FIGURE – Plots of different feature points in the paraconsistent plane (a) UA-Speech (b) TORGO

Results using PFE Ranking

TABLE – Paraconsistent framework on features (best values in bold)

-	UA-Speech			TORGO		
Feature set	α	β	D	α	β	D
Prosody	0.83	**0.49**	**0.73**	0.81	**0.53**	**0.79**
Articulation	**0.97**	0.74	1.05	**0.97**	0.78	1.11
Glottal	0.73	0.75	1.12	0.86	0.79	1.14
Phonation	0.49	0.75	1.27	0.45	0.79	1.36

- Prosody : P lies closest to the ideal point (1,0), hence best β and D values => inter-class dissimilarity is the greatest
- Articulation : highest α value => intra-class similarity is the greatest

Results using Different ML Classifiers

TABLE – Classification accuracy (%) obtained on different ML classifiers (best values in bold)

Database	Classifier	Phonation	Glottal	Prosody	Articulation
TORGO	SVM	62.88	55.60	60.18	**83.18**
	RF	69.14	76.45	81.49	**85.65**
	kNN	60.09	50.44	69.23	**73.99**
	NB	**54.29**	44.84	39.90	45.74
UA-Speech	SVM	60.81	55.91	61.68	**77.98**
	RF	65.82	70.86	67.72	**77.64**
	kNN	53.38	43.33	54.90	**60.69**
	NB	46.12	43.40	46.89	**54.02**

- Articulatory features gave the best results among subsets on all the classifiers, except NB.
- Same trend observed on deep learning classifiers in our previous experiments [2].
- Articulation deficits and reduced vowel articulation index efficiently mapped the stage of Parkinsons disease in [11]

Results of Feature Importance Calculation

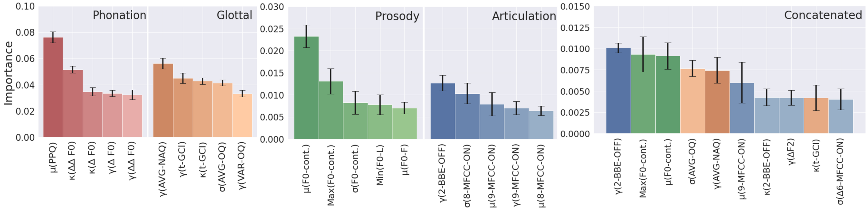

FIGURE – Feature importance graph using permutation on the UA-Speech database(X-axis shows important features from each set)

- Most discriminating feature = skewness of 2^{nd} Bark Band Energy(BBE) on offset transitions.
- BBEs - reduced in dysarthics compared to healthy speakers [6].
- Top four in articulatory feature set = MFCCs, affirming results of [12], [2].

Conclusion

Findings

- Prosody and articulation features are found to be best useful, which was supported by the classification accuracies obtained on using different ML classifiers.
- As previously reported in [2] and supported by the findings in [13], the classification accuracy does not improve with the mere increment in feature dimension.

Relevance of the Study

- Low resource of impaired speech data
- Extendable to other speech disorders like apraxia, and to specific cases of dysarthria like hypokinetic dysarthria exhibited by Parkinson's disease.
- Helps in implementing simple predictors based on ranking results without the problem of over-fitting, to aid SLPs.

References I

R. Palmer and P. Enderby.
Methods of speech therapy treatment for stable dysarthria : A review.
Int. J. Speech, Lang. Pathology, pages 140–153, 2007.

Amlu Anna Joshy and Rajeev Rajan.
Automated dysarthria severity classification : A study on acoustic features and deep learning techniques.
IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30 :1147–1157, 2022.

Rodrigo Capobianco Guido.
Paraconsistent feature engineering [lecture notes].
IEEE Signal Processing Magazine, 36(1) :154–158, 2018.

Ankur T Patil, Rajul Acharya, Hemant A Patil, and Rodrigo Capobianco Guido.
Improving the potential of enhanced teager energy cepstral coefficients (etecc) for replay attack detection.
Computer Speech & Language, 72 :101281, 2022.

References II

Alex Marino GonÃ§alves de Almeida, Claudineia Helena Recco, and Rodrigo Capobianco Guido.
Use of paraconsistent feature engineering to support the long term feature choice for speaker verification.
The International FLAIRS Conference Proceedings, 34, Apr. 2021.

Juan Rafael Orozco-Arroyave, Juan Camilo Vásquez-Correa, Jesús Francisco Vargas-Bonilla, Raman Arora, Najim Dehak, Phani S Nidadavolu, Heidi Christensen, Frank Rudzicz, Maria Yancheva, H Chinaei, et al.
Neurospeech : An open-source software for parkinson's speech analysis.
Digital Signal Process., 77 :207–221, 2018.

Elkyn Alexander Belalcázar-Bolanos, Juan Rafael Orozco-Arroyave, Jesús Francisco Vargas-Bonilla, Tino Haderlein, and Elmar Nöth.
Glottal flow patterns analyses for parkinson's disease detection : acoustic and nonlinear approaches.
In *Proc. Int. Conf. Text, Speech, and Dialogue*, pages 400–407. Springer, 2016.

कर्म ज्यायोह्य कर्मण

References III

Tomas Arias-Vergara, Juan Camilo Vásquez-Correa, and Juan Rafael Orozco-Arroyave.
Parkinson's disease and aging : analysis of their effect in phonation and articulation of speech.
Cognitive Computation, 9(6) :731–748, 2017.

Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas S Huang, Kenneth Watkin, and Simone Frame.
Dysarthric speech database for universal access research.
In *Ninth Annual Conference of the International Speech Communication Association*, 2008.

Frank Rudzicz, Aravind Kumar Namasivayam, and Talya Wolff.
The torgo database of acoustic and articulatory speech from speakers with dysarthria.
Language Resources and Evaluation, 46(4) :523–541, 2012.

Sabine Skodda, Wenke Visser, and Uwe Schlegel.
Vowel articulation in parkinson's disease.
J. Voice, 25(4) :467–472, 2011.

References IV

Amlu Anna Joshy and Rajeev Rajan.
Automated dysarthria severity classification using deep learning frameworks.
In *Proc. 28th Eur. Signal Process. Conf.*, pages 116–120, 2021.

Henrik Holmström and Victor Zars.
Effect of feature extraction when classifying emotions in speech-an applied study.
Ph.D dessertation, Umeà University, Sweden, 2018.

THANK YOU