Robust & complex approach of pathological speech signal analysis

Publication in ScienceDirect Neuro Computing Elsevier journal

What the paper is about

- Study of approaches in the state of art in the field of **pathological speech signal analysis, special focus on parameterization techniques**
- Description of **92 speech features**
- **36 new speech features** introduced that pathological based on modulation spectra, inferior colliculus coefficients, bispectrum, sample and approximate entropy and empirical mode decomposition
- The significance of these features was tested on **3 (English, Spanish and Czech) pathological voice databases** with respect to classification accuracy, sensitivity and specificity

Main goals of this research-

1. According to complex parameterization and consequent robust testing **identify features** that have the **largest discriminative power** in the field of pathological speech analysis.
2. **Design new features** that can quantify **hoarseness, breathiness and non-linearities** in pathological speech signals

Main Goals

3. Prove that the **proposed large set parameterization** approach can provide better classification results (with respect to classification accuracy, sensitivity and specificity) than those published in the field of pathological speech analysis by the other researchers.

4. **Select a database** that has high potential for the future, especially in terms of **speech features design, tuning and testing**

Related terms

. Phonation refers to vocal fold vibration

- **Aphonic**- no vocal fold vibration
- **Normophonic** - no anomalies
- **Dysphonic** - anomalous vocal fold vibration pattern, pathology, something wrong in phonation organs(mainly the larynx)

Classification of Voice pathologies/ disorders

- **Tissue infection** - laryngitis, bronchitis, croup.
- **Systemic changes**- dehydration, pharmacological and drug effects, hormonal changes.
- **Mechanical stress**- vocal nodules, polyps,ulcers, granulomae, laryngocele, hemorrhage.
- **Surface irritation**- laryngitis, leukoplakia, gastroesophageal reflux.
- **Tissue changes**- laryngeal carcinoma, keratosis, papillomas, cysts.
- **Neurological and muscular changes**- bilateral and unilateral vocal fold paralysis, Parkinson's Disease (PD), Amyotrophic Lateral Sclerosis (ALS), myotonic dystrophy, Huntington's Chorea, myasthenia gravis.
- **Abnormal muscle patterns**- conversion aphonia or dysphonia, spasmodic dysphonia, mutational dysphonia, Ventricular phonation.

Acoustic Voice Quality Analysis (AVQA)

- **Set of different methodologies** designed to **quantify acoustic correlates** giving a definition of the **quality of phonation** or speech production
- **Objectives** established in order of difficulty:
  - Dysphonic voice detection,
  - Dysphonic voice grading, and dysphonic voice classification

Aim of AVQA- Aim of AVQA is to design the best methodology to use Voice Correlates in

Voice Pathology detection, grading and classification

Different sets of features

2.1 Features describing phonation
2.2.Features describing tongue movement
2.3 Features describing speech quality
2.4 Segmental features
2.5 Features based on bispectrum
2.6 Features based on wavelet decomposition
2.7 EMD (Empirical Mode Decomposition)

2.1 Features describing phonation

- Explore **significance** of the **most**ly used **speech features** when focusing on the ability of **differentiation between healthy and pathological speech**

Popular features describing pathological voice are

- Fundamental frequency F0
- Parameters describing variability in time (jitter):
  - **PPQ5** (five-point Pitch Perturbation Quotient),
  - **RAP** (Relative Average Perturbation),
  - **jittloc** (average absolute difference between consecutive periods, divided by the average period),
  - **jittabs** (average absolute difference between consecutive periods),
  - **jittddp** (average absolute difference between consecutive differences on neighbor glottal periods, divided by the average period) [39,41–43].

2.2. Features describing tongue movement

- Frequencies of first three formants F1, F2, F3 and

  their bandwidths B1, B2, B3
- Related to volumes of vocal tract cavities.
- Especially the volume of throat and oral cavity is modified by the tongue position.

2.3 Features describing speech quality

Signs of vocal fold dysfunctions are usually associated with breathiness or hoarseness

- **ZCR** (ZeroCrossing Rate) simplest.
- **HZCRR** (High Zero-Crossing Rate Ratio)- takes into account a variation of ZCR in time.
- **FLUF** (Fraction of Locally Unvoiced Frames) describe an impossibility of carrying out periodical glottal closure.

Features based on the **variation of spectrum values** between adjacent frames

- **SF** (Spectral Flux) [56]
- **SDBM (**Spectral Distance Based on Module)
- **SDBP**(Spectral Distance Based on Phase)

2.4 Segmental features

Segmental features are considered as matrices (not only vectors) calculated from the whole signal.

MFCC the **most popular** segmental features in the field of speech signal analysis

MSC (Modulation Spectra Coefficients) provide **information complementary** to MFCC [36,63]. These features can **capture a class of source mechanism** characteristics related to voice quality

Features based on linear prediction

- **LPC** (Linear Predictive Coefficients) [42],
- **PLP** (Perceptual Linear Predictive coefficients) [64],
- **LPCC** (Linear Predictive Cepstral Coefficients) [65],
- **LPCT** (Linear Predictive Cosine Transform coefficients) [42] and
- **ACW** (Adaptive Component Weighted coefficients) [65] were tested

Segmental features(contd)

ICC (Inferior Colliculus Coefficients) parameters that **analyze amplitude modulations** in voice using a biologically inspired model of the inferior colliculus [66].

2.5 Features based on bispectrum

Greater presence of **quadratic coupling** is observed in healthy voice when comparing it to the pathological one

Quadratic coupling can be appropriately described by **bispectrum and features derived from this 2D signal**

- **BII** (Bicoherence Index interference)
- **HFEB** (High Frequency Energy of one-dimensional Bicoherence)
- **LFEB** (low Frequency Energy of one-dimensional Bicoherence)
- **BMII** (Bispectrum Module Interference Index)
- **BPII** (Bispectrum Phase Interference Index)

2.6 Features based on wavelet decomposition

- Wavelet transform is widely used in **coding and speech denoising**.
- Application in **pathological speech analysis** [67].
- Method of **voice quality measurement**
- **Detail coefficients** after the decomposition can be used to **estimate the present noise**
- Calculate **SNR (Signal-to-Noise ratio)**.

2.7 EMD (Empirical Mode Decomposition)

Decompose the arbitrary non-linear and time-varying signal into countable and usually a small number of IMF (Intrinsic Mode Functions).

These functions are modulated in **amplitude and frequency** and their sum gives the original signal

New features introduced

- IMFSNRTKEO (based on TeagerKaiser Energy Operator)
- IMFSNRSEO (based on Squared Energy Operator)
- IMFSNRSE (based on Shannon Entropy)
- IMFNSRTKEO
- IMFNSRSEO
- IMFNSRSE

3. Databases

- **English, Spanish and Czech** databases used during testing procedure.
- Each database represents a **different language group** (Germanic, Romanic and Slavic).
- Advantageous from the **cultural difference point of view.**
- Speakers of different languages exhibit especially different **prosodic characteristics**.
- The **aim of** this **work** is to **find features significant** for the **particular language, selection of language independent**.

3.1 MEEI disordered voice database

- Massachusetts Eye and Ear Infirmary (MEEI) database
- Benchmark in the field of pathological speech analysis
- Commercially available database
- Consists of 53 healthy and 657 pathological speakers
- Class labels- different pathologies (e.g. adductor spasmodic dysphonia, conversion dysphonia, erythema, and hyperfunction).
- The recordings are sampled at f s ¼ 50 kHz or f s ¼ 25 kHz.

3.2 PdA database

- Príncipe de Asturias (PdA) database
- 239 healthy and 200 pathological speakers
- Classes with different organic pathologies (e.g. nodules, polyps, oedemas, and carcinomas).
- The recordings are sampled at f s25 kHz.
- Accuracies are not so high as in the case of MEEI database

3.3 Parkinsonian Speech Database (PARCZ)

- Czech Parkinsonian Speech Database (PARCZ)
- Recorded at St. Anne's University Hospital in the Czech Republic
- 52 healthy speakers and 57 speakers with Parkinson's disease (PD) who suffer from hypokinetic dysarthria
- Contains 91 speech tasks (free speech, reading text, maintained vowels, and diadochokinetic tasks) which are used for an analysis of speech dysfunctions that usually accompany PD
- Sampled at f s ¼ 48 kHz.

Rule of 30

- To decide whether the **corpus size is sufficient for the robust conclusions**
- Comes directly from the binomial distribution, assuming independent trials [85].
- The rule is "**To be 90% confident that the true error rate is within +-30% of the observed error rate, there must be at least 30 errors.**"

Experiments

- All the databases were resampled to f s ¼ 16 kHz
- The data have been divided into 9 groups.
- Two approaches have been considered: gender dependent and gender independent.
- Each database is randomly divided into 75% and 25% training and testing subsets respectively.
- The classifier is evaluated consequently.
- This procedure (data split, classifier tuning and evaluation) is repeated 100 times.
- The resulting accuracy (ACC), sensitivity (SEN) and specificity (SPE) are calculated according to standard formulas

Results

Table 6
Summary of pathological speech detection results represented as mean ± std (%) (SVM – Support Vector Machine with a radial kernel, RF – Random Forest, F – female, M – male, MF – all genders).

ID	Scenario		No. of sel. features	Accuracy		Sensitivity		Specificity	
	Dataset	Gender	(Mean ± std)	SVM	RF	SVM	RF	SVM	RF
M1	MEEI	F	13,996 ± 288	99.5 ± 1.5	**100.0 ± 0.0**	99.3 ± 2.0	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**
M2	MEEI	M	13,561 ± 398	99.2 ± 1.7	**100.0 ± 0.0**	99.1 ± 2.1	**100.0 ± 0.0**	99.3 ± 3.3	**100.0 ± 0.0**
M3	MEEI	MF	15,521 ± 231	99.9 ± 0.4	**100.0 ± 0.0**	99.8 ± 0.5	**100.0 ± 0.0**	99.9 ± 0.7	**100.0 ± 0.0**
P1	PdA	F	9,726 ± 447	75.7 ± 4.3	**78.5 ± 4.9**	72.8 ± 6.5	**77.2 ± 7.7**	78.4 ± 6.8	**79.6 ± 7.3**
P2	PdA	M	9,721 ± 458	78.6 ± 5.1	**80.9 ± 5.1**	71.0 ± 10.1	**74.7 ± 9.8**	84.2 ± 6.1	**85.4 ± 6.7**
P3	PdA	MF	11,540 ± 419	77.7 ± 3.2	**82.1 ± 3.3**	74.9 ± 5.3	**80.0 ± 5.9**	80.1 ± 5.0	**83.8 ± 5.1**
C1	PARCZ	F	2,141 ± 415	65.9 ± 11.9	**67.1 ± 8.3**	**35.3 ± 30.3**	10.3 ± 16.9	79.0 ± 15.3	**91.4 ± 11.3**
C2	PARCZ	M	2,121 ± 489	**67.3 ± 10.7**	66.5 ± 10.3	**50.4 ± 20.2**	42.6 ± 21.6	79.4 ± 14.5	**83.6 ± 16.0**
C3	PARCZ	MF	1,750 ± 331	65.4 ± 7.6	**67.9 ± 6.0**	**39.3 ± 14.9**	31.0 ± 14.6	79.3 ± 10.0	**87.5 ± 8.5**

Assignment

1. What is AVQA? What are its characteristics?
2. What is the rule of 30?
3. What are the different sets of features? Give an example of a set.
4. What are the different types of voice pathologies?