Voice as a Vehicular Tool to Organic and Neurological Disease Tracking: How far we may go?

Pedro Gómez-Vilda, PhD

Member of IEEE Life Sciences and Systems Society
Neuromorphic Speech Processing Lab
Biomedical Technology Center
Universidad Politécnica de Madrid
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid, Spain
Index

- Motivation
- **Gain in Knowledge** from PD Phonation
- Phonation Basics
- Modeling and Simulation: Three Inverse Problems
- Characterizing tremor in PD
- Study Cases
- Future lines
- Conclusions
• The neurological diseases are **affecting a larger segment** of the population in the western world due to increasing life expectancy. For instance PD prevalence in Spain is 1.7/1000 inh., Alzheimer's disease is the sixth-leading cause of death in the United States and the only cause of death among the top 10 in the United States that cannot be prevented, cured or even slowed (5.4 million people diagnosed) http://www.alz.org.

• The **costs of treatment** to grant a minimum life quality will **become unbearable**

• Responses to this critical situation as **early detection** and **treatment monitoring** are to be sought from medicine and engineering

• It is well known that many neurological diseases induce **speech, voicing and phonation** impairments or problems

• Aim: Explore if algorithms developed for the detection and grading of pathological voice **may be extended** to detect and monitor neurological diseases resulting in phonation impairments

• http://www.med.harvard.edu/AANLIB/
• This presentation concentrates mainly in PD, although many conclusions can be extended to other ND types.
• Observable effects of ND in voicing and speech are at least the following:
  – Voicing Dysfunction: Over or under-tone voice (f0), Tremor in sustained phonations (unstable of F0), monotonous voice, poor prosody
  – Impaired fluency: slow leading trails, frequent pauses, excess fillers
  – Dysarthria: reduced ability to produce sharp nasopharyngeal, lingual and bilabial transitions
  – Impaired articulation planning, clumsy articulation, elisions, metatesis
• Many of these effects are caused by damage to cortical areas and descending pathways to the vocal muscles and nerves, affecting the dynamics of the whole system\textsuperscript{1,2}

\textsuperscript{2}Goberman, A., Coelho, C., Robb, M.: Phonatory characteristics of Parkinsonian speech before and after morning medication: the ON and OFF states, J. Com. Disorders, 35, 217-239, 2002
The objective of the present work will be to find new ways of characterizing ND dysfunction in voicing (hypertension, dispersion, unbalance, tremor). Mel-cepstrum parameters classically used in organic voice pathology detection can also be used in detecting ND. Studies have been published on the detection of Parkinson from the degradation of the vowel triangle.

Parkinson’s Disease is a neuro-degenerative illness due to deterioration of neuro-motor centers and pathways in mid-brain. Its manifestation is rigidity of limbs, akinesia, bradikinesia, tremor... It is well known that this disease leaves clues also in voice and speech: hypokinetic dysarthria (phonatory impairment, higher fO due to rigidity of laryngeal muscles resulting in increased vocal fold stiffness: Gobermann et al. J. Com. Dis. 2002)

This raises BURNING QUESTIONS:

Q: May we use dysarthric voice to help measuring PD progress by a simple test?

A: Probably measuring vocal fold stiffness
Q: Do we have a way to measure vocal fold stiffness?
A: Yes, we do!
Q: How?

Main Hypothesis: Meaningful correlates of PD disorder in voice may be derived from vocal fold biomechanics (stiffness)
Secondary Hypothesis: Vocal fold stiffness is affected differently by Organic Larynx Diseases than by Neuromotor Disorders
Is it possible to obtain robust, reliable and semantic features in voice to detect, model and track neurological disease?
Problem framework: How to measure vocal fold stiffness?

What Can We Do?

Parameters Telling Us

Gain in Knowledge
But WHAT IS ‘Gain in Knowledge’?

• Formulating a model to understand how voice is produced and HOW V.F. STIFFNESS INFLUENCE VOICE, by:
  – Analyzing top-down the neurophysiological pathways
    • how are built and connected
    • what do they do
    • building small models explaining each step (divide and conquer)
  – Synthesizing bottom-up a chain of models:
    • inspired in top-down analysis
    • to deconstruct voice to its components
    • related with phonation physiology

• A Gain in Knowledge is any added portion of semantics which may help us in estimating, validating and explaining any phenomenon by the addition of semantics, i.e. understanding better the process
Examples of PD Dysarthria (with PangurBan)
Referencia
• **1st Utterance:**
  - The 1st stop consonant /k/ is long and affricate
  - The 1st /a/ is velarized
  - The interval previous to the 2nd stop consonant is 0.38-0.24=0.14 s
  - The 2nd stop consonant /p/ lacks the vertical plosive burst
  - The 2nd /a/ has a long and slow vowel onset
  - As a result the 2nd stop consonant is configured as a /b/
  - The 2nd /a/ is nasalized, as well as the last one
  - The utterance may be transcribed as /kʰam.ʳãnã/

• **2nd Utterance:**
  - The silence previous to the 2nd stop consonant is 0.32-0.37=0.05 s
  - The 2nd stop consonant /p/ has a vertical plosive burst
  - The nasalization is only evident in the 3rd /a/
  - The utterance may be transcribed as /kʰampanã/
First conclusions

• Fatigue in sustaining phonation possibly due to weak diaphragm control
• Tremor due to a combination of weak laryngeal and diaphragm control (tremor is larger at the end of phonation)
• Diaphragm tremor produces changes in amplitude, but not in the position of formants
• Jaw tremor produces changes in formant positions
• Laryngeal tremor produces changes in biomechanical tension, but not in formant positions
Comp Sp Lab Kay Elemetrics

Jitter, shimmer, HNR and other organic pathol clues

Moderate success

• Uses sustained phonations of /a/ and contrasts them against a UPDRS evaluation by clinicians on 47 patients monitored 6 months every week
• Gender-dependent modelling
Unified Parkinson’s Disease Rating Scale:

- Quantifies subjectively 5 levels of intensity from each behavioral feature:
  - I. Mentation, Behavior and Mood
  - II. Activities of daily living
  - III. Motor Examination

<table>
<thead>
<tr>
<th>1-Intellectual Impairment</th>
<th>12-Turning in bed and adj. bed clothes</th>
<th>23-Finger taps</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Thought Disorder</td>
<td>13-Falling</td>
<td>24-Hand Movements</td>
</tr>
<tr>
<td>3-Depression</td>
<td>14-Freezing when walking</td>
<td>25-Rapid Alternating Hand Movement</td>
</tr>
<tr>
<td>4-Motivation/Initiative</td>
<td>15-Walking</td>
<td>26-Leg Agility</td>
</tr>
<tr>
<td>5-Speech (intelligibility)</td>
<td>16-Tremor</td>
<td>27-Arising from chair</td>
</tr>
<tr>
<td>6-Salivation</td>
<td>17-Sensory complaints</td>
<td>28-Posture</td>
</tr>
<tr>
<td>7-Swallowing</td>
<td>18-Speech (motor)</td>
<td>29-Gait</td>
</tr>
<tr>
<td>8-Handwriting</td>
<td>19-Facial Expression</td>
<td>30-Postural Stability</td>
</tr>
<tr>
<td>9-Cutting food and handling utensils</td>
<td>20-Tremor at rest</td>
<td>31-Body Bradykinesia and Hypokinesia</td>
</tr>
<tr>
<td>10-Dressing</td>
<td>21-Action or Postural Hand Tremor</td>
<td></td>
</tr>
<tr>
<td>11-Hygiene</td>
<td>22-Rigidity</td>
<td></td>
</tr>
</tbody>
</table>
Sapir’s Approach

Vowel Triangle Estimation

Formant-based Parameters

VT Shrinkage, may serve in AD

Vowel Space Area:

\[ VSA = \frac{f_{1i}(f_{2a} - f_{2u}) + f_{1a}(f_{2i} - f_{2u}) + f_{1u}(f_{2i} - f_{2a})}{2} \]

Vowel Articulation Index:

\[ VAI = \frac{f_{2i} + f_{1a}}{f_{2u} + f_{2a} + f_{1u} + f_{1i}} \]

Formant Centralization Ratio:

\[ FCR = \frac{f_{2u} + f_{2a} + f_{1u} + f_{1i}}{f_{2i} + f_{1a}} \]
& More on Sapir’s Approach

<table>
<thead>
<tr>
<th></th>
<th>f1u</th>
<th>f1a</th>
<th>f1i</th>
<th>f2u</th>
<th>f2a</th>
<th>f2i</th>
<th>VSA</th>
<th>VAI</th>
<th>FCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Av. Male</td>
<td>380</td>
<td>650</td>
<td>340</td>
<td>800</td>
<td>1400</td>
<td>2200</td>
<td>709000</td>
<td>0.9760</td>
<td>1.0246</td>
</tr>
<tr>
<td>Av. Female</td>
<td>400</td>
<td>780</td>
<td>380</td>
<td>800</td>
<td>1500</td>
<td>2600</td>
<td>1055000</td>
<td>1.0974</td>
<td>0.9112</td>
</tr>
<tr>
<td>F 75y PD</td>
<td>420</td>
<td>850</td>
<td>280</td>
<td>800</td>
<td>1500</td>
<td>2600</td>
<td>1094000</td>
<td>1.1500</td>
<td>0.8696</td>
</tr>
</tbody>
</table>
These correlates are ambiguous, but together may produce robust markers in ND (PD, AD, LAS, HC, ET, etc.) Even emotional and intentional tremor may be treated.
Voicing is produced by the vocal folds

a) Cover
   Upper Lip (supraglottal)
   Lower Lip (subglottal)
   Body

b) Body mass
   k-1 Cover masses

x y
Data taken from Story (2002)

Total Volume of the fold:
1.5x0.3x0.3x\(\pi/4\)=0.106 cm\(^3\)

Volume of right section:
0.011 cm\(^3\)

This section behaves more as a reactive spring

This section behaves more as an inertial mass
Descending Neural Pathways to Larynx
The innervation of the **transverse and oblique arytenoid muscles** by the **superior, inferior and transversal** laryngeal nerves is responsible of the vocal fold tension by enlargement or shortening of the musculus vocalis, see: Luschei, E. S., Ramig, L. O., Baker, K. L., Smith, M. E., "Discharge characteristics of laryngeal single motor units during phonation in young and older adults and in persons with Parkinson disease", *J. Neurophysiol.*, Vol. 81, 1999, pp. 2131-2139.
Systemic View: From Analysis to Synthesis

Neuromotor Cortex
Articulatory Planning

Hypothalamic Relay & Feedback Loop

Vocal Fold Adduction and Abduction Control

Vocal & Nasal Tract Acoustic Filtering

Laryngeal Nerve Activation

Temporal Phonation Activation

Phonation Bursts

Voiced Segments

Vowel Triangle Aspect Ratios (VAI, FDR)

Glottal Source Correlate

Vocal Fold Strain Estimation

Tremor Amplitude & Frequency Ratios

Biomechanical Estimates

Mechanical & Neural Asymmetry

Vocal & Nasal Tract Estimation & Removal

Mechano-Neural Unbalance
Model Inversion: Three inverse problems

- **First level inversion**
  - Vocal Tract Model: Voice → Glottal Source

- **Second level inversion**
  - 2-mass vocal fold biomechanics: Glottal Source PSD → mass, viscoelasticity

- **Third level inversion**
  - Tremor, over-tension: Vocal Fold Body Stiffness → Cyclicality

In the voice production model of Gunnar Fant it is assumed that the glottal source is produced by a train of delta pulses $\delta(n)$ which are modeled by a Glottal Function $F_g(z)$ to reproduce the glottal source $u(n)$. This signal, when injected in the vocal tract composed by a chain of tubes $F_v(z)$ produces voice $s_l(n)$ which is radiated as $s(n)$. 
• Glottal Profiler: Low-order Adaptive Paired Lattice (typ: 1, 2, 3)
• Vocal Tract Profiler: High-Order Adaptive Paired Lattice (typ: 36 for 16 kHz)
• Wiener Filter: Extra High-Order Adaptive Lattice (typ: 96 for 16 kHz)
Estimation of Glottal Source by Inverse Filtering

(1) Radiation Inverse Filter $H_r(z)$

(2) Mirror Filter $s_w(n)$

(3) Inverse Glottal Filter $H_{gl}(z)$

(4) Vocal Tract Inverse Filter $H_{vt}(z)$

(5) Mirror Filter $s_v(n)$

De-glottalized voice $s_v(n)$

Glottal Residual $s_{gl}(n)$

Recorded voice $s(n)$

Radiation-compensated voice $s_r(n)$

a) Input Voice

b) Derivative of the Forward Pressure Wave

Plenary Session BIOSTEC 2013, Barcelona, 11-15 February
Glottal Correlates

a) Input Voice

b) Derivative of the Forward Pressure Wave

c) Glottal Source

d) Glottal Flow

- time (sec.)
This example shows an order-1 estimation with good profiling.
This example shows an order-1 estimation with good profiling.
First Inversion Validation Results

Glottal Source and Mucosal Wave Correlates for RegVoX.35

Voice PSD and Vocal Tract Transfer Function

Glottal Source Clipping

Glottal Source PSD
• It is a biometrical signature which can be compared in semantic power with the ElectroCardioGram
• Easy to obtain, store and match
• As the electrocardiogram describes semantic correlates in its singularities (both in time and amplitude), the glottogram points to a similar semantics in the biomechanics of the vocal folds
• The technology Glottex® estimates the different temporal and biomechanical patterns from the observed glottal source profile
Glottal Source

Glottal Source and Mucosal Wave Correlates for regvoz-267

ContGap: 0.065
AdduGap: -0.000
PermGap: 0.006

Glottal Source Clipping
Defective Contact and Permanent Gaps

a) Inspiration/Espiration
b) Adduction
c) Abduction
d) Defective Permanent Gap
e) Defective Asymmetric Gap
f) Defective Contact Gap
Second Problem Inversion

David A. Berry, Modal and nonmodal phonation, J. Phonetics, (29) 431-450, 2001
Results for a typical adjustment (body and cover): Case 346 F 34y N
Data from 100 subjects, normophonic, assessed in Hospital Universitario Gregorio Marañón, Madrid, sustained /a/ phonations, gender balanced, 20-50 y
Descriptive statistics avail intuitive results (first level of analysis)
Some parameters show clear cyclicality

Q: How may we measure cyclicality?
A: How about adaptive AR modelling?

\[ \xi_n = \sum_{i=1}^{K} a_i \xi_{n-i} + \varepsilon_n \]
Third Inversion Process

Voiced speech trace \(- S_v(n)\)

Glottal source estimation \(- S_g(n)\)

Power spectral density estimation and matching \(- |S_r(\omega)|\)

Biomechanical parameter estimation
- Dynamic mass \(\mu_{bm}\), stiffness \(\xi_{bm}\), losses \(\sigma_{bm}\)

Error-driven adaptation rule
- \(e_{km}\), \(c_{km}\)

Cyclic parameters estimation from stiffness
- \(c_1\), \(c_2\), \(c_3\)

Tremor estimation
- Frequency \(f_t\), amplitude \(\eta_t\), poles \(r_t\)
### Third Inversion Process: AR Model

**Cyclic Parameter Adaptive Estimation from Stiffness**

\[ \xi_n \rightarrow \{\xi_{Kn}, c_{Kn}\} \]

\[ \{c_{Kn}\} \]

---

**AR Model Hypothesis:**

\[ \xi_n = \sum_{i=1}^{K} a_i \xi_{n-i} + \epsilon_n \]

**Adaptive Model Estimation:**

\[ \{\xi_{Kn}, c_{Kn}\} = \Phi_{Kn} \{\xi_n, W_K, \beta\} \]

**Parameter Disclosing:**

\[ a_{kn} = a_{k-1n} - c_{kn} \tilde{a}_{k-1n} \]

**Behavior in the Freq. Domain:**

\[ H(z) = \frac{1}{1 - \sum_{i=1}^{K} a_i z^{-i}} = \prod_{i=1}^{K} \frac{z}{z - Z_i} \]
Third Inversion Process: Properties

Frequency estimate:

\[ f_{ti} = \frac{\varphi_i}{2\pi} f_s; \]

Amplitude estimate:

\[ \rho_{ti} = \frac{1}{1 - r_i}; \]

\[ \eta_t = \frac{1}{N_k} \sum_{n \in W_k} \left[ \xi_{Kn} - \bar{\xi}_K \right]^2 \]

A 3rd order model grants a real pole and two complex conjugate ones:

\[ c_1 = \frac{a_1 - a_2 a_3}{1 + a_2 - a_1 a_3 - a_3^2}; \quad c_2 = \frac{a_2 - a_1 a_3}{1 - a_3^2}; \quad c_3 = a_3 \]
Why do this model detect cyclicality?

- It may be shown that if the moduli of the poles $r_1 \rightarrow 1$ (larger peak amplitude) the first coefficient $c_1 \rightarrow -1$

- Therefore $c_1$ may be an indicator of cyclicality in v.f. stiffness, i.e., of tremor in voice

- The accompanying coefficients $c_2$ and $c_3$ are used also as co-descriptors, although they do not share the same properties as $c_1$
$z_1 = r_1; \quad z_2 = r_2 e^{j\varphi_2}; \quad z_3 = r_3 e^{j\varphi_3}$

$\zeta = \{z_2, z_3\}; \quad \hat{\zeta} = \{\hat{z}_2, \hat{z}_3\}$

$r_1 \in \mathbb{R}; \quad r_2 = r_3 \in \mathbb{R}; \quad r_2, r_3 > 0$

$0 \leq \varphi_2 \leq \pi; \quad \varphi_3 = -\varphi_2$

$\varepsilon_z = \hat{\zeta} - \zeta = \{\hat{z}_2 - z_2, \hat{z}_3 - z_3\}$

$\left\| \varepsilon_z \right\| = \left[ 2(\hat{r}_2^2 + r_2^2 - 2\hat{r}_2 r_2 \cos(\hat{\varphi}_2 - \varphi_2)) \right]^{1/2}$

$\varepsilon_r = \frac{\left\| \varepsilon_z \right\|}{\left\| \zeta \right\|}$
Normalizing for a large population

Normophonic
50 Males
50 Females
200 ms of vowel /a/
The lack of a tendency (regression line) indicates that c1, c2 and c3 are statistically independent (under 2nd order stat.)

Is this good or bad?
Quartile values indicate balanced distributions
Parametric normalization of these distributions could be possible
Clustering could be carried out based on these distributions

Q: What are these coefficients useful for?
A: They may be used to classify subjects by the tremor in voice
Q: But tremor in voice may not be always associated to pathology, may it be?
### Male Normals (no tremor and tremor)

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Cond.</th>
<th>Tremor</th>
<th>Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100508</td>
<td>Male</td>
<td>Norm.</td>
<td>No</td>
<td>a-b</td>
</tr>
<tr>
<td>100503</td>
<td>Male</td>
<td>Norm.</td>
<td>Yes</td>
<td>c-d</td>
</tr>
<tr>
<td>223211</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>No</td>
<td>e-f</td>
</tr>
<tr>
<td>334866</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>g-h</td>
</tr>
<tr>
<td>100040</td>
<td>Female</td>
<td>Norm.</td>
<td>No</td>
<td>i-j</td>
</tr>
<tr>
<td>100350</td>
<td>Female</td>
<td>Norm.</td>
<td>Yes</td>
<td>k-l</td>
</tr>
<tr>
<td>333282</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>No</td>
<td>m-n</td>
</tr>
<tr>
<td>337523</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>o-p</td>
</tr>
</tbody>
</table>
### Male PD (no tremor and tremor)

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Cond.</th>
<th>Tremor</th>
<th>Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100508</td>
<td>Male</td>
<td>Norm.</td>
<td>No</td>
<td>a-b</td>
</tr>
<tr>
<td>100503</td>
<td>Male</td>
<td>Norm.</td>
<td>Yes</td>
<td>c-d</td>
</tr>
<tr>
<td>223211</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>No</td>
<td>e-f</td>
</tr>
<tr>
<td>334866</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>g-h</td>
</tr>
<tr>
<td>100040</td>
<td>Female</td>
<td>Norm.</td>
<td>No</td>
<td>i-j</td>
</tr>
<tr>
<td>100350</td>
<td>Female</td>
<td>Norm.</td>
<td>Yes</td>
<td>k-l</td>
</tr>
<tr>
<td>333282</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>No</td>
<td>m-n</td>
</tr>
<tr>
<td>337523</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>o-p</td>
</tr>
</tbody>
</table>
### Female Normals (no tremor and tremor)

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Cond.</th>
<th>Tremor</th>
<th>Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100508</td>
<td>Male</td>
<td>Norm.</td>
<td>No</td>
<td>a-b</td>
</tr>
<tr>
<td>100503</td>
<td>Male</td>
<td>Norm.</td>
<td>Yes</td>
<td>c-d</td>
</tr>
<tr>
<td>223211</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>No</td>
<td>e-f</td>
</tr>
<tr>
<td>334866</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>g-h</td>
</tr>
<tr>
<td>100040</td>
<td>Female</td>
<td>Norm.</td>
<td>No</td>
<td>i-j</td>
</tr>
<tr>
<td>100350</td>
<td>Female</td>
<td>Norm.</td>
<td>Yes</td>
<td>k-l</td>
</tr>
<tr>
<td>333282</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>No</td>
<td>m-n</td>
</tr>
<tr>
<td>337523</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>o-p</td>
</tr>
</tbody>
</table>

---

![Graphs showing body stiffness and cyclicity coefficient](image)
Female PD (no tremor and tremor)

<table>
<thead>
<tr>
<th>Case</th>
<th>Gender</th>
<th>Cond.</th>
<th>Tremor</th>
<th>Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>100508</td>
<td>Male</td>
<td>Norm.</td>
<td>No</td>
<td>a-b</td>
</tr>
<tr>
<td>100503</td>
<td>Male</td>
<td>Norm.</td>
<td>Yes</td>
<td>c-d</td>
</tr>
<tr>
<td>223211</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>No</td>
<td>e-f</td>
</tr>
<tr>
<td>334866</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>g-h</td>
</tr>
<tr>
<td>100040</td>
<td>Female</td>
<td>Norm.</td>
<td>No</td>
<td>i-j</td>
</tr>
<tr>
<td>100350</td>
<td>Female</td>
<td>Norm.</td>
<td>Yes</td>
<td>k-l</td>
</tr>
<tr>
<td>333282</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>No</td>
<td>m-n</td>
</tr>
<tr>
<td>337523</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>o-p</td>
</tr>
</tbody>
</table>
Intra- and inter-subject scatter plots ($c_2vsc_1$) for female control and tremor affected PD Patients
Observation vectors are based on parameter averages of body and cover stiffness and first cyclicality parameter:

\[ \bar{x}_{1s} = \mu_{bs} = \left\langle \xi_{bs} \right\rangle_n ; \quad \bar{x}_{2s} = \mu_{cs} \left\langle \xi_{cs} \right\rangle_n ; \quad \bar{x}_{3s} = \left\langle c_{1s} \right\rangle_n \]

Observation matrices are composed of concatenated observation vectors for male or female normophonic subjects according to gender from a database:

\[ X_{sm} = \left[ \bar{x}_{1sm} , \bar{x}_{2sm} , \bar{x}_{3sm} \right] ; \quad X_{sf} = \left[ \bar{x}_{1sf} , \bar{x}_{2sf} , \bar{x}_{3sf} \right] \]

Covariance Matrices are directly derived from observation matrices giving a description of the observation statistical distributions:

\[ C_m = X_{sm}^T X_{sm} ; \quad C_f = X_{sf}^T X_{sf} \]

Parameter averages by the dimension of subjects are all what is needed to proceed with classification:

\[ \chi_m = E \left\langle X_{sm} \right\rangle_s ; \quad \chi_f = E \left\langle X_{sf} \right\rangle_s \]
Classification is based on conditional probabilities of an observation vector from a new subject (patient) of being produced by the model considered.

\[
\Pr(x_q \mid \Gamma_m) = \frac{1}{(2\pi)^{3/2} |C_m|^{1/2}} \int_{-\infty}^{x_q} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-1/2 (\zeta - \chi_m)^T C_m^{-1} (\zeta - \chi_m)} d\zeta
\]

\[
\Pr(x_q \mid \Gamma_f) = \frac{1}{(2\pi)^{3/2} |C_f|^{1/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-1/2 (\zeta - \chi_f)^T C_f^{-1} (\zeta - \chi_f)} d\zeta
\]

The membership of a given subject characterized by an observation vector relative to the group of normophonics or not is given by a Log Likelihood Ratio of the odds:

\[
\lambda_{Nm}(x_q) = \log \frac{\Pr(x_q \mid \Gamma_m)}{1 - \Pr(x_q \mid \Gamma_m)} = \log \left\{\Pr(x_q \mid \Gamma_m)\right\} - \log \left\{1 - \Pr(x_q \mid \Gamma_m)\right\}
\]

\[
\lambda_{Nf}(x_q) = \log \frac{\Pr(x_q \mid \Gamma_f)}{1 - \Pr(x_q \mid \Gamma_f)} = \log \left\{\Pr(x_q \mid \Gamma_f)\right\} - \log \left\{1 - \Pr(x_q \mid \Gamma_f)\right\}
\]
## Detection Results combining stiffness and tremor

### Cyclicality parameters and likelihood ratios

<table>
<thead>
<tr>
<th>Case</th>
<th>G</th>
<th>Cond.</th>
<th>Trem.</th>
<th>$f_t$</th>
<th>$\eta_t$</th>
<th>$\mu_b$</th>
<th>$\mu_c$</th>
<th>$c_1$</th>
<th>$\sigma_{c1}$</th>
<th>$c_2$</th>
<th>$\sigma_{c2}$</th>
<th>$c_3$</th>
<th>$\sigma_{c3}$</th>
<th>$\lambda_{c1}$</th>
<th>$\lambda_{T1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100508</td>
<td>Male</td>
<td>Norm.</td>
<td>No</td>
<td>17,55</td>
<td>0,0049</td>
<td>10,180</td>
<td>5,393</td>
<td>-0,6</td>
<td>0,19</td>
<td>-0,08</td>
<td>0,33</td>
<td>0,05</td>
<td>0,15</td>
<td><strong>0.48</strong></td>
<td>-0,44</td>
</tr>
<tr>
<td>100503</td>
<td>Male</td>
<td>Norm.</td>
<td>Yes</td>
<td>5,59</td>
<td>0,009</td>
<td>12,134</td>
<td>7,010</td>
<td>-0,85</td>
<td>0,03</td>
<td>-0,03</td>
<td>0,1</td>
<td>0,14</td>
<td>0,16</td>
<td><strong>-1.53</strong></td>
<td>-0,28</td>
</tr>
<tr>
<td>223211</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>No</td>
<td>10,78</td>
<td>0,0136</td>
<td>14,145</td>
<td>19,498</td>
<td>-0,57</td>
<td>0,05</td>
<td>-0,16</td>
<td>0,11</td>
<td>0,28</td>
<td>0,19</td>
<td><strong>0.70</strong></td>
<td>-13,9</td>
</tr>
<tr>
<td>334866</td>
<td>Male</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>5,39</td>
<td>0,0342</td>
<td>13,777</td>
<td>14,498</td>
<td><strong>-0,89</strong></td>
<td>0,02</td>
<td>0,04</td>
<td>0,19</td>
<td>0,26</td>
<td>0,13</td>
<td><strong>-1.88</strong></td>
<td>-6,03</td>
</tr>
<tr>
<td>100040</td>
<td>Female</td>
<td>Norm.</td>
<td>No</td>
<td>10,66</td>
<td>0,0034</td>
<td>19,227</td>
<td>14,023</td>
<td>-0,63</td>
<td>0,04</td>
<td>0</td>
<td>0,06</td>
<td>0,24</td>
<td>0,07</td>
<td><strong>-0.21</strong></td>
<td>-0,03</td>
</tr>
<tr>
<td>100350</td>
<td>Female</td>
<td>Norm.</td>
<td>Yes</td>
<td>7,49</td>
<td>0,006</td>
<td>20,247</td>
<td>20,373</td>
<td>-0,84</td>
<td>0,02</td>
<td>-0,25</td>
<td>0,07</td>
<td>0,27</td>
<td>0,12</td>
<td><strong>-1.82</strong></td>
<td>-0,45</td>
</tr>
<tr>
<td>333282</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>No</td>
<td>16,34</td>
<td>0,0597</td>
<td>17,276</td>
<td>12,815</td>
<td>-0,52</td>
<td>0,04</td>
<td>0,03</td>
<td>0,04</td>
<td>-0,01</td>
<td>0,11</td>
<td><strong>1.19</strong></td>
<td>-11,4</td>
</tr>
<tr>
<td>337523</td>
<td>Female</td>
<td>Par. Dis.</td>
<td>Yes</td>
<td>5,15</td>
<td>0,0383</td>
<td>25,314</td>
<td>30,274</td>
<td><strong>-0,91</strong></td>
<td>0,03</td>
<td>-0,25</td>
<td>0,06</td>
<td>0,13</td>
<td>0,13</td>
<td><strong>-2.91</strong></td>
<td>-5,36</td>
</tr>
</tbody>
</table>

- If tremor is above 8-10 Hz it is not perceived as tremor anymore
- $c_1$ is a good detector of tremor
- PD condition is best assessed combining tremor detection ($c_1 > 0.85$) and cover stiffness
- Of course, a significance study is required
• Other ND alterations in voice (AD, LAS, HC, etc.)
• Combination of larynx biomechanical, diaphragmal, nasopharyngeal, and mandibular, having in mind that evidence shows that:
  • Larynx biomechanics alters $f_t$ and rMSA (tremor frequency and amp.)
  • Diaphragm dystonia alters voicing volume
  • Nasopharyngeal alters consonant stops, vowel onset and nasalization
  • Mandibular, lingual and mental dystonia induce tremor in formants, reduces vowel triangle and alters articulation positions
• Emotion alteration detection from running speech
• Emotion in singing voice (stage fright: kind of mild transitory ND disorder)
• Side effects of drug dosage in Psychiatric Disease Treatment showing PD syndrome
To conclude

- PD leaves important clues in vocal fold body stiffness: over-tenseness and tremor
- Overtenseness and tremor may be graded using normophonic databases
- Tremor may be not perceived over 10 Hz
- Tremor may be characterized using 3rd order all-pole systems
- Overtenseness and tremor may grant PD patient evaluation
- It is important to distinguish organic from PD overtenseness
- It is important to distinguish essential from emotional, intentional or pathological tremor
- Larger and more specific databases are required (careful evaluation of inter-pathological definitions)
- Better modelling of upper neural pathways have to be investigated
• Techniques:
  – Pitch and energy contours
  – Articulatory: distortions in formant space
  – Temporal description of dysarthria: velo-pharyngeal switch, lip
    coordination, VOT
  – Alterations in vocal fold biomechanics
  – Further research in mid-brain and upper pathway deterioration from
    phenomena timing and qualifying: MEG

• Applications:
  – Emotional state description
  – Organic pathology monitoring
  – Neurological deterioration evaluation
  – Speech and Singing education and rehabilitation
Pedro Gómez-Vilda, PhD
Head of Research Group GIAPSI

Neuromorphic Speech Processing Lab
Center for Biomedical Technology
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid, Spain

Tel: +34.913367384; +34.913364661
Fax: +34.913366601

E-mail:
pedro.gomez@ctb.upm.es
pedro@fi.upm.es

URL: www.ctb.upm.es