Organic and Neurologic Disease Detection and Monitoring from Voice:
Where we are and how far we may go?

Pedro Gómez, Víctor Nieto, Victoria Rodellar, Rafael Martínez, Agustín Álvarez, Luis M. Mazaira, E. Bartolomé

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

Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Motivation

- The neurological diseases are **affecting a larger segment** of the population in the western world due to increasing life expectancy (PD prevalence in Spain is 1.7/1000 inh.)
- 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

Objectives

- Study the alterations of biomechanical parameters of vocal folds with neurologic pathology
- Elaborate protocols for monitoring treatment impact in longitudinal studies
- Produce a database of voice records from neurologic pathology patients
- Discriminate effects due to other neurologic pathologies and emotional arousal
- Study the possibility of extending biomechanical feature descriptors to other neurological and/or cognitive disorders (Alzheimer, Hunttington's Korea, Amyotrophic Lateral Schlerosis, etc.)

Considerations

- Observable effects of neurological disease in voicing and speech are at least the following:
  - Tremor in sustained phonations (unstable of F0), monotonous, low tone, poor prosody
  - Impaired fluency, slow leading trails, frequent pauses, excess fillers
  - Disarthia, reduced ability to produce sharp nasopharyngeal, lingual and bilabial transitions
  - Impaired articulation planning, clumsy articulation, elisions, metatesis
- Many of these effects are caused by impairments to the vocal muscles and nerves, affecting the dynamics of the whole system[1]

[1]Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., Ramig, L. O.: Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, 56 (4), 1015-1022, 2009.

Problem Definition

- 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 dysartrhia (phonatory impairment, higer f0 due to rigidity of laryngeal muscles resulting in increased vocal fold stiffness, Gobermann et al. J. Com. Dis. 2005)

- This raises BURNING QUESTIONS:
  Q: Are v. f. overtension and tremor unique to PDP (Parkinson's Disease Patients)?
  A: Possibly not → Differentiation methods are needed
  Q: Is PD dysarthria related to PD progress?
  A: Possibly there is a change in v. f. stiffness related to f0
  Q: May we use dysarthric voice to measure PD progress?
  A: Probably so if we can differentiate PD changes from other pathologies with similar effects in voice
  Q: And if YES, HOW?
  A: Possibly measuring v. f. rigidity
  Q: How to measure vocal fold rigidity?
  A: Probably measuring vocal fold stiffness

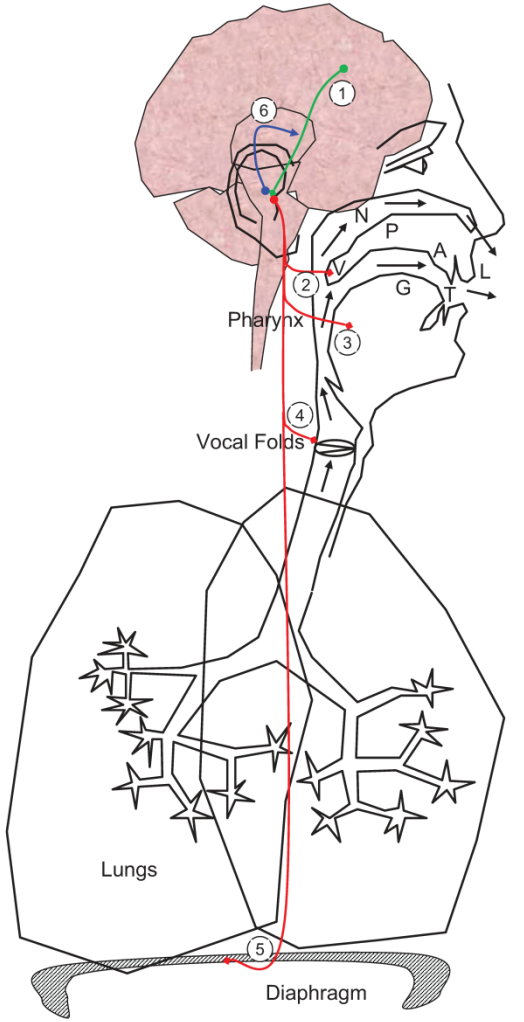
Stem Question

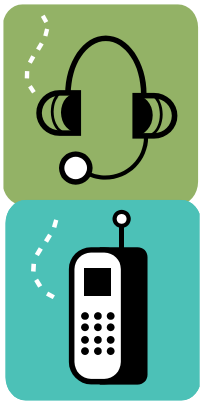
Q: Do we have a way to measure vocal fold stiffness?

A: Yes, we do!

Q: How?

Problem framework: How to measure vocal fold stiffness?





WhatCanWeDo?

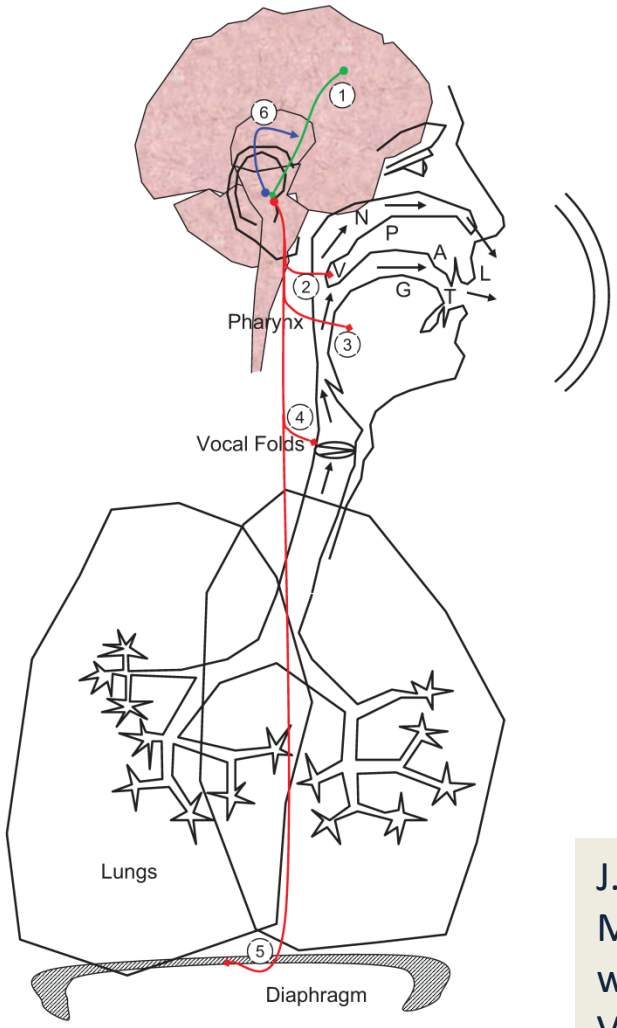
ParametersTellingUs

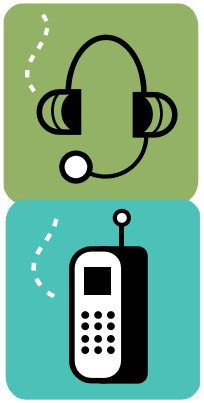
Gain in Knowledge

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
- It seems too easy to be TRUE!
- Where is the TRICK?

Early work from Gamboa et al.





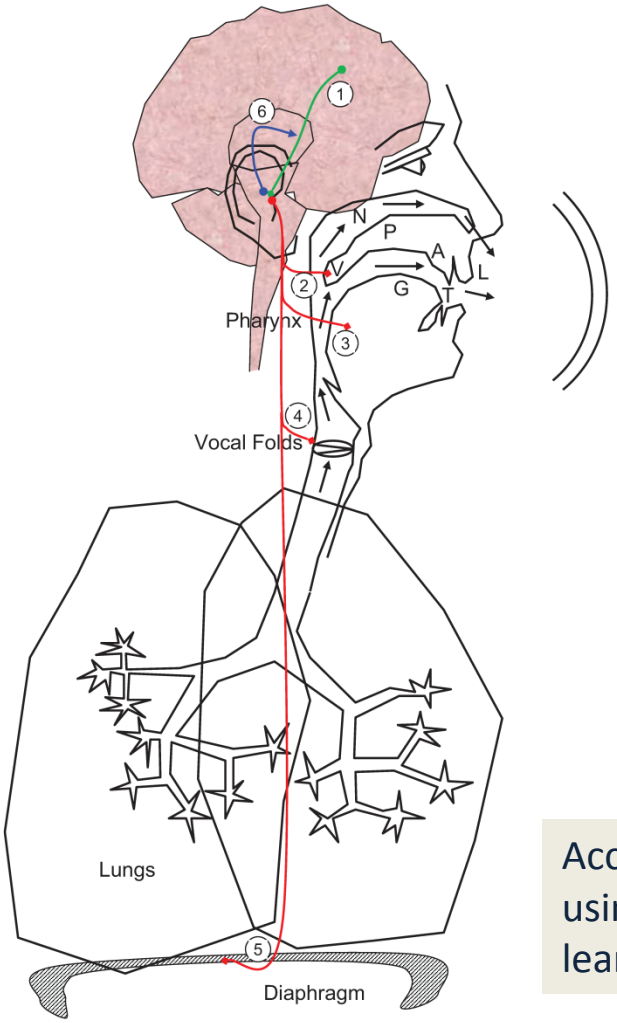
Comp Sp Lab Kay Elemetrics

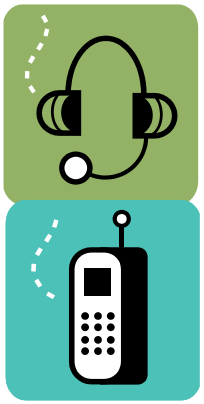
Jitter, shimmer, HNR and other organic pathol clues

Moderate success

J. Gamboa, F. J. Jiménez, A. Nieto, J. Montojo, M. Ortí, J. A. Molina, E. García, I. Cobeta. Acoustic Voice Analysis in Patients with Parkinson's Disease Treated with Dopaminergic Drugs, J. Voice (11) 3, 314-320, 1977

Tsanas' Approach





MFCC extraction, GMM, SVM Clustering

MFCC's

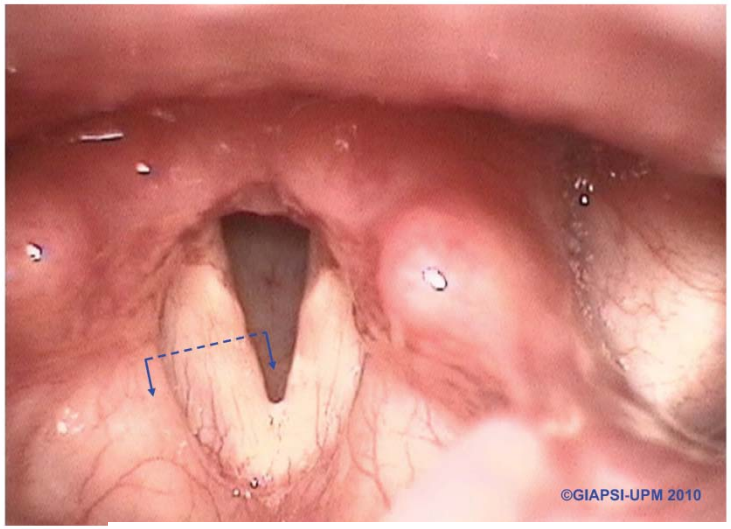
Robust, although Low Semantics

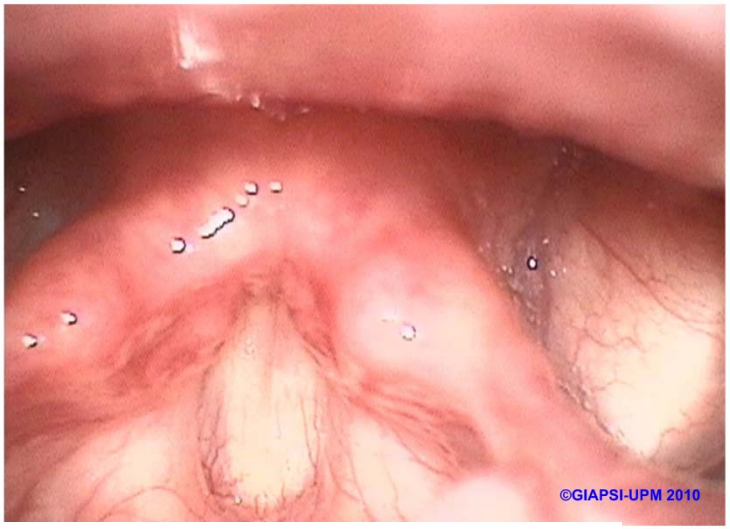
Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning, A. Tsanas, D. Phil. Thesis, University of Oxford, UK, 2012.

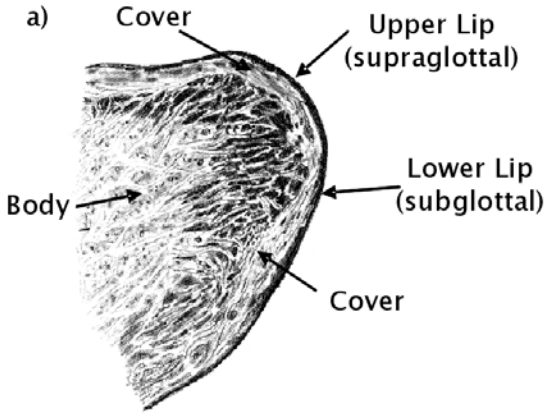
Summary

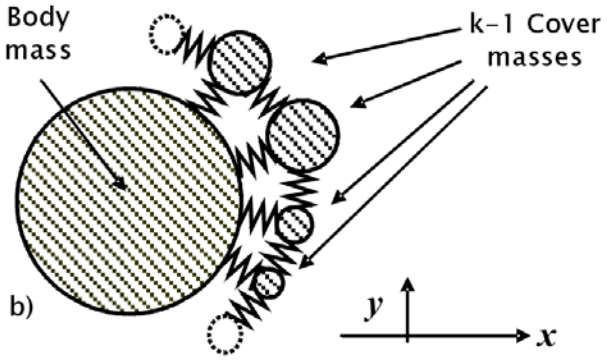
- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Voicing is produced by the vocal folds

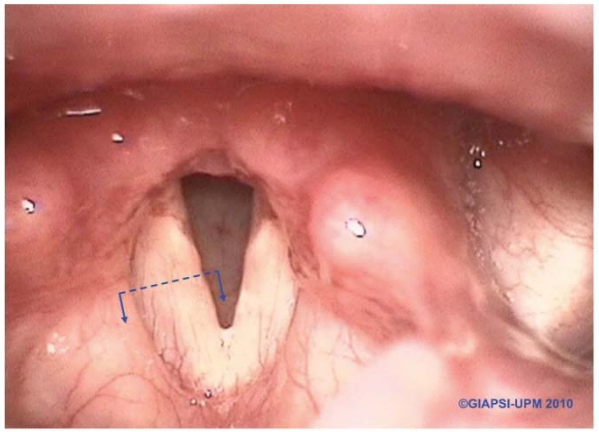


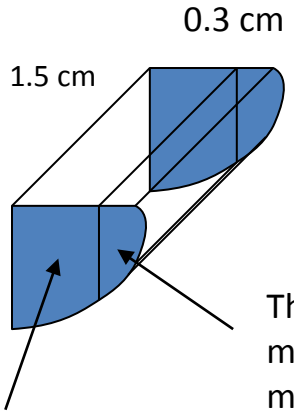






The phonation cycle





This section behaves more as a reactive spring

This section behaves more as an inertial mass

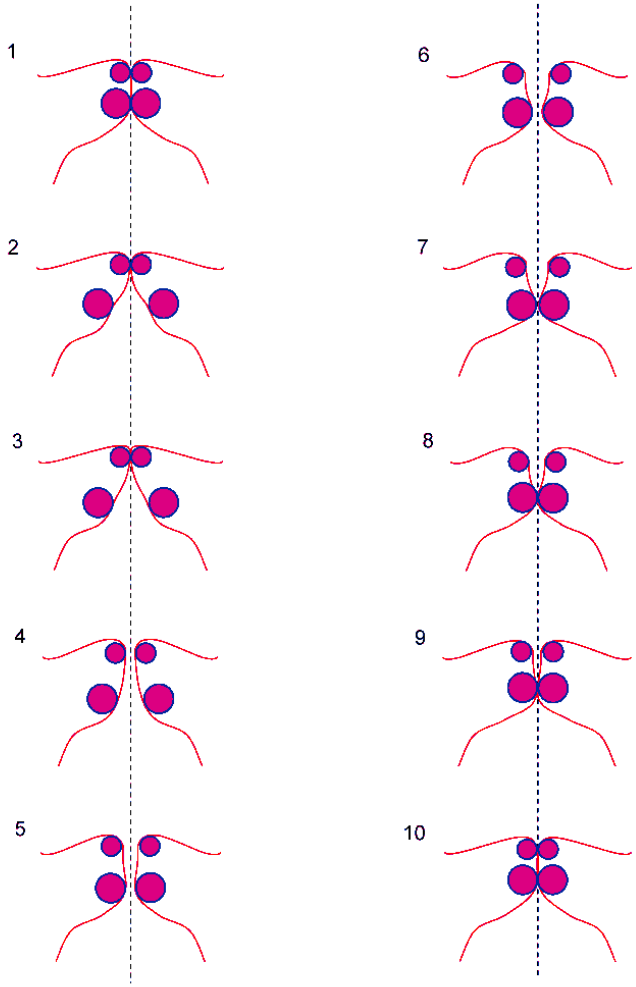
Data taken from Story (2002)

Total Volume of the fold:

$1.5 \times 0.3 \times 0.3 \times \pi/4 = 0.106\ cm^3$

Volume of right section:

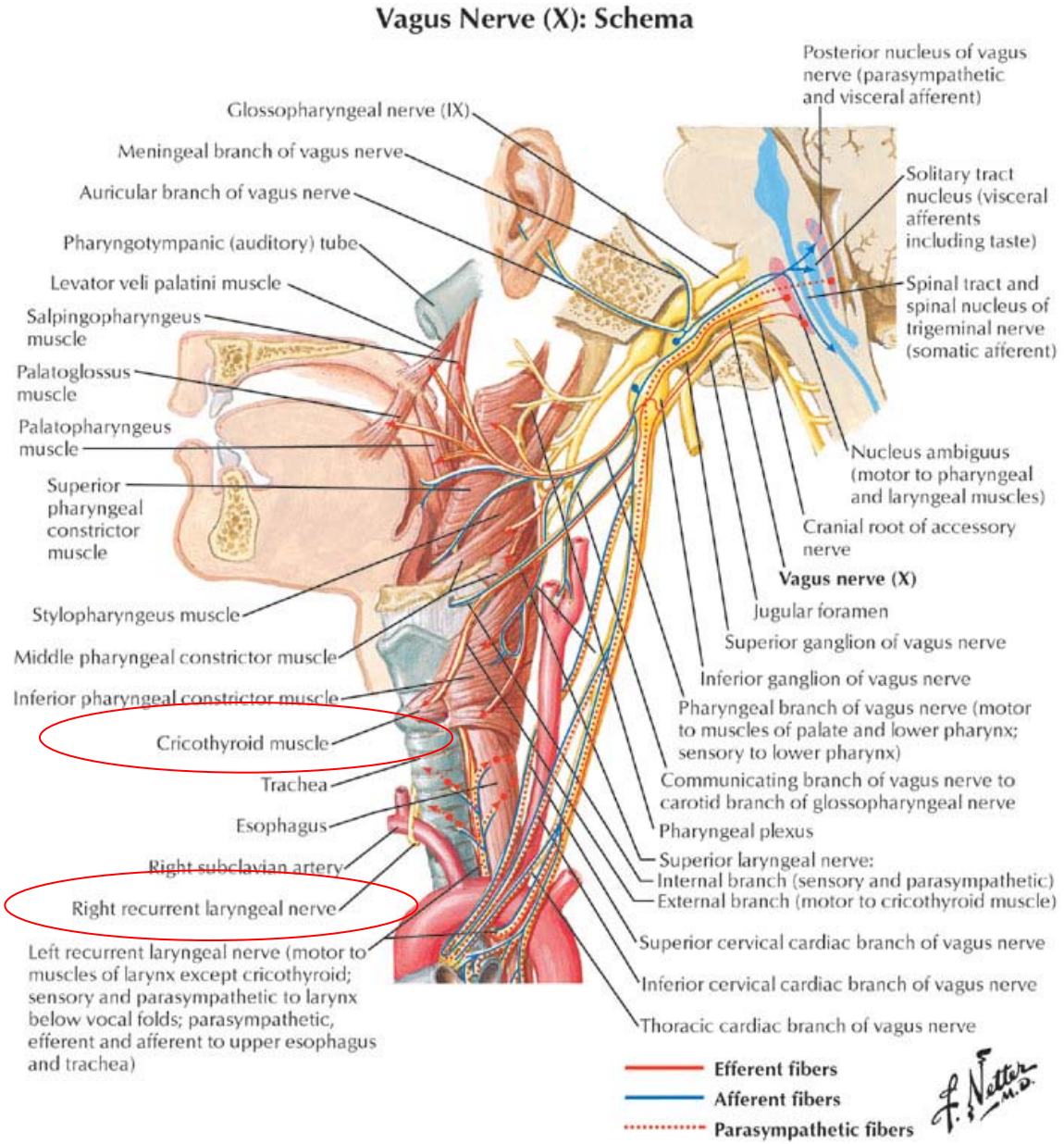
$0.011\ cm^3$



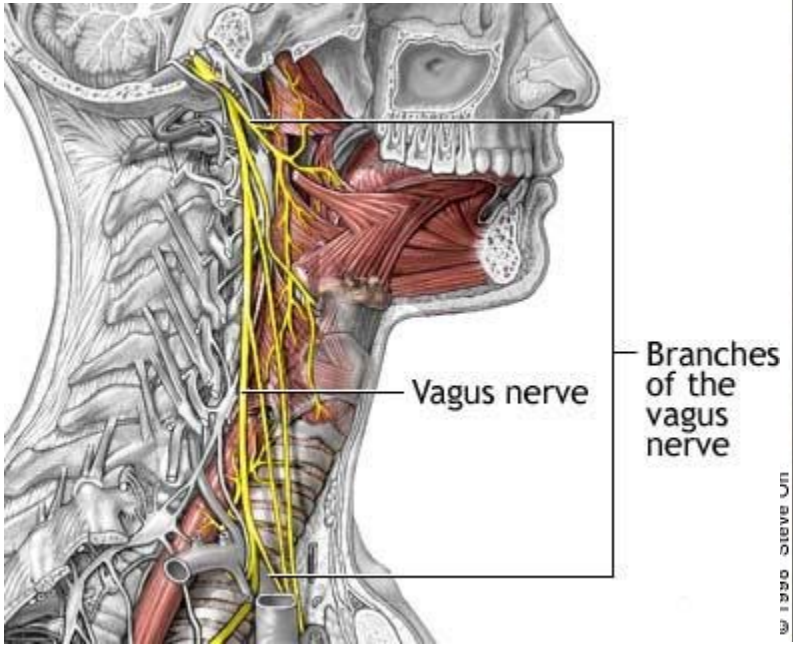
Summary

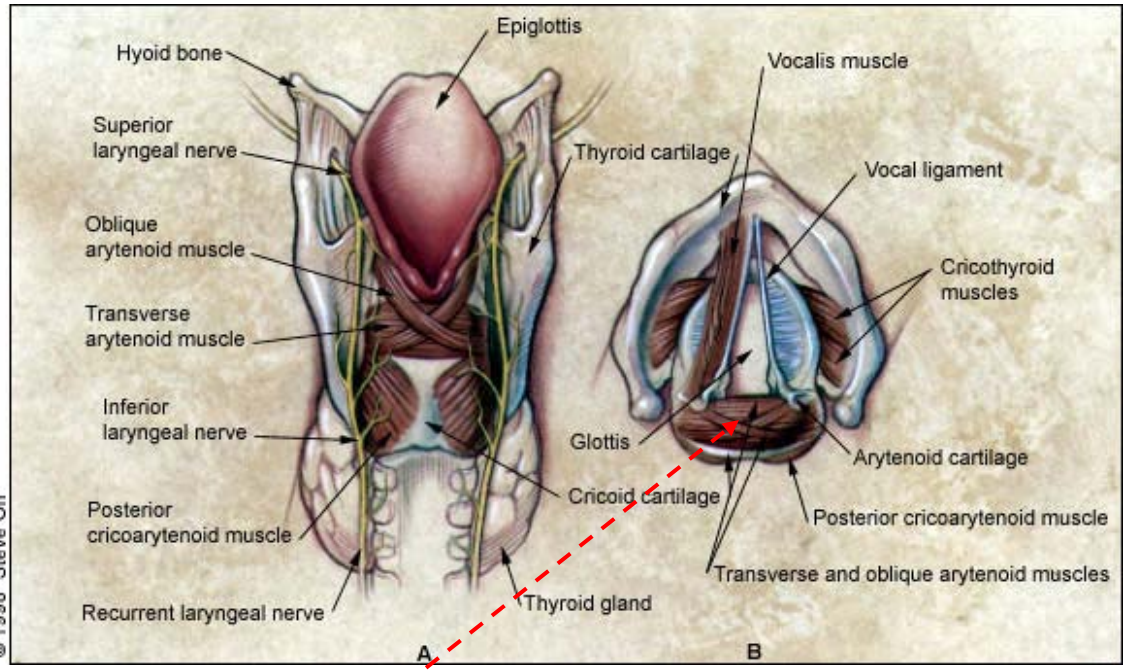
- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Descending Neural Pathways to Larynx



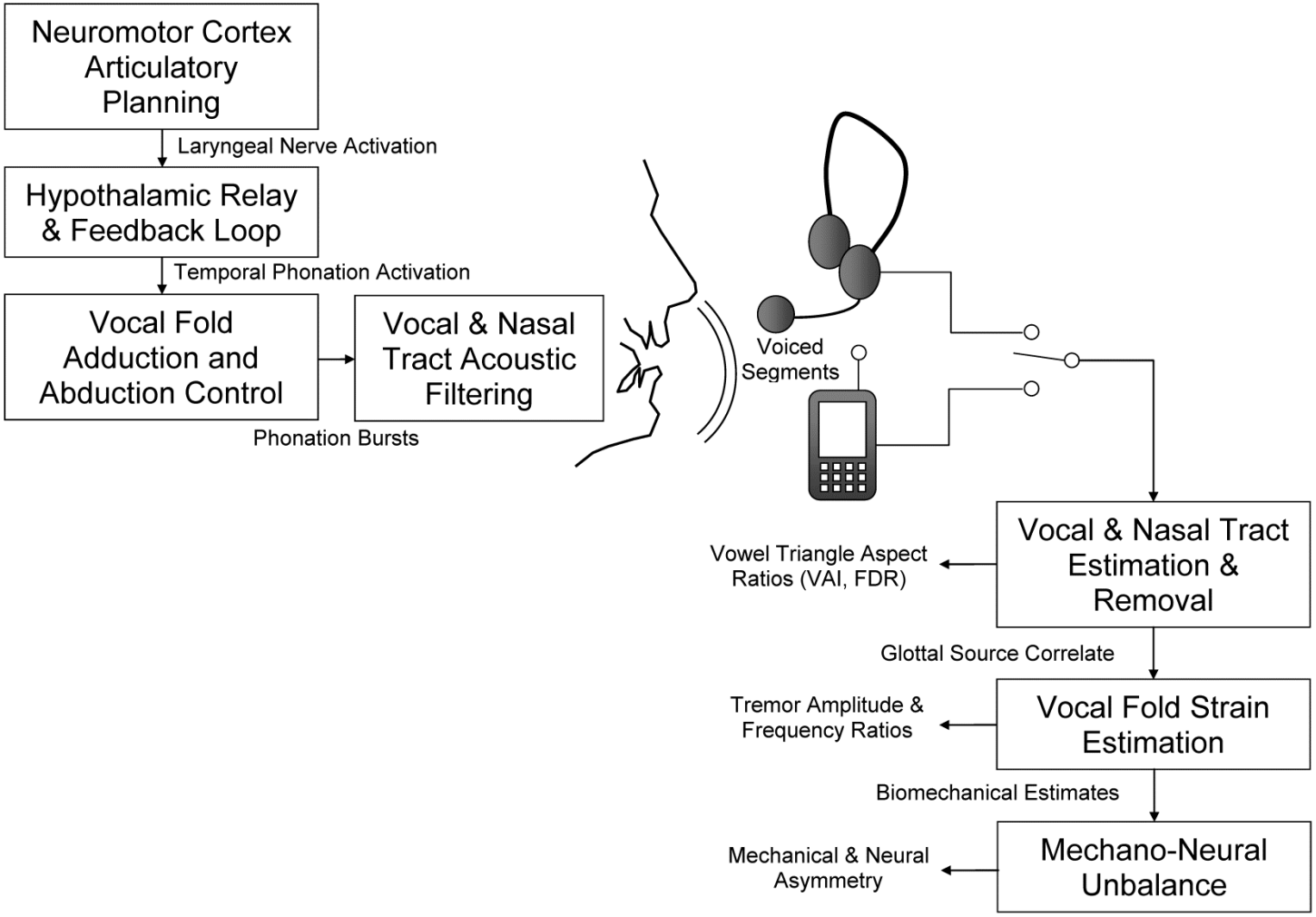
Larynx Innervation





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

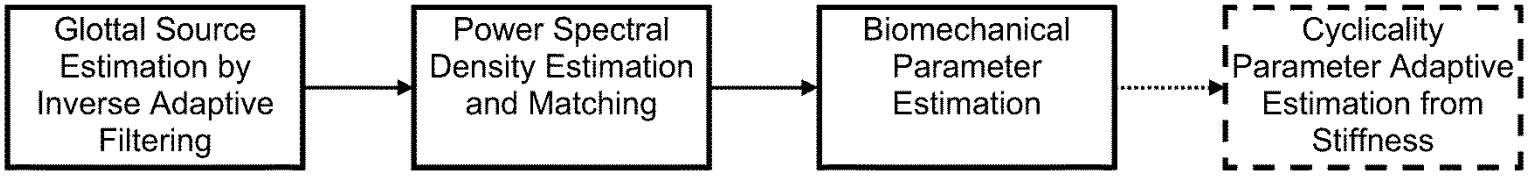


Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Model Inversion: Three inverse problems

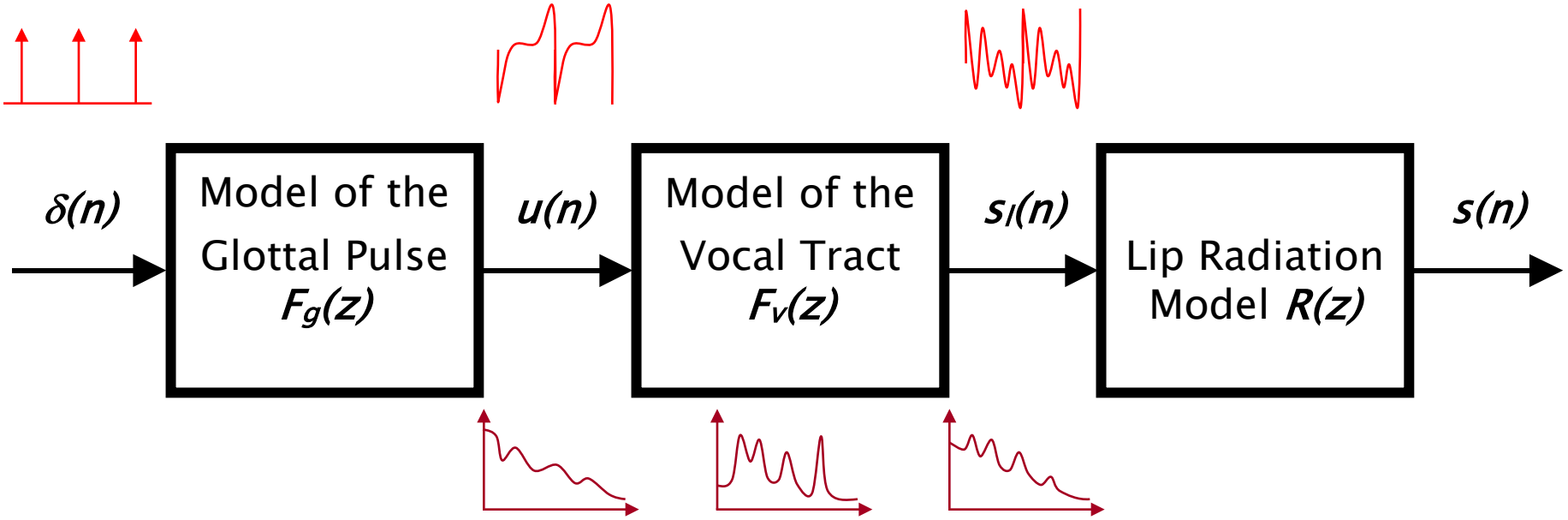
- First level inversion
  - Vocal Tract Model: Voice→Glottal Source
- Second level inversion
  - 2-mass vocal fold biomechanics: Glottal Source PSD→mass, viscoelasticity
- Third level model
  - Tremor, over-tension: Vocal Fold Body Stiffness→Cyclicality



Gómez, P., Fernández, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira, L. M., Martínez, R, Godino, J. I.: Glottal Source Biometrical Signature for Voice Pathology Detection, Speech Communication, Vol. 51, pp. 759-781 (2009)

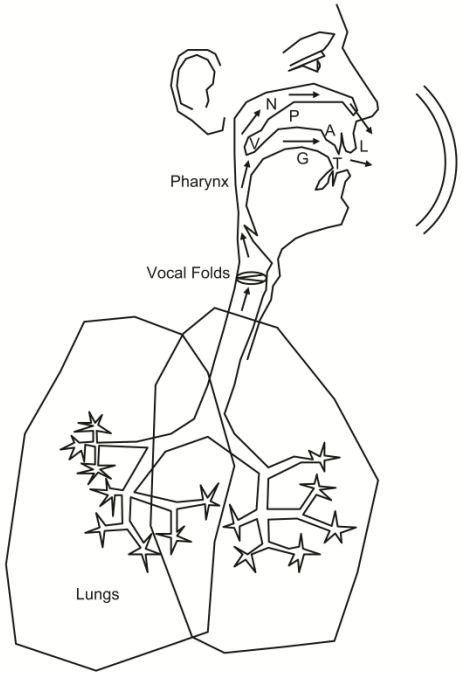
First Model Inversion: Glottal Source Reconstruction

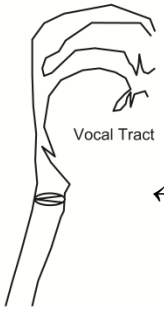
In the voice production model of Gunnar Fant it is assumed that the glottal source is produced by a train of delta pulses δ(n) which are modeled by a Glottal Function $F_g(z)$ to reproduce the glottal source u(n). This signal, when injected in the vocal tract composed by a chain of tubes $F_v(z)$ produces voice $s_l(n)$ which is radiated as s(n).

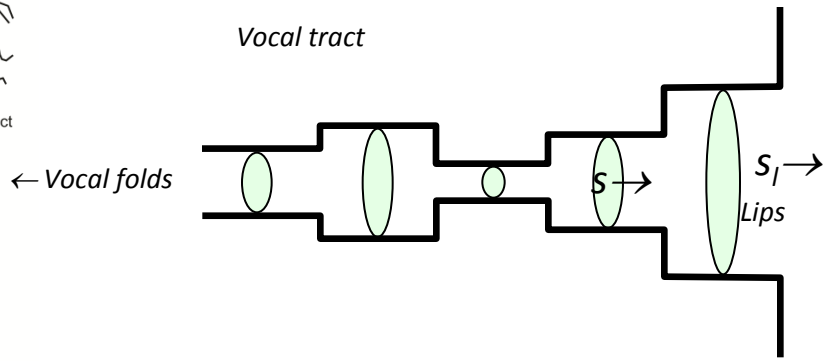


The Vocal Tract as a chain of tube sections

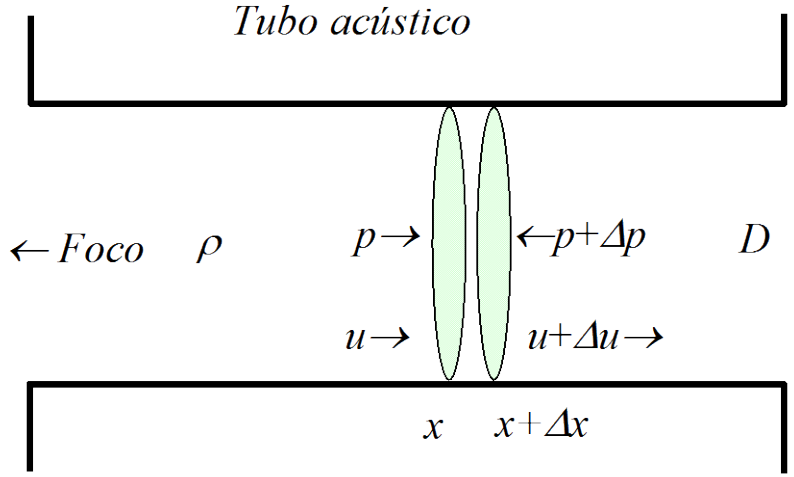
Under the plane wave hypothesis the phonatory apparatus may be seen from vocal folds as a series of concatenated tubes of variable section.







Plane waves in tubes



From Newton's View:

$$pS-(p+\Delta p)S=\rho\Delta x\frac{\partial u}{\partial t}\Rightarrow\frac{\partial p}{\partial x}=-\rho\frac{\partial u}{\partial t}$$

$$\frac{\partial^2 u}{\partial x^2}=-K\frac{\partial^2 p}{\partial x\partial t}$$

From Hooke's View:

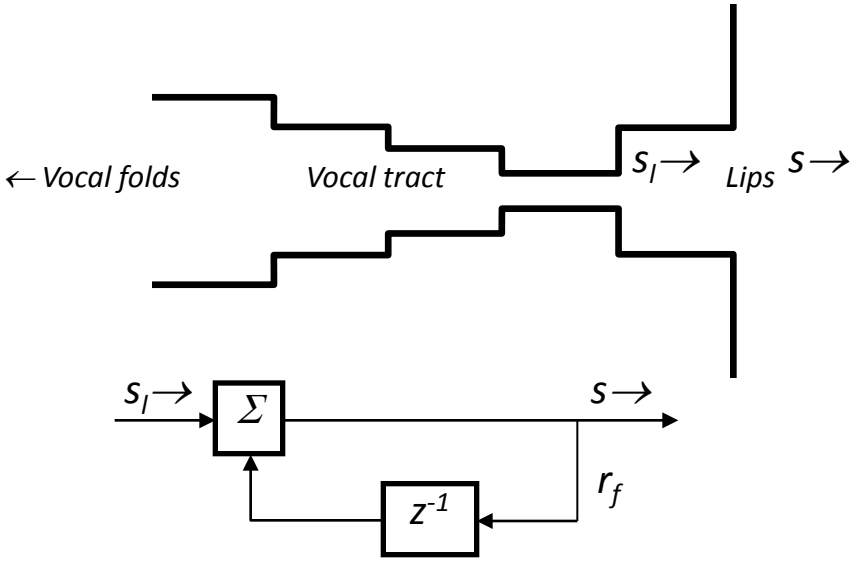
$$uS-(u+\Delta u)S=K\Delta x\frac{\partial p}{\partial t}\Rightarrow\frac{\partial u}{\partial x}=-K\frac{\partial p}{\partial t}$$

$$\frac{\partial^2 p}{\partial t\partial x}=-\rho\frac{\partial^2 u}{\partial t^2}$$

$$\frac{\partial^2 p}{\partial t^2}=\frac{\rho}{K}\frac{\partial^2 u}{\partial t^2}$$

Wave Equation

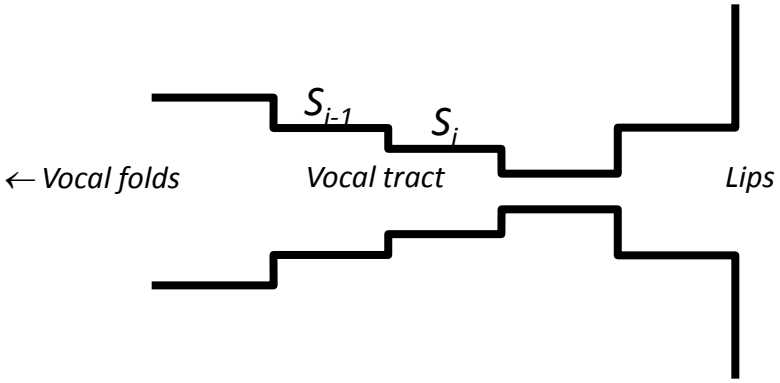
Radiation Model



$$r_f = \frac{Z_f - Z_l}{Z_f + Z_l} = \frac{\frac{\rho c}{S_f} - \frac{\rho c}{S_l}}{\frac{\rho c}{S_f} + \frac{\rho c}{S_l}} = \frac{S_l - S_f}{S_l + S_f}$$

$$R(z) = \frac{S(z)}{S_l(z)} = \frac{1}{1 - r_f z^{-1}}$$

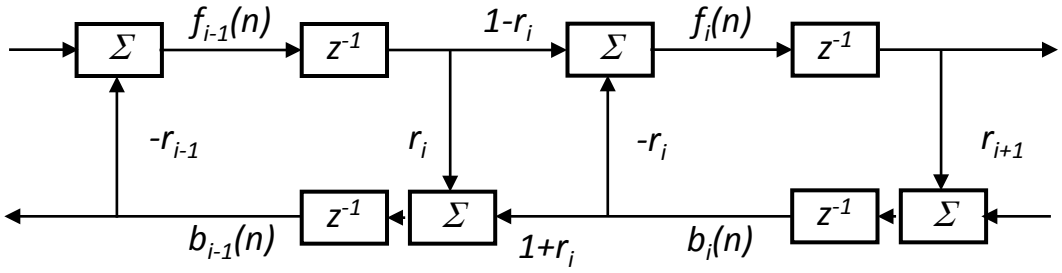
Vocal Tract Model



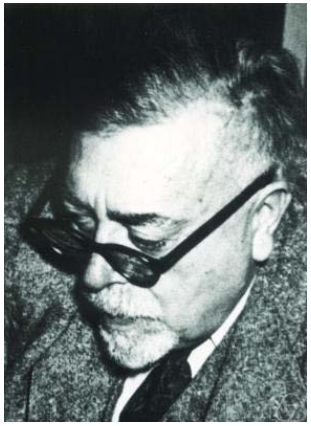
$$f_i(n) = (1 - r_i) f_{i-1}(n-1) - r_i b_i(n)$$

$$b_{i-1}(n) = r_i f_{i-1}(n-2) + (1 + r_i) b_i(n)$$

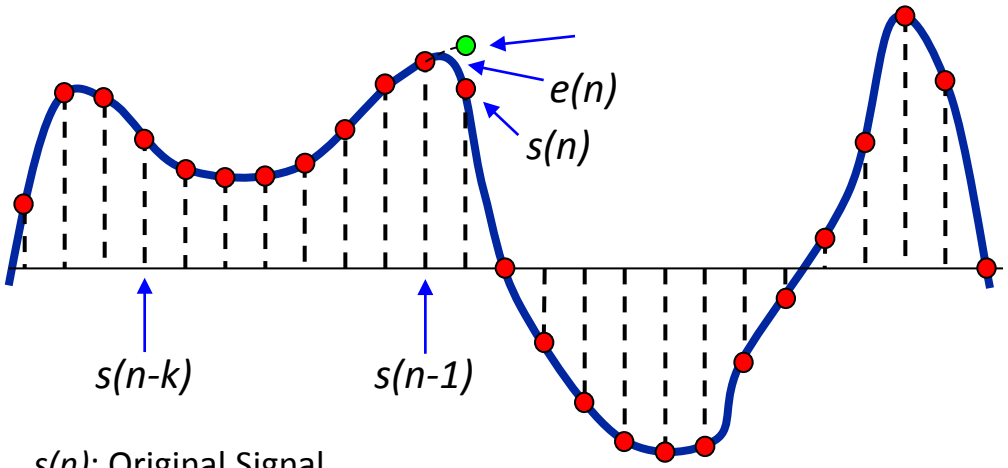
$$r_i = \frac{Z_i - Z_{i-1}}{Z_i + Z_{i-1}} = \frac{\frac{\rho c}{S_i} - \frac{\rho c}{S_{i-1}}}{\frac{\rho c}{S_i} + \frac{\rho c}{S_{i-1}}} = \frac{S_{i-1} - S_i}{S_{i-1} + S_i}$$



Linear Prediction goes back to Gauss (C. F.)



Linear Prediction



s(n): Original Signal

$\hat{s}(n)$: Predicted Signal

e(n): Prediction Error

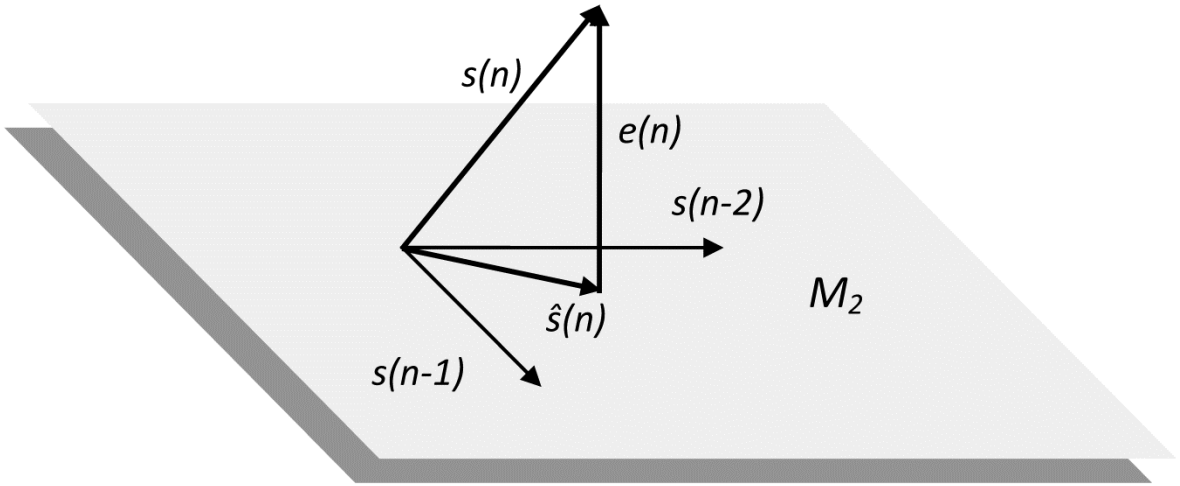
$$\hat{s}(n) = a_1 s(n-1) + a_2 s(n-2) + \ldots + a_K s(n-K) = \sum_{i=1}^{K} a_i s(n-i)$$

$$e_K(n) = s(n) - \hat{s}(n) = \sum_{i=0}^{K} h_i s(n-i); \quad h_0 = 1; \quad h_i = -a_i \quad 1 \le i \le K$$

Normal Equations: Orthogonality Principle

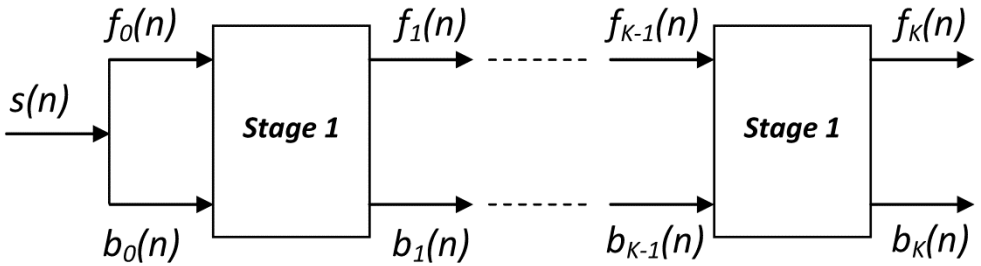
$$\sum_n s(n-j)s(n) = \sum_{i=1}^{K} a_i \sum_n s(n-j)s(n-i)$$

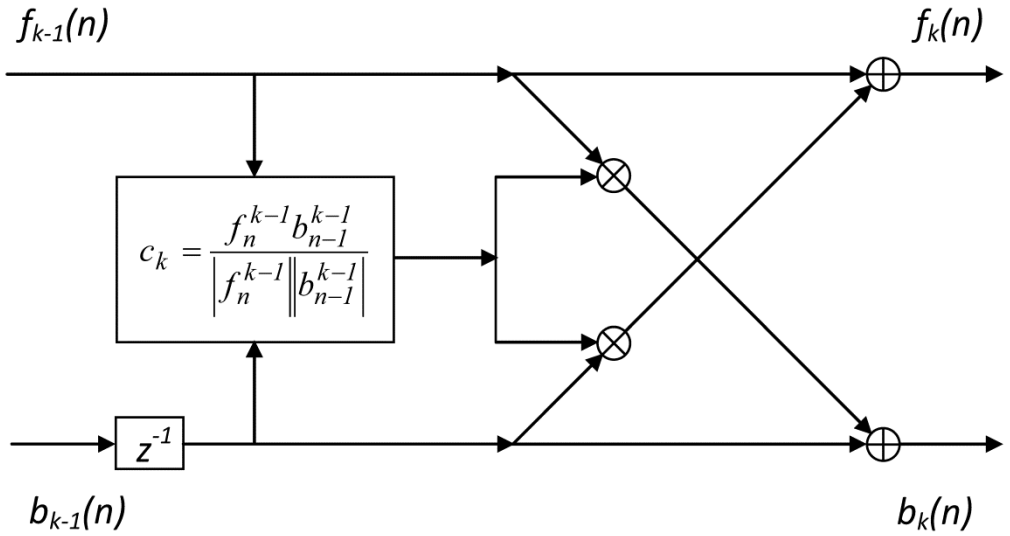
$$\sum_n s(n-j)e(n) = 0; \quad 1 \le j \le K$$



Lattice Prediction-Error Filters

- 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)



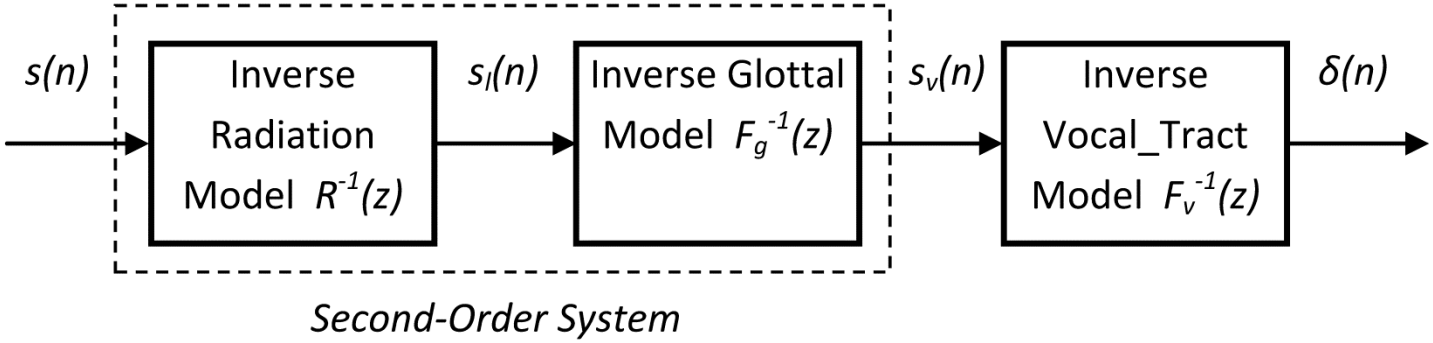


Excitation Model Compensation

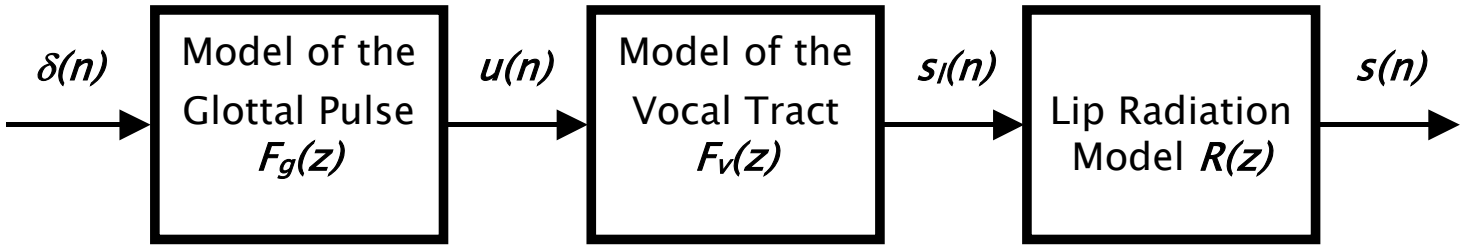
$$e(n) = u(n) - \gamma u(n-1)$$

$$H_g(z) = 1 - \gamma z^{-1}$$

Radiation and Excitation Models may be compensated by a PEF of second order



Glottal Pulse Estimation



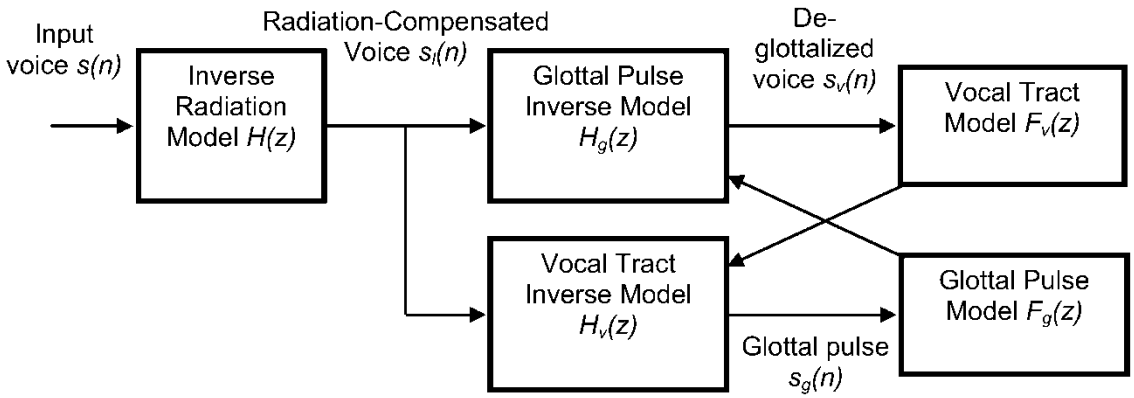
$$\mathbf{s} = \{\{\delta * \mathbf{f_g}\} * \mathbf{f_v}\} * \mathbf{r} = \{\mathbf{f_g} * \mathbf{f_v}\} * \mathbf{r} = \mathbf{s_l} * \mathbf{r}$$

$$\mathbf{s} * \mathbf{h}_r = \{\mathbf{s_l} * \mathbf{r}\} * \mathbf{h}_r = \mathbf{s_l} * \{\mathbf{r} * \mathbf{h}_r\} \cong \mathbf{s_l}$$

$$\mathbf{s_l} = \mathbf{f_g} * \mathbf{f_v}$$

$$\mathbf{s_l} * \mathbf{h_{v0}} = \{\mathbf{f_g} * \mathbf{f_v}\} * \mathbf{h_{v0}} = \mathbf{f_g} * \{\mathbf{f_v} * \mathbf{h_{v0}}\} \cong \mathbf{f_g} = \mathbf{u_0}$$

Glottal Signal Estimation from Voice Inverse Filtering

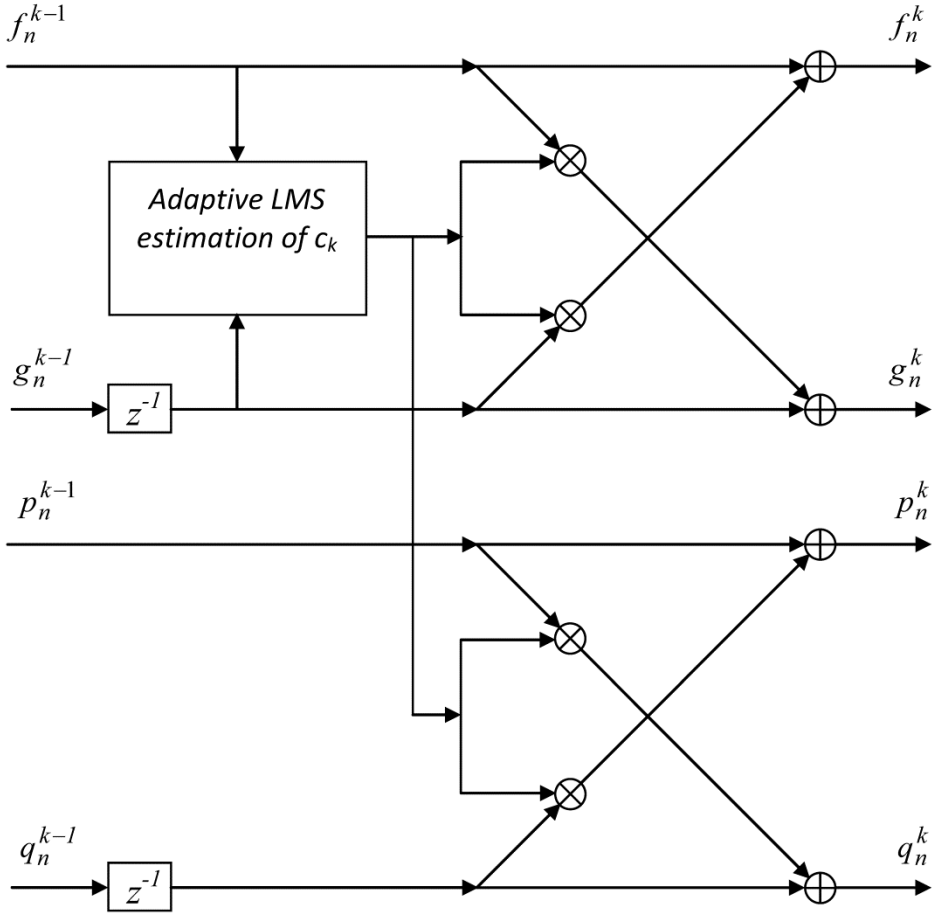


Inspired in P. Alku's well-known methodology

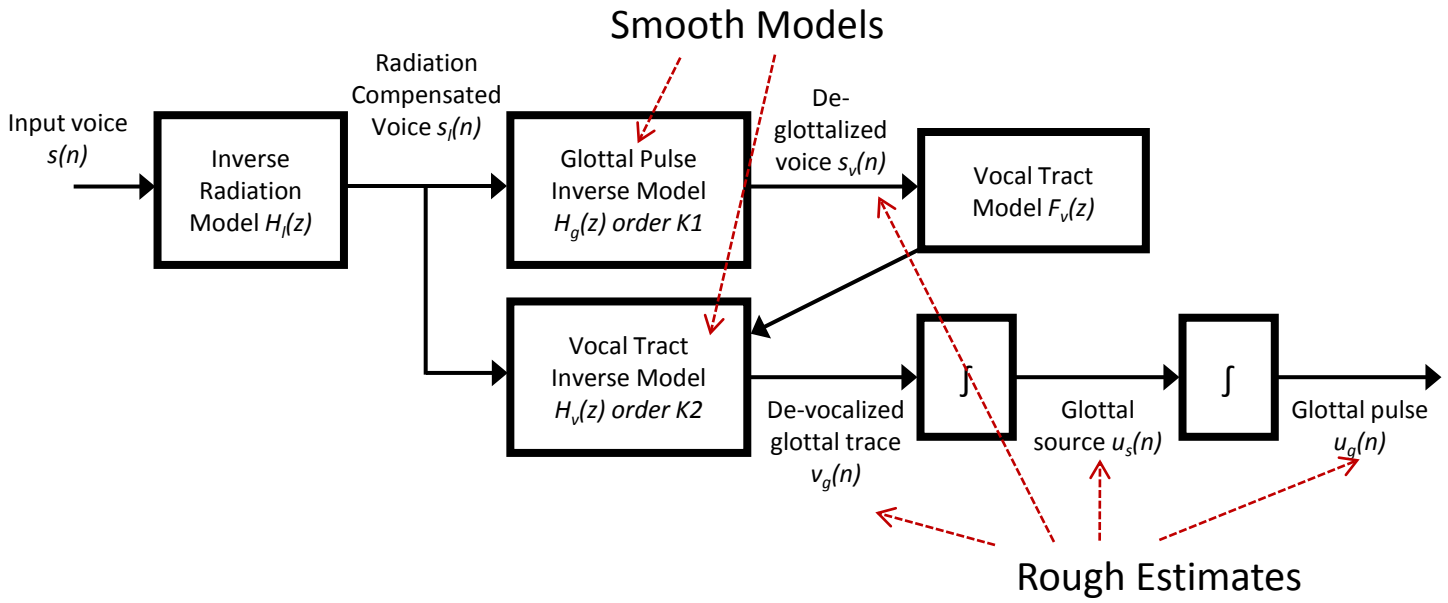
The iterative inverse filtering and deconvolution:

- Estimates the low-order (K_1) behaviour of the glottal correlate
- Removes this low-order behaviour from voice to estimate the vocal tract impulse response
- Models the vocal tract transfer function with accuracy large enough (K_2)
- Removes this high-order behaviour from voice to estimate the glottal residual (derivative of the glottal source)
- Re-builds the glottal source and flow applying successive integrations

Inverse Filters used in the study



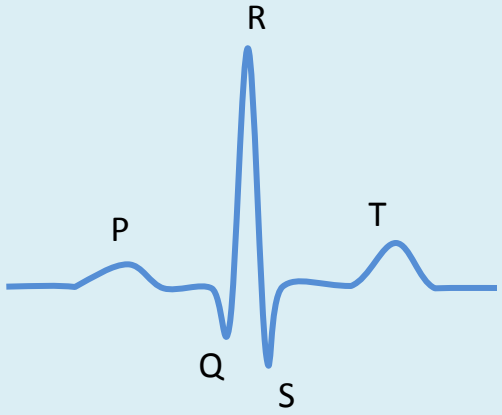
Modeling Process



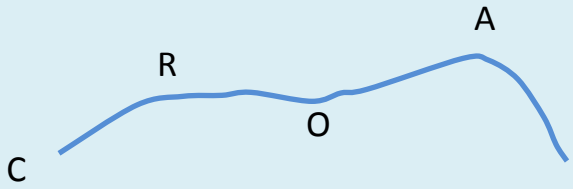
- The methodology is based on the use of ideal all-pole filters which model the spectral envelopes of the glottal pulse and the vocal tract impulse responses as smooth functionals to a K1 and K2 order extent
- Their residuals produce rough estimates of the counterpart signals (vocal→glottal and glottal→vocal) from where second order effects may be inferred: over-ringing, ripples, etc: these can be assigned a meaning

The Glottal Source

- It is a biometrical signature which can be compared in semantic power with the ElectroCardioGram
- Easy to obtain, store and match
- Once parameterized it may be used in Medical Studies (Organic + Neurologic Pathology Monitoring), Forensics (Police Voice Line-Up Systems), Customer Research Management (Telephone Call Centers, Customer Oriented Services, etc.)

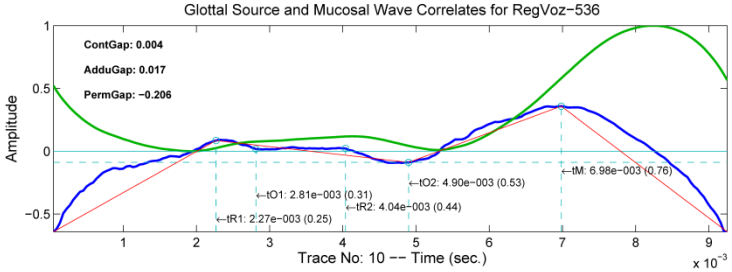


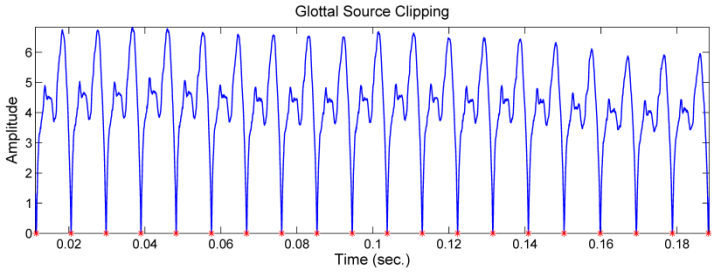
ElectroCardioGram (1/s)

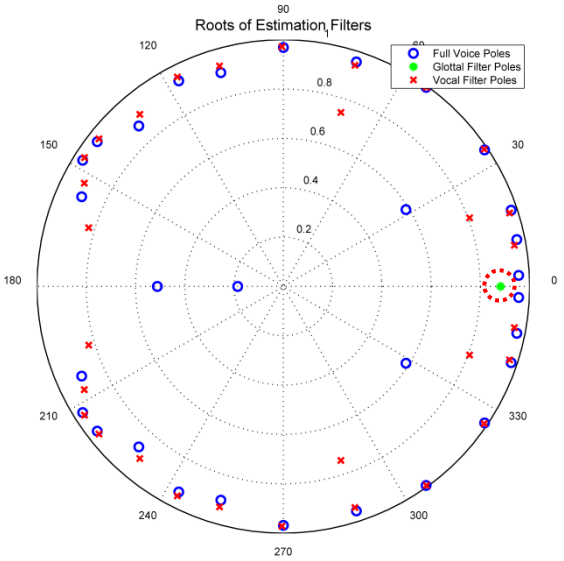


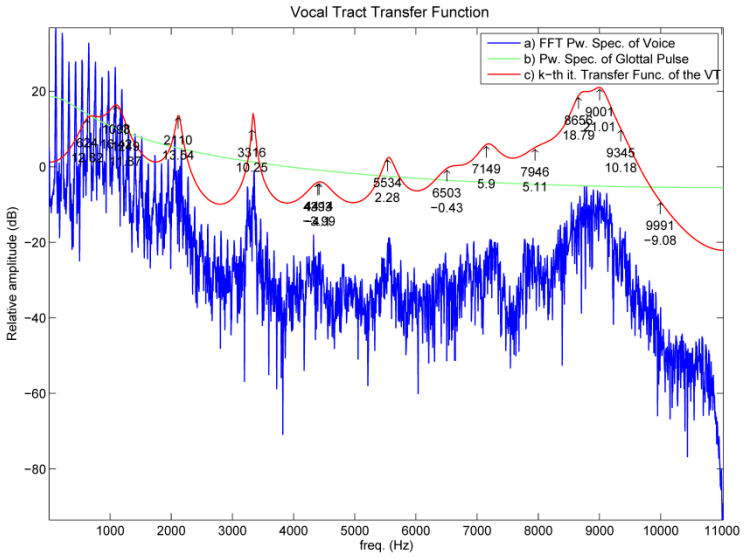
AudioGlottoGram (100-200/s)

Prototype Male Speaker (1)



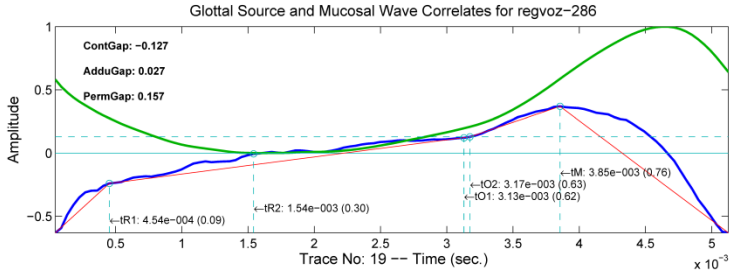


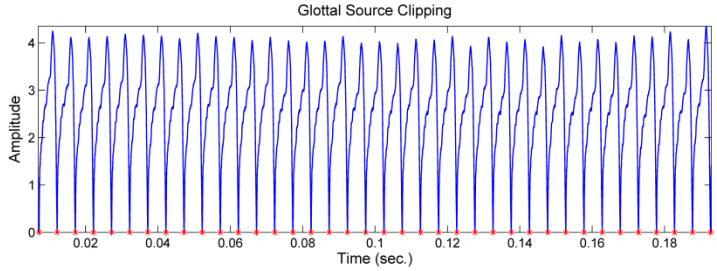


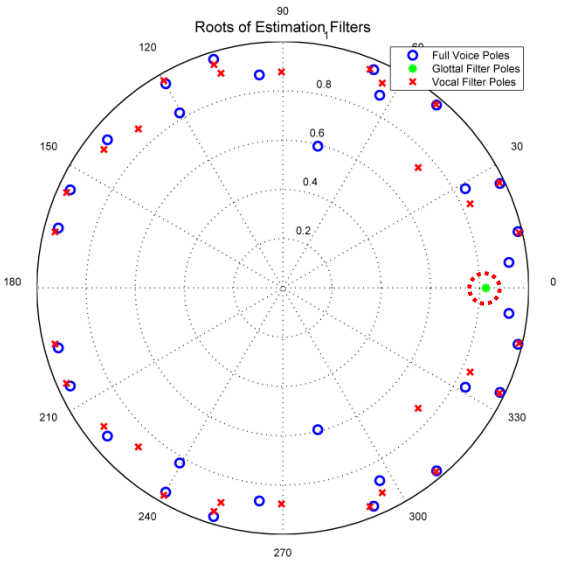


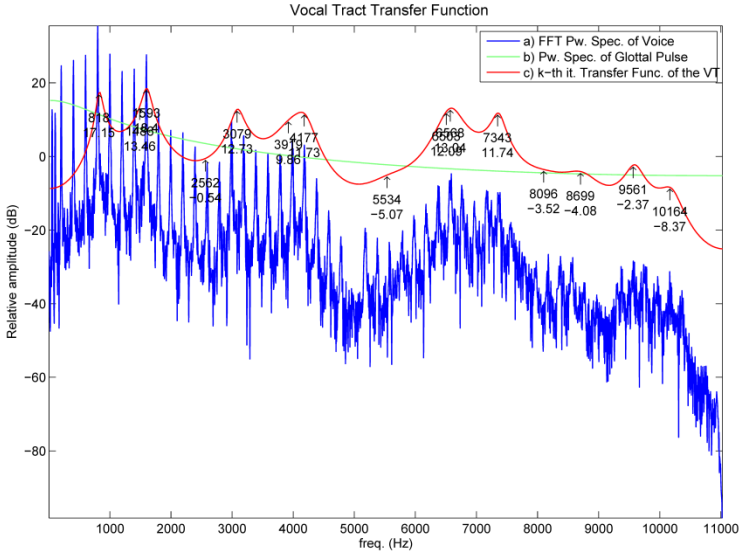
This example shows an order-1 estimation with good profiling

Prototype Female Speaker (1)



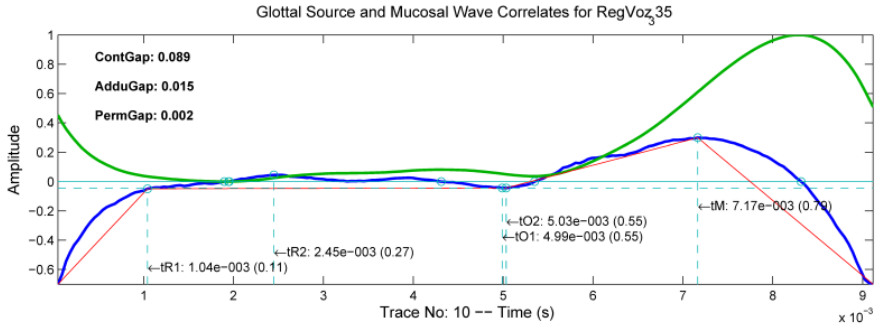


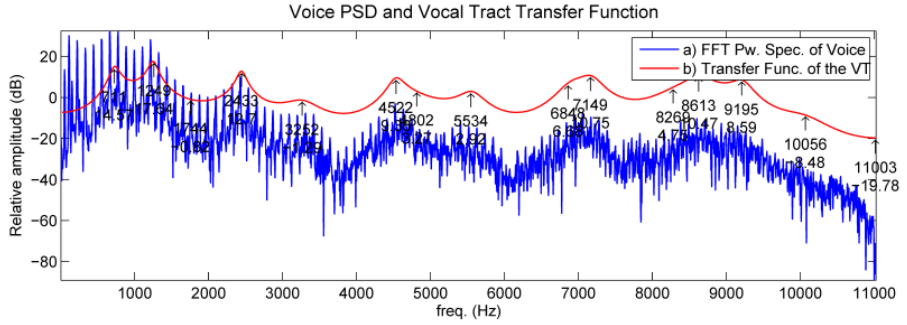


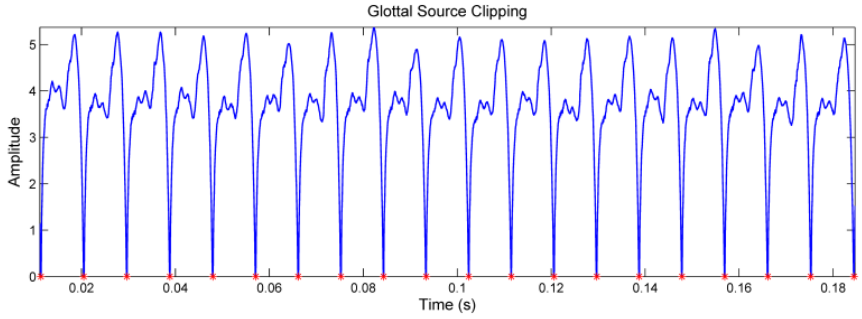


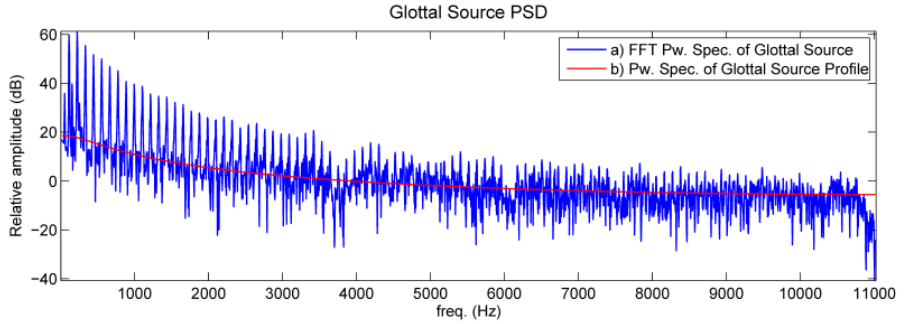
This example shows an order-1 estimation with good profiling

First Inversion Validation Results





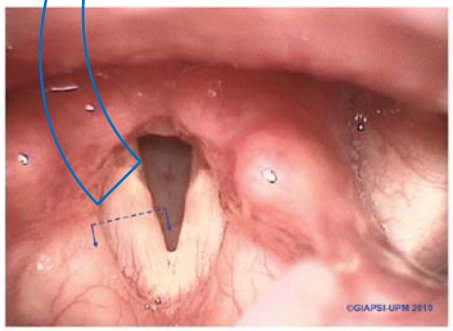


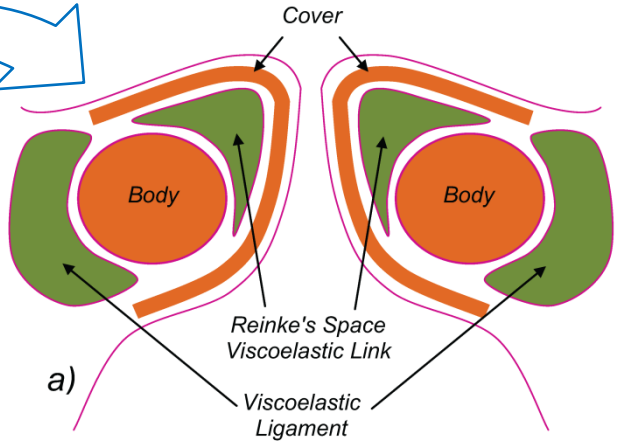


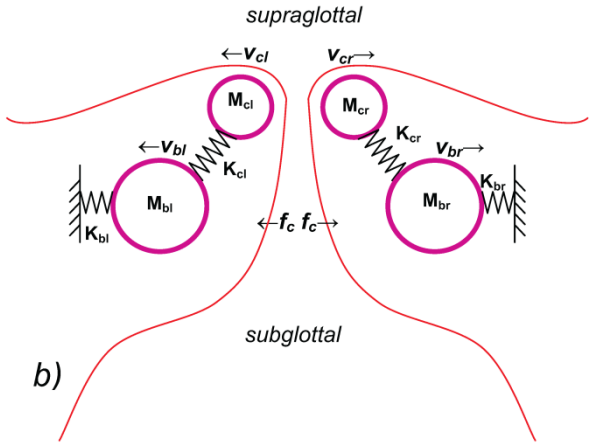
Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

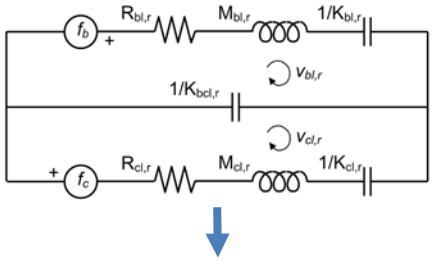
Second Problem Inversion

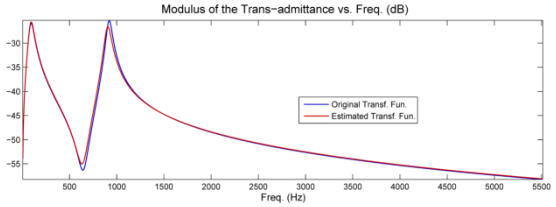


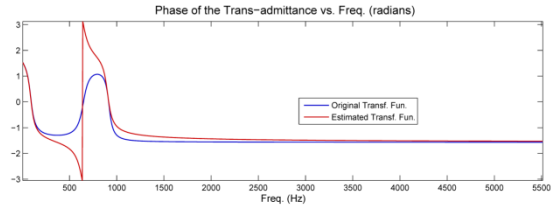




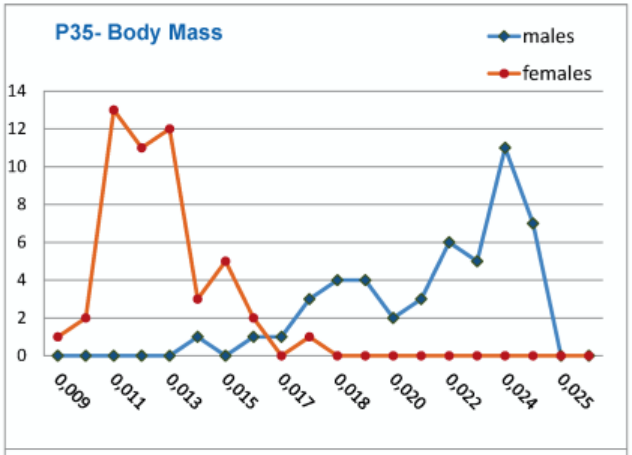
David A. Berry, Modal and nonmodal phonation, J. Phonetics, (29) 431-450, 2001

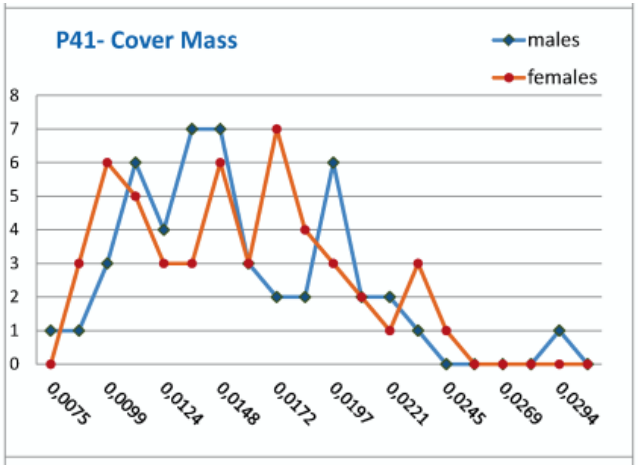


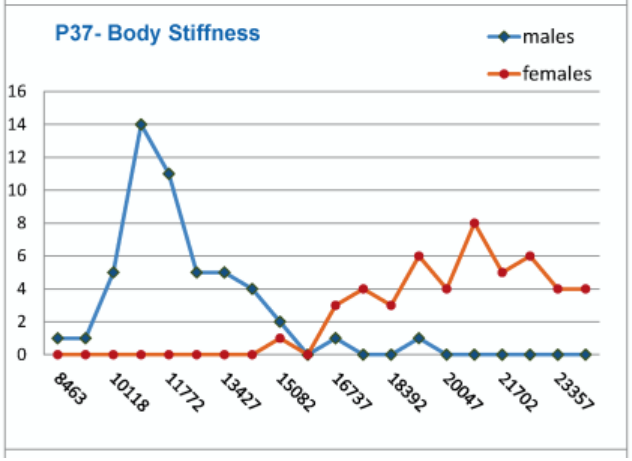


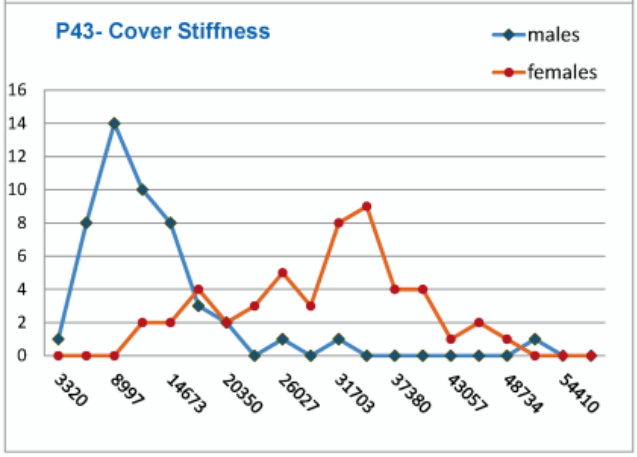


Second Inversion Validation Results









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)

Why do we accept these distributions as valid?

- Because they show a clear gender-dependent distinction:
  - Body mass is larger in males (almost twice) than in females
  - Body stiffness is larger in females than in males (almost twice)
  - In the average they are reciprocally distributed: subjects with larger mass have smaller stiffness and vice-versa
  - This last observation need not be true in pathological cases
  - Cover stiffness is also gender-dependent, but this is not true for cover mass, which apparently is not gender-differentiated
- Because distributions are modal
  - Although they are not apparently normally distributed
  - Most of them show skewness
  - Male distributions are leptokurtic, females tend to platikurtic
- Because mass estimations are plausible:
  - An average vocal fold may be in the range 80-150 mg
  - Mass estimates are a fraction of this total mass ranging from ¼ to ½ of the total
  - This is compliant with fold vibration rather than with cord vibration dynamics

And because it works: Clinical Voice Quality Assessment

Study Case:

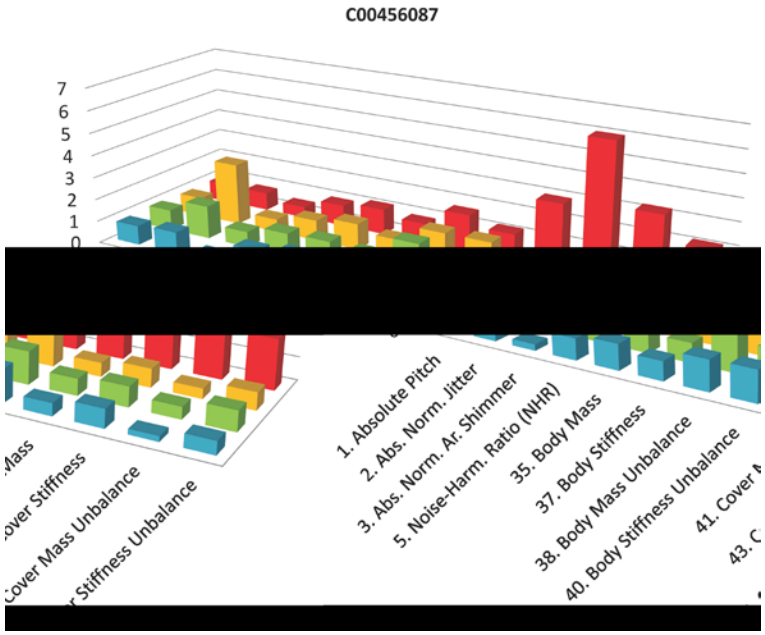
Female patient 65 years-old

Post-Thyroidectomic Vocal Fold Recurrent Paralysis (pTVFRP)

Treated by infiltration of fat from the patient in the vocal folds

The patient's voice was examined during almost a year (2011):

Once before the intervention (pre: March) and three times after the intervention (post1: May; post2: September; and post3: November).



Parameter	Pre	Post1	Post2	Post3
2-Jitter (%)	2.8	5.4	0.6	0.6
3-Shimmer (%)	10.5	3.3	1.5	1.0
38-Body M. Unb. (%)	4	21	<1	<1
40-Body S. Unb. (%)	10	30	1	1
41-Cover M. (mg)	26	8	8	6
43-Cover S. ($g.s^{-2}$)	91,746	24,228	14,175	11,808
44-Cover M. Unb. (%)	47	14	2	1
46-Cover S. Unb. (%)	43	26	3	3

And gives a lot of side information:

Why Jitter and Body Unbalances suffer an increment after intervention and other features do not?

Summary

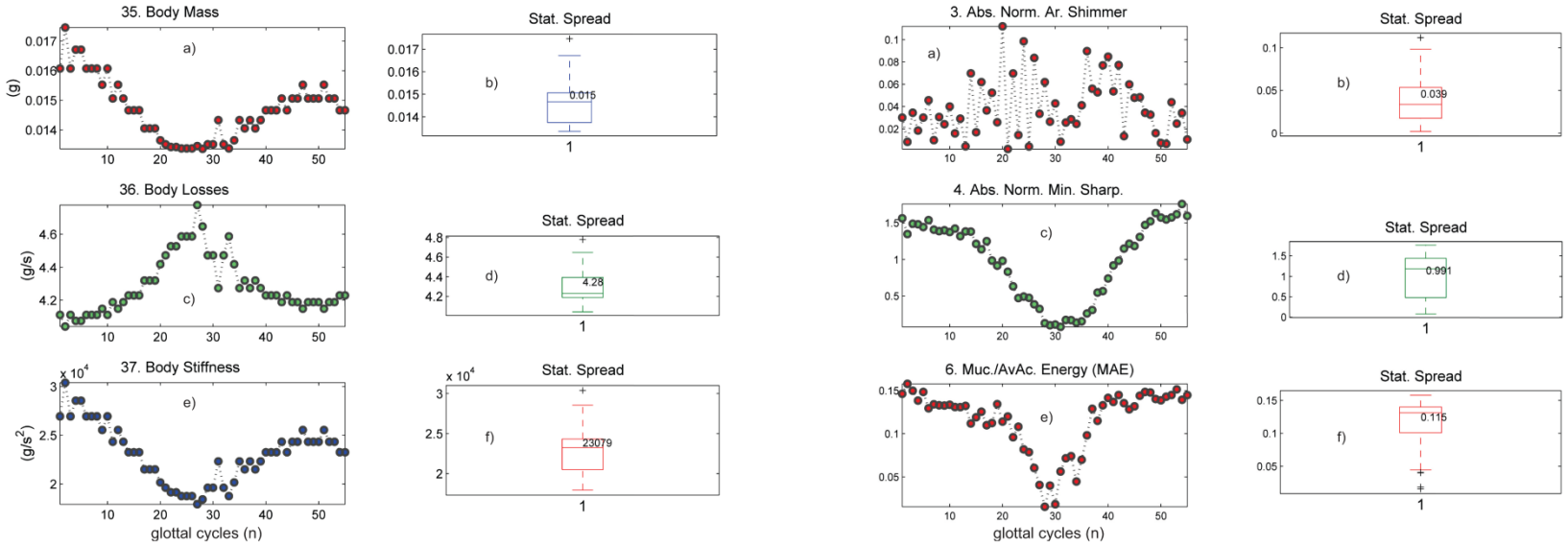
- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Third Model Inversion

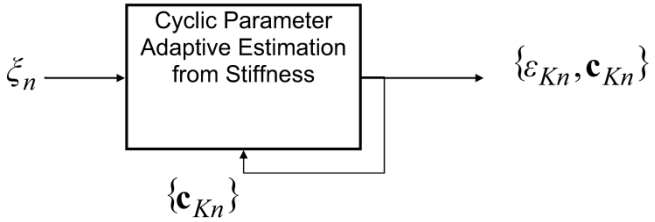
Some parameters show clear cyclicality

Q: How may we measure cyclicality?

A: How about adaptive AR modelling?



Third Inversion Process: Maths



AR Model Hypothesis:

$$\xi_n = \sum_{i=1}^{K} a_i \xi_{n-i} + \varepsilon_n$$

Adaptive Model Estimation:

$$\{\varepsilon_{Kn}, \mathbf{c}_{Kn}\} = \Phi_{Kn}\{\xi_n, W_K, \beta\}$$

Parameter Disclosing:

$$\mathbf{a}_{kn} = \mathbf{a}_{k-1n} - c_{kn}\tilde{\mathbf{a}}_{k-1n}$$

Behavior in the Freq. Domain:

$$H(z) = \frac{1}{1 - \sum_{1}^{K} a_i z^{-i}} = \prod_{i=1}^{K} \frac{z}{z - z_i}$$

Properties

Frequency estimate:

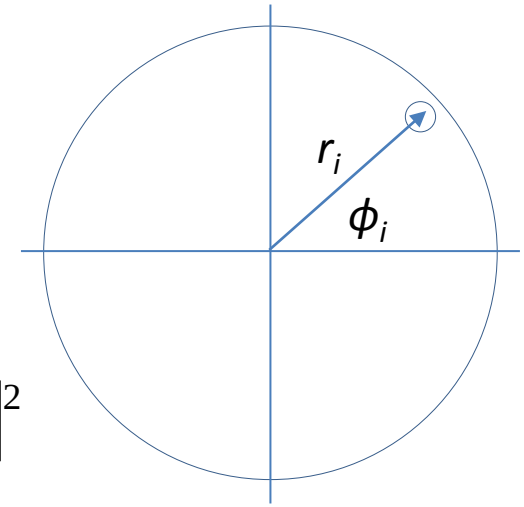
$$f_{ti} = \frac{\varphi_i}{2\pi} f_s;$$

Amplitude estimate:

$$\rho_{ti} = \frac{1}{1 - r_i};$$

Amplitude relevance:

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



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

$$c_1 = \frac{a_1 - a_2 a_3}{1 + a_2 - a_1 a_3 - a_3^2};$$

$$c_2 = \frac{a_2 - a_1 a_3}{1 - a_3^2};$$

$$c_3 = a_3$$

Why do this model detect cyclicality?

- It may be shown that if the moduli of the poles $r_i \rightarrow 1$ (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, althouth they do not share the same properties as c_1

3rd Model Inversion Validation Results: 3-pole estimatin error

$$z_1 = r_1; \quad z_2 = r_2 e^{j\varphi_2}; \quad z_3 = r_3 e^{j\varphi_3}$$

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

$$r_1 \in \Re; \quad r_2 = r_3 \in \Re; \quad r_2, r_3 > 0$$

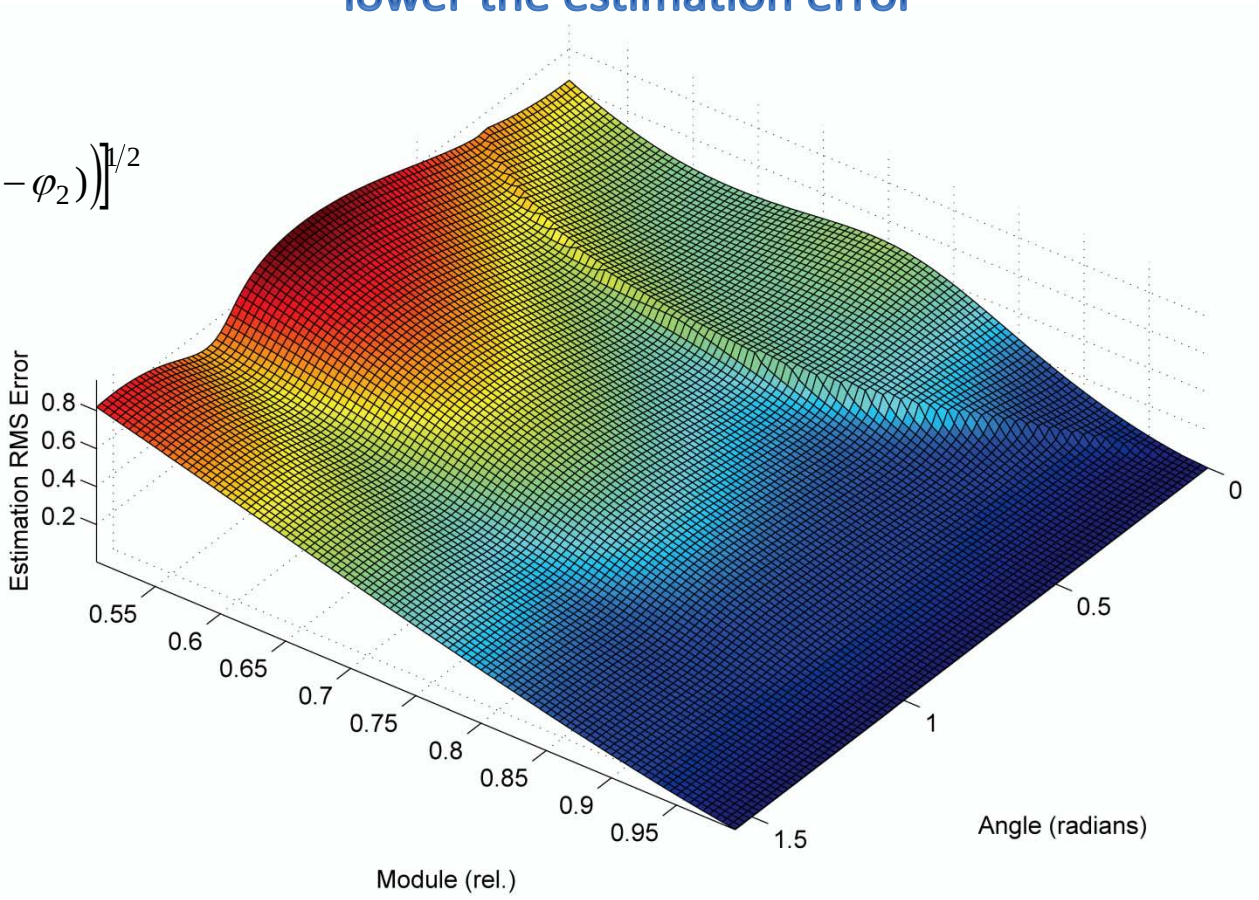
$$0 \le \varphi_2 \le \pi; \quad \varphi_3 = -\varphi_2$$

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

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

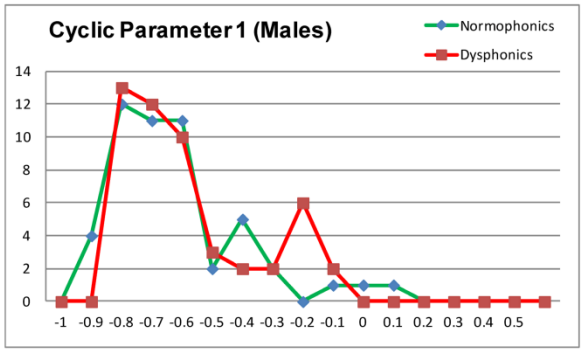
$$\varepsilon_r = \frac{\|\varepsilon_z\|}{\|\zeta\|}$$

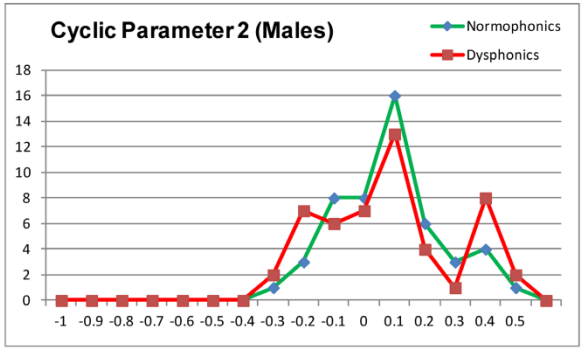
The good news is that the closer the zeroes to the unity circle the lower the estimation error

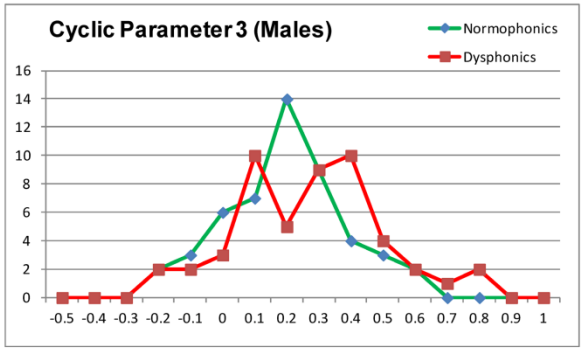


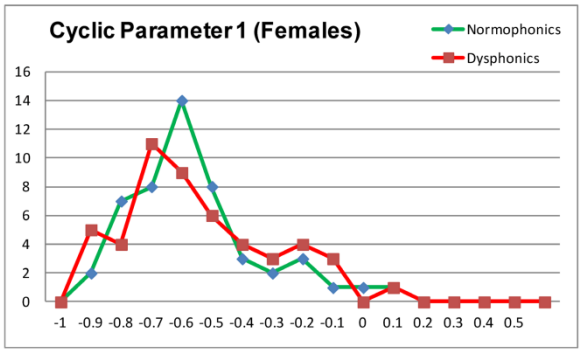
Normalizing for a large population

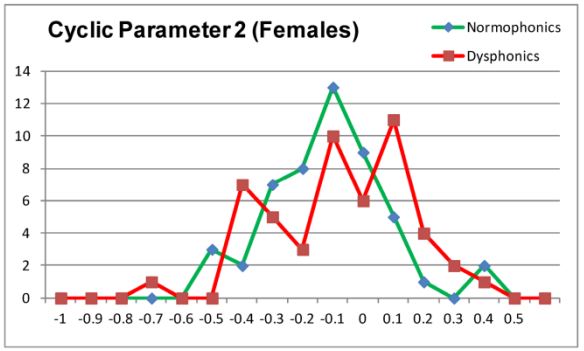
Normophonic
50 Males
50 Females
200 ms of vowel /a/

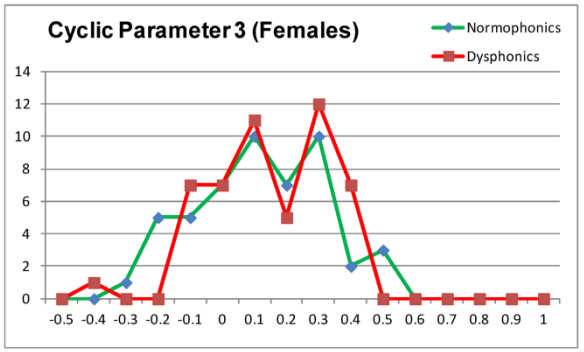












Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Statistical Analysis (1st level): Dist. Quartiles

- Quartile values show that distributions are balanced
- Parametric normalization of these distributions could be possible
- Clustering could be carried out based on these distributions

Three quartiles of the male distribution cyclicality coefficients

Parameters	c_1	c_2	c_3
1st Quartile Normophonics	-0,8126	-0,0882	0,0298
2nd Quartile Normophonics	**-0,7226**	**0,0426**	**0,1532**
3rd Quartile Normophonics	-0,6170	0,1181	0,2777
1st Quartile Dysphonics	-0,8076	-0,1272	-0,0306
2nd Quartile Dysphonics	**-0,7077**	**0,0272**	**0,1436**
3rd Quartile Dysphonics	-0,5256	0,1759	0,2699

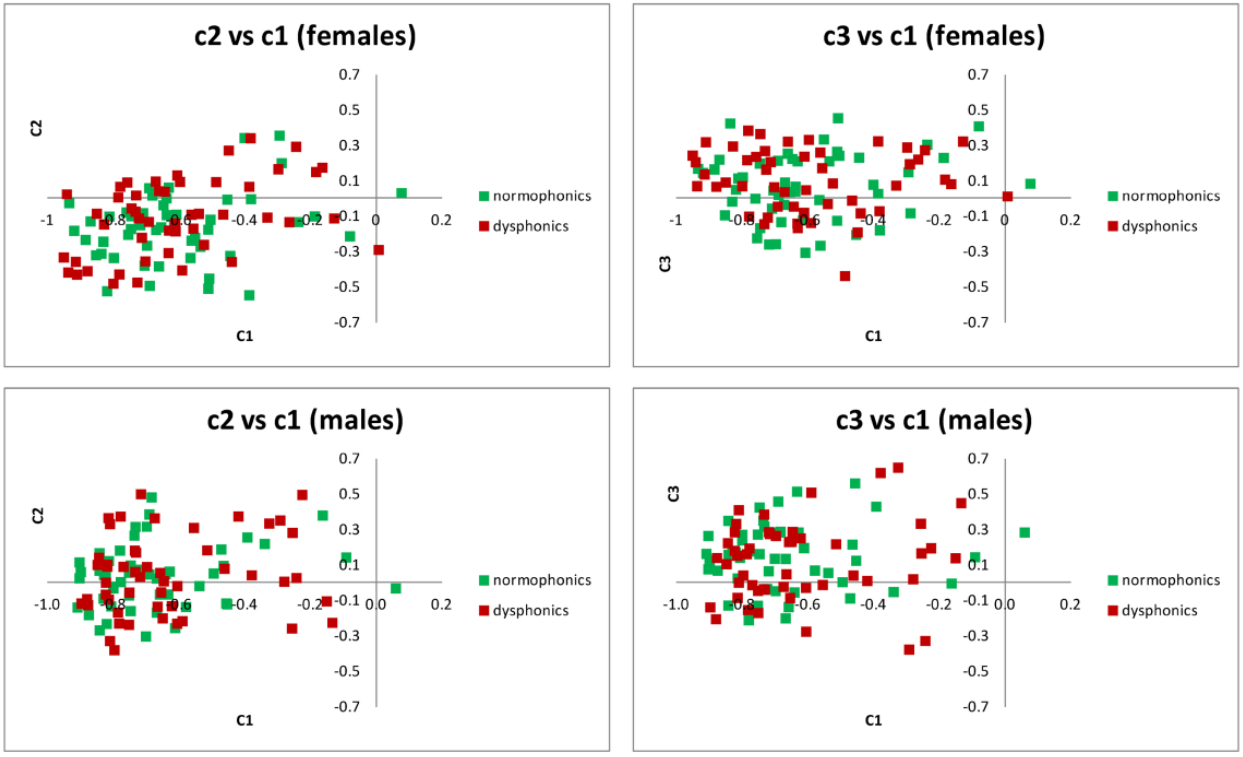
Three quartiles of the female distribution cyclicality coefficients

Parameters	c_1	c_2	c_3
1st Quartile Normophonics	-0,7468	-0,2738	-0,0802
2nd Quartile Normophonics	**-0,6521**	**-0,1654**	**0,0709**
3rd Quartile Normophonics	-0,5099	-0,0519	0,2109
1st Quartile Dysphonics	-0,7553	-0,3067	-0,0455
2nd Quartile Dysphonics	**-0,6326**	**-0,1131**	**0,0862**
3rd Quartile Dysphonics	-0,4540	0,0579	0,2391

- 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?

Statistical Analysis (1st level): Scatter Plots

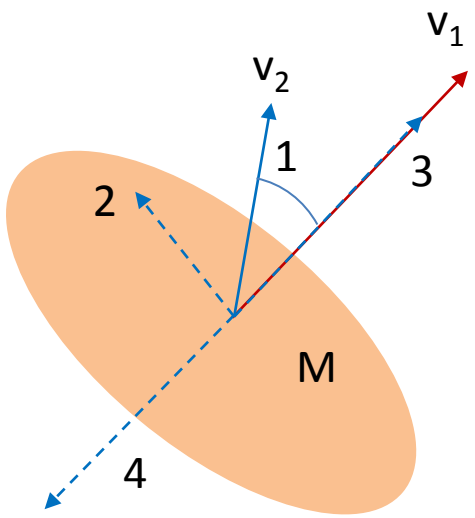
- The lack of a tendency (regression line) indicates that c1, c2 and c3 are statistically independent (under 2nd order stat.)
- Is this good or bad?



How to read correlation (Pearson's)

- Let's asume that v_1 and v_2 are observation vectors on the same set of parameters
- The estimation of correlation will be drawn using Pearson's coeff.:

$$\rho = \frac{\sum_i v_{1i} v_{2i}}{\left[\sum_i v_{1i}^2 \sum_i v_{2i}^2\right]^{1/2}} = \frac{\langle v_1 \cdot v_2 \rangle}{\|v_1\| \|v_2\|} = \cos\beta$$



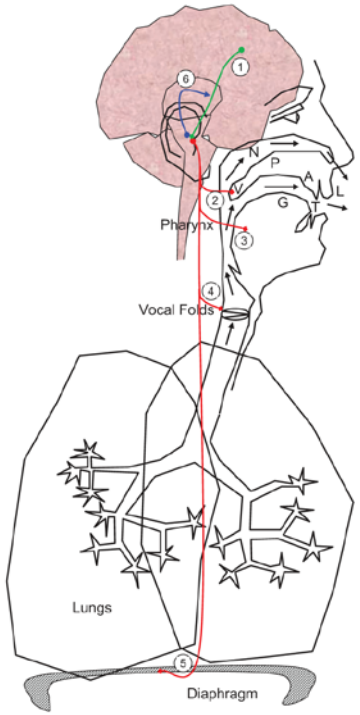
Case	Angle	PC	Meaning
1	0<β<π/4	√2/2<ρ<1	High corr.
2	β=π/2	ρ=0	No corr.
3	β=0	ρ=1	Same Highest
4	β=-π/2	ρ=-1	Contrary Highest

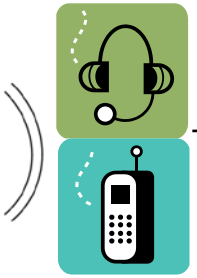
Zone	Angle	PC	Meaning
A	0<β≤π/4	0.707≤ρ<1	High corr.
B	π/4<β<3π/4	|ρ|<0.707	Low corr.
C	3π/4<β≤π	-1≤ρ<-0.707	High. cont. corr.

What is a confidence test?

- It is a methodological procedure to ensure that the results obtained in detection tasks (binary decision) fulfill certain guarantees under statistical foundations
- It is a method to estimate the probability of the errors produced in a detection task
- A detection task is a procedure by which having a given estimate of observations or features x, it must be inferred if based on these observations a specific hypothesis is fulfilled (H_1: H_D) or not (H_0: H_N)
- Based on the a priori knowedge on how the observations distribute over the two hypotheses $f_D(x)$ and $f_N(x)$
- In what follows it will be assumed that the distributions $f_D(x)$ and $f_N(x)$ are Gaussian with means μ_D and μ_N, and standard deviations σ_D and σ_N

Possible outcomes and errors





3 Inversion Steps leading to Biomechanics

A priori Knowledge: μ_D, μ_N, σ_D, σ_N

Body & Cover Stiffness

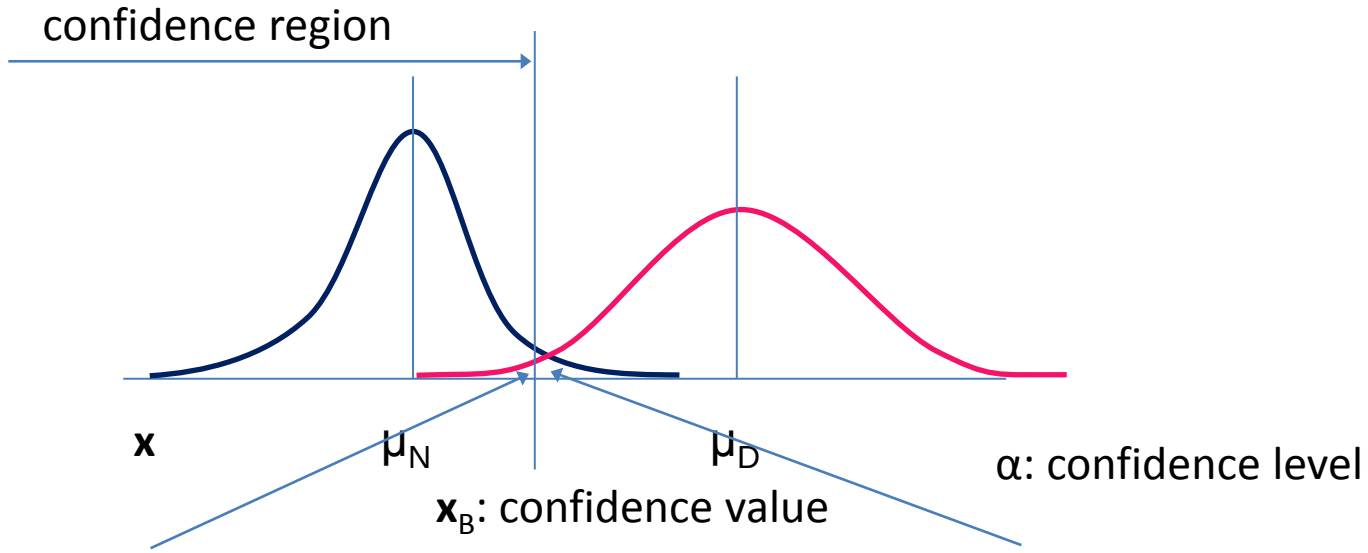
Detector of Dysphonia: YES or NOT

Confidence Test: Up to which level the Detector IS RELIABLE?

Gain in Knowledge

Real/Detected	Normophonic	Dysphonic
Normophonic	Correct Detection	False Positive (Type I Error)
Dysphonic	False Negative (Type II Error)	Correct Detection

Now Statistics comes to help



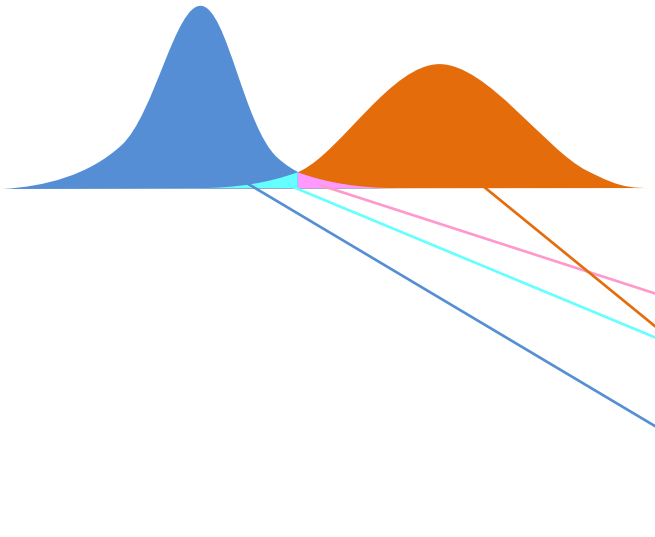
β - Area under red curve:
Probability of accepting H_N
for dysphonic subjects
Type II error: False Negatives

α - Area under blue curve:
Probability of rejecting H_N
for normophonic subjects
Type I error: False Positives

THE SENTENCE:
Evidence supports the rejection of the null hypothesis H_0 (in this case H_N) under a confidence level of α (generally given as a %) at a value (confidence region) given by x_B

Some Maths to estimate factors

Real/Detected	Normophonic	Dysphonic
Normophonic	True Negative: 1-α	False Positive: α
Dysphonic	False Negative: β	True Positive: 1-β



$$f_N(x) = \frac{1}{\sigma_N\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_N}{\sigma_N}\right)^2}; \quad f_D(x) = \frac{1}{\sigma_D\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_D}{\sigma_D}\right)^2}$$

$$F_N(x) = p_N(\zeta \le x) = \int_{-\infty}^{x} f_N(\zeta)d\zeta$$

$$F_D(x) = p_D(\zeta \le x) = \int_{-\infty}^{x} f_D(\zeta)d\zeta$$

$$FP: \quad \alpha = 1 - F_N(x_\alpha)$$

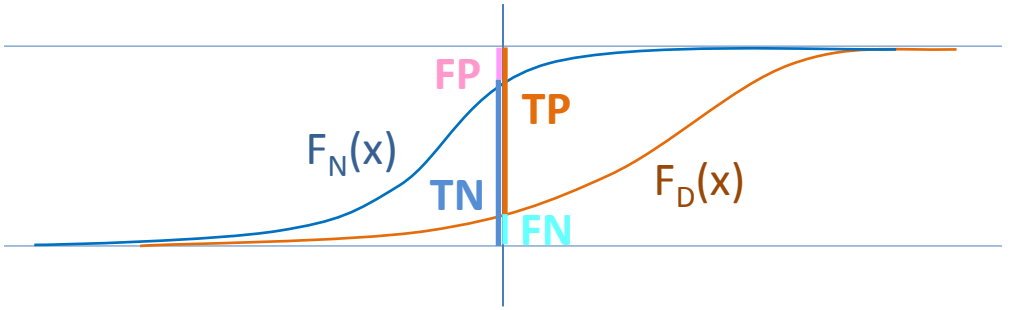
$$FN: \quad \beta = F_D(x_\alpha)$$

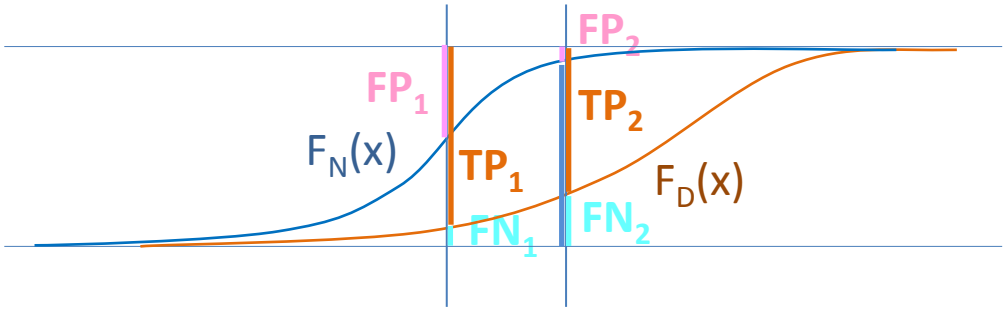
$$TP: \quad 1-\beta = 1 - F_D(x_\alpha)$$

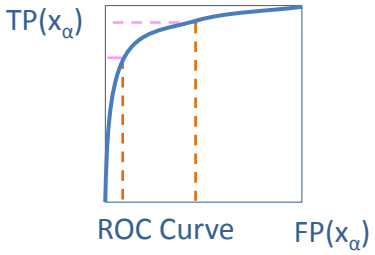
$$TN: \quad 1-\alpha = F_N(x_\alpha)$$

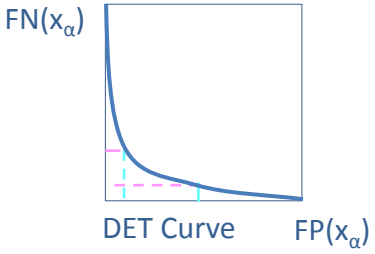
Different representations

Tippett's Plots: give detection probabilities in terms of Cumulative Probability Functions after **Leonard Henry Caleb Tippett** (8 May 1902 - 9 November 1985)

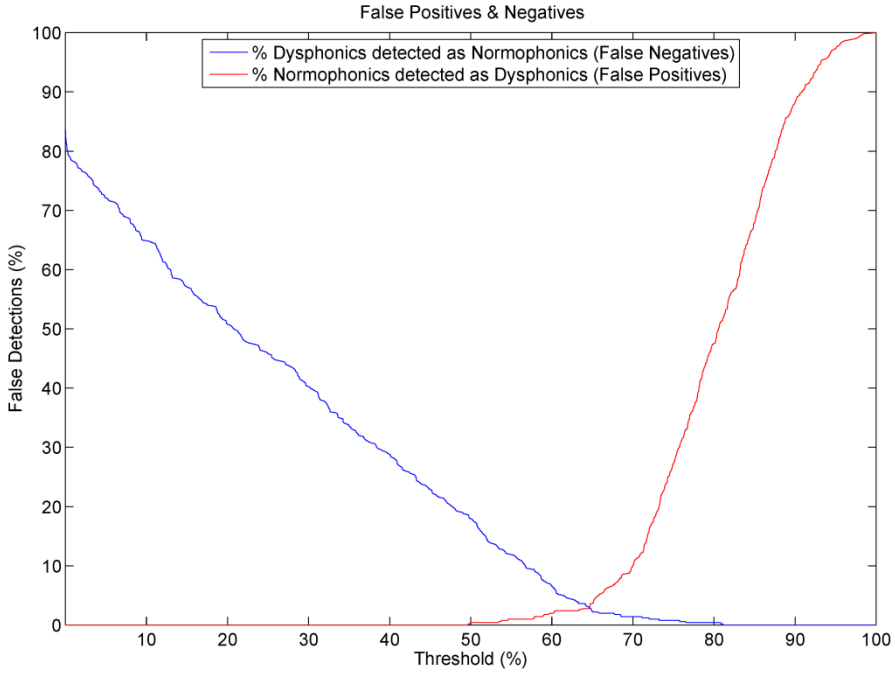


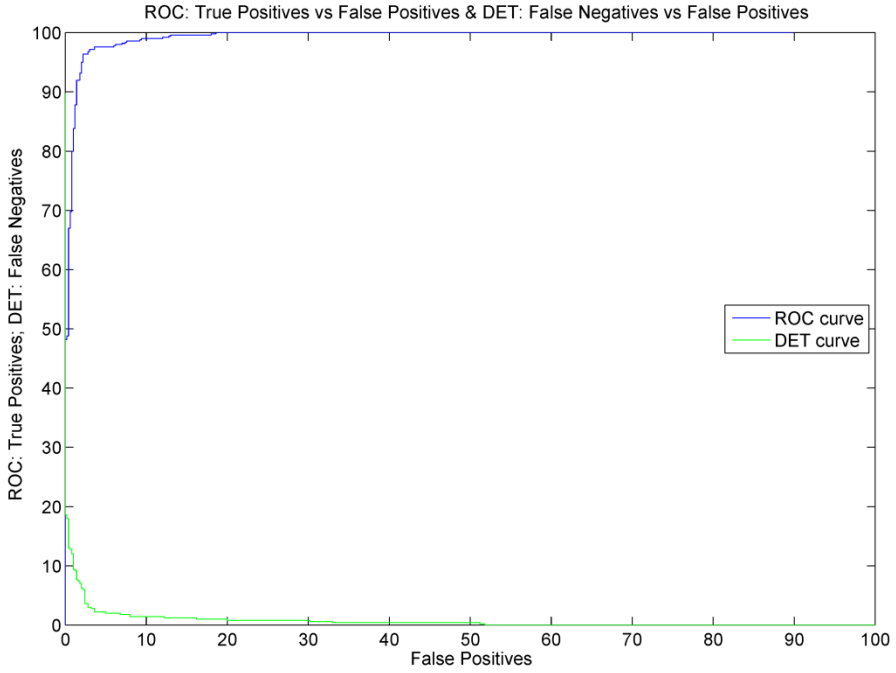






Some real results: male subjects

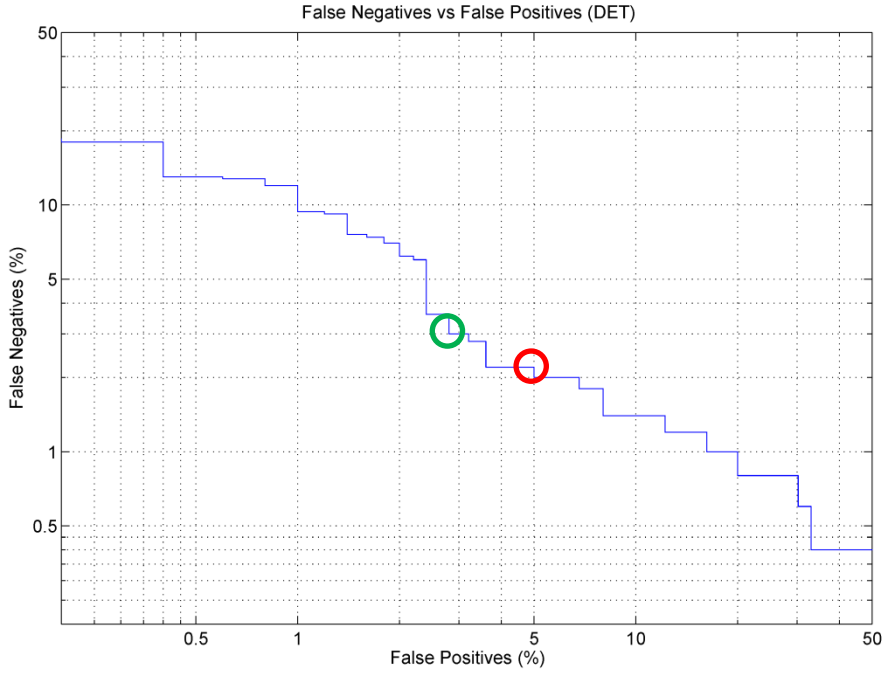


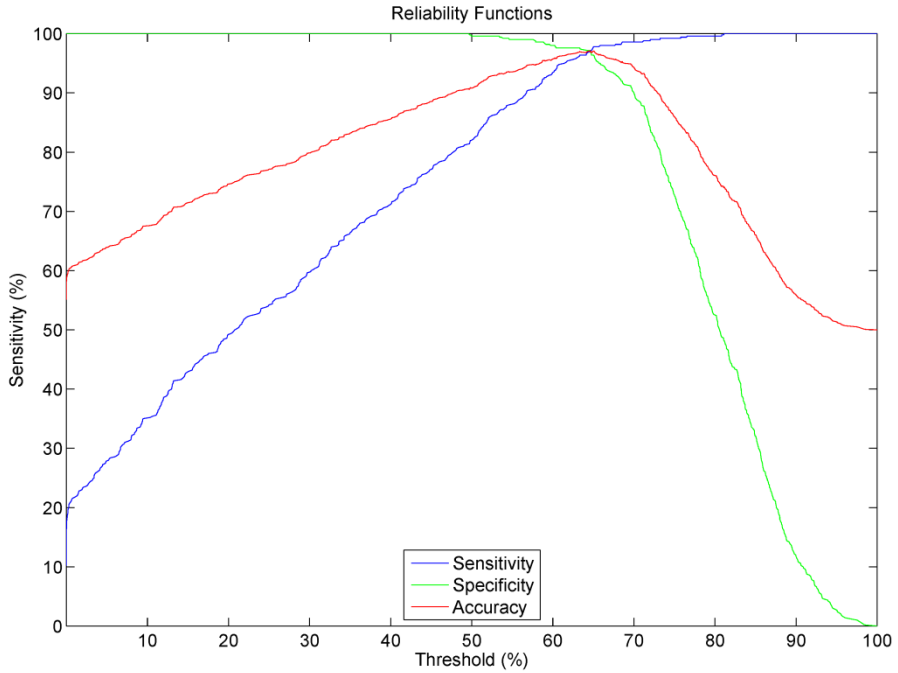


Merit factors: male subjects

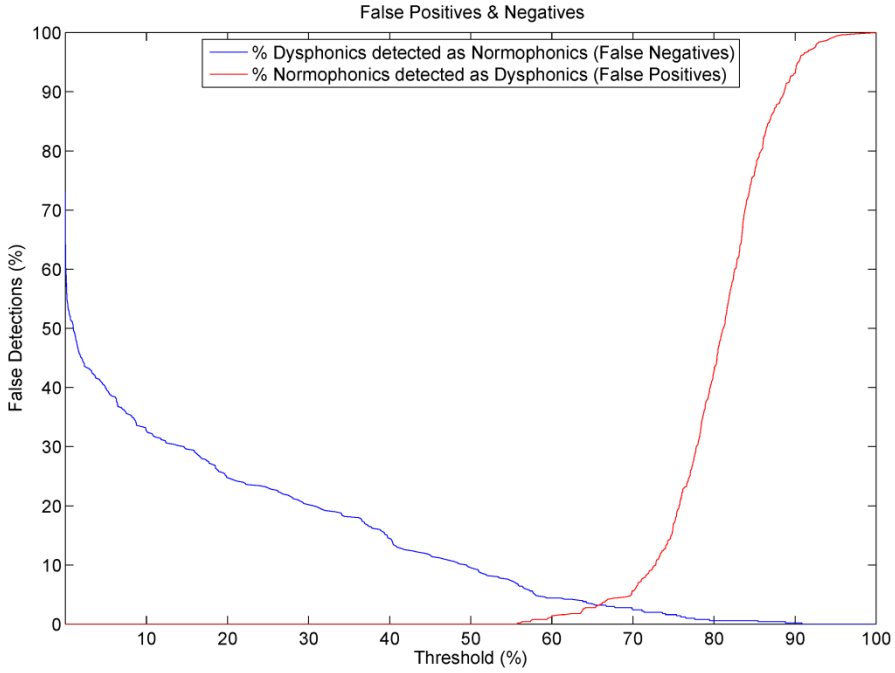
Q: How good our detector is to catch true positives?
A: Sensitivity=TP/(TP+FN)
Q: How good our detector is to catch true negatives?
A: Selectivity=TN/(TN+FP)
Q: How good is the detection of trues?
A: Accuracy=(TP+TN)/(TP+FN+TN+FP)

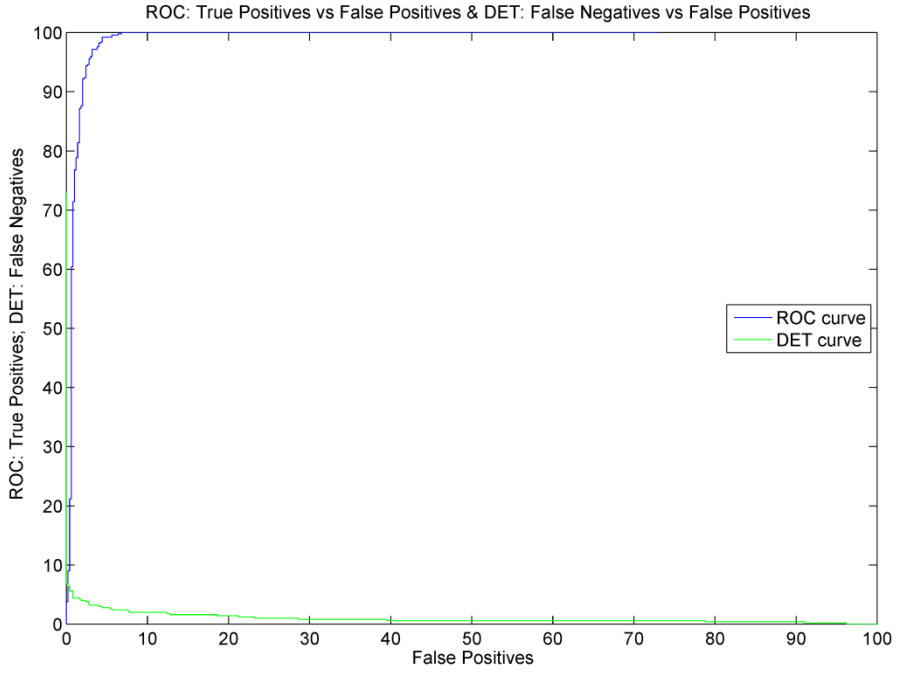
DET: Deteccion-Error-Trade-off





Some real results: female subjects

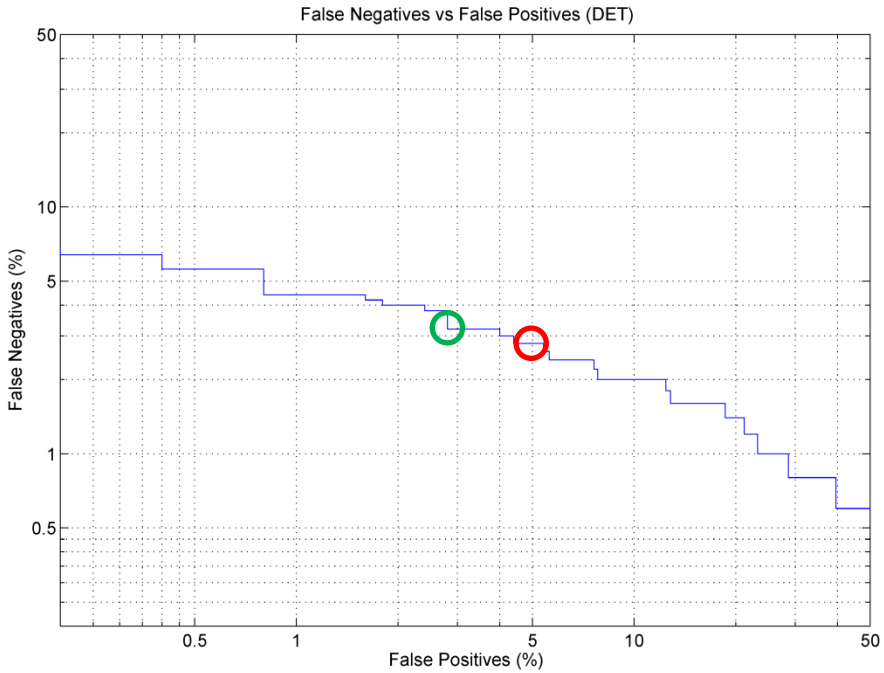


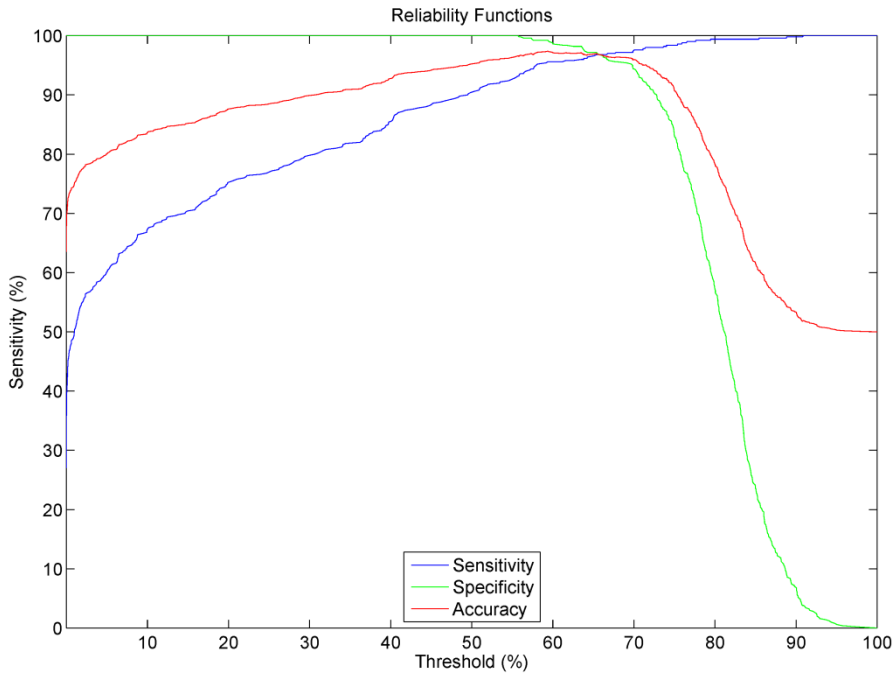


Merit factors: female subjects

Another merit factor is the point in the DET curve closest to the Equal Error Rate where FP=FN

The interval between the confidence level of α=5% and the EER is also a merit factor





Weaknesses and Solutions

Weaknesses:

- In many occasions $f_N(x)$ is well known, or may be accurately modelled, but $f_D(x)$ cannot, or is very difficult to model
- In most occasions neither $f_N(x)$ nor $f_D(x)$ may be modelled by Gaussian Distributions

Solutions:

- If $f_D(x)$ is not well-known a detection process may be implemented based in odds: $p(x|H_D)=1-p(x|H_N)$
- If $f_N(x)$ or $f_D(x)$ cannot be modelled as single Gaussian Distributions, mixtures of different Gaussians may be used to approximate the real distributions (in case that there is enough data availability)

Statistical Analysis (2nd level) dependence among Cyc. Par.

- Correlation (Pearson) may show if two each two coefficients are related under second order statistics. It may be considered that all the pairs of cyclicality coefficients c_i-c_j are not significantly correlated
- Confidence tests over a certain threshold (let's say 0.05) indicate plausibility in hypothesis testing: H0 is that distributions from Normal and Dysphonic are not generated by the same statistical models. As p-values are over 0.05 it means that H0 has to be disregarded: Normal and Dysphonic distributions may be generated by the same models under cyclicality.

Distribution Characteristics and Statistical dependence among c1, c2 and c3

	σ_{c1}	σ_{c2}	σ_{c3}	χ_{c1}	χ_{c2}	χ_{c3}	$\rho_{c2\text{-}c1}$	$\rho_{c3\text{-}c1}$	$\rho_{c3\text{-}c2}$	p_{c1NvsD}	p_{c2NvsD}	p_{c3NvsD}
Male Norm.	1,621	0,356	0,041	2,909	-0,025	-0,326	0,208	0,019	0,063	0.382	0.884	0.493
Male Dysph.	0,932	0,289	0,051	-0,371	-0,691	-0,056	0,153	0,110	0,096	---	---	---
Female Norm.	1,029	0,317	-0,096	1,172	0,650	-0,654	0,259	0,144	-0,155	0.862	0.368	0.357
Female Dysph.	0,618	-0,304	-0,537	-0,214	-0,033	0,024	0,294	-0,070	-0,067	---	---	---

Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Study Cases

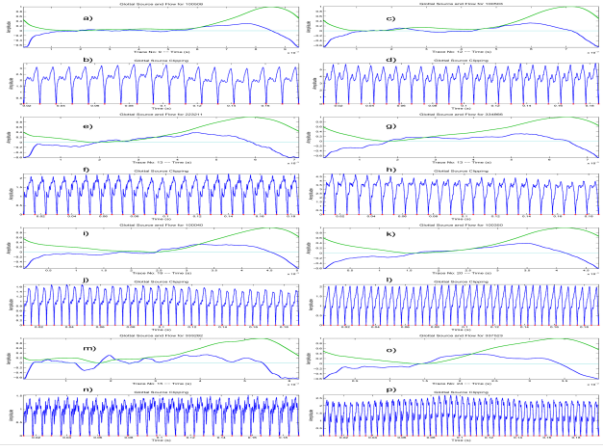
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Parkinson's Dis.	No	e-f
334866	Male	Parkinson's Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Parkinson's Dis.	No	m-n
337523	Female	Parkinson's Dis.	Yes	o-p

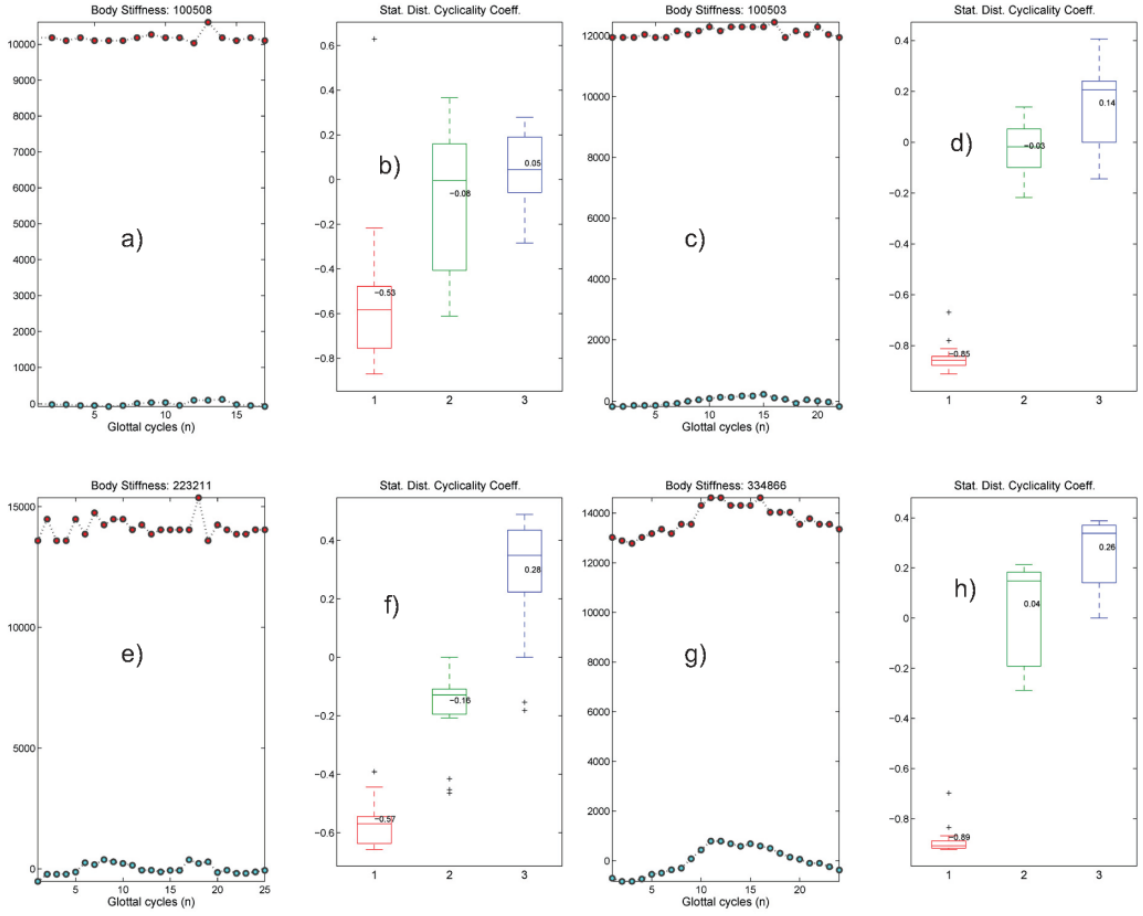
Some distortion and biomechanical parameter estimations. Conditions: NPNT-normphonic, no tremor; NPYT-normophonic, tremor; PDNT-Parkinson's Disease, no tremor; PDYT-Parkinson's Disease, tremor.

Case	Condition	Gender	Age	Grade	f_0 (Hz)	σ_{f0} (Hz)	Jitter (rel.)	Shimmer (rel.)	μ_{Kb} (g)	σ_{Kb} ($g.s^{-2}$)	μ_{Kc} (g)	σ_{Kc} ($g.s^{-2}$)
100508	NPNT	M	23	0	105.44	0.69	0.007	0.027	10,180	134	5,393	135
100040	NPNT	F	28	0	200.67	0.98	0.005	0.053	19,227	154	14,023	942
100503	NPYT	M	28	0	125.79	0.93	0.006	0.020	12,134	156	7,010	399
100350	NPYT	F	24	0	211.70	1.31	0.006	0.014	20,247	192	20,373	544
223211	PDNT	M	65	1	142.52	1.79	0.013	0.016	14,145	407	19,498	1,573
333282	PDNT	F	70	2	162.83	5.31	0.038	0.022	17,276	2,172	12,815	1,945
334866	PDYT	M	74	2	135.84	2.20	0.006	0.017	13,777	576	14,498	2,380
337523	PDYT	F	72	2	248.63	4.31	0.008	0.015	25,314	1.098	30,274	2,525

Phenomenological Studies (I)

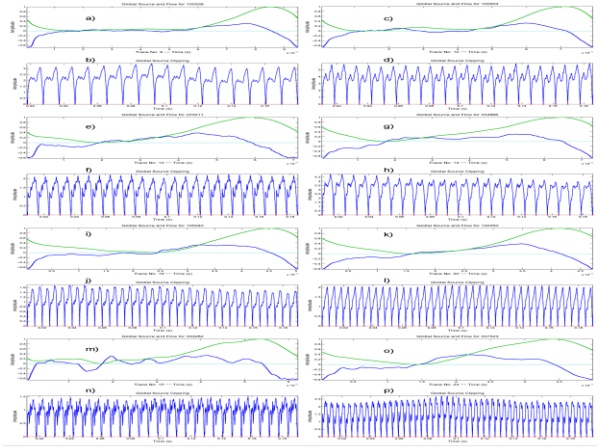
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

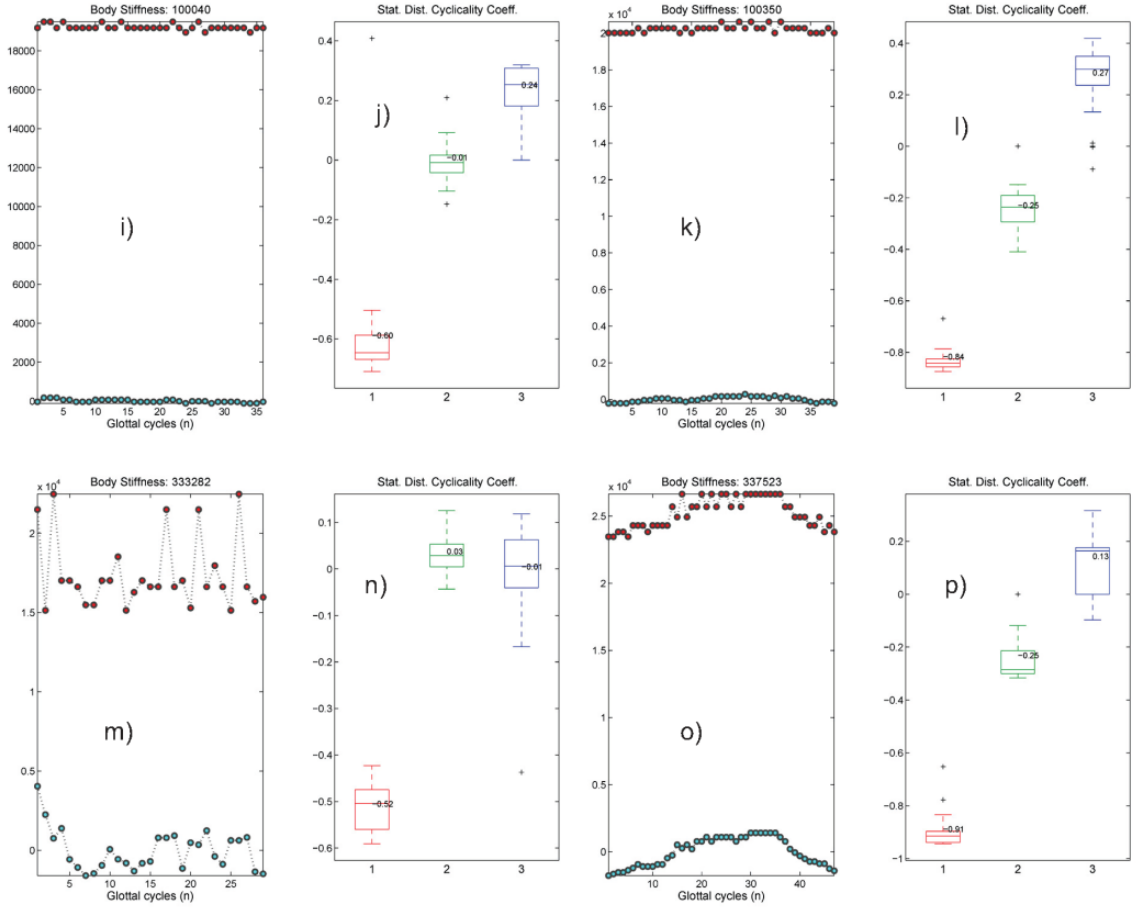




Phenomenological Studies (II)

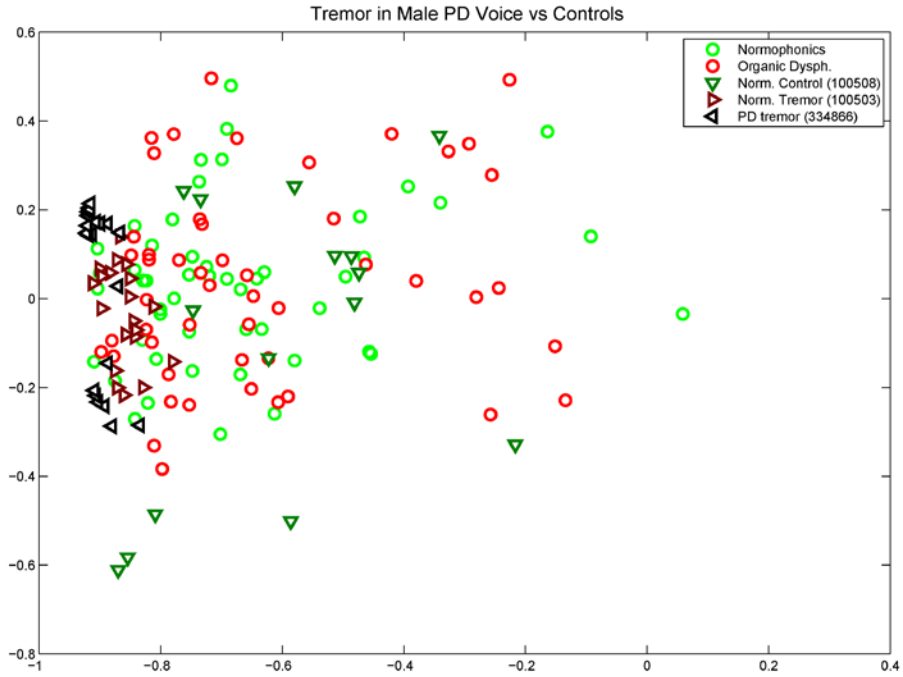
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

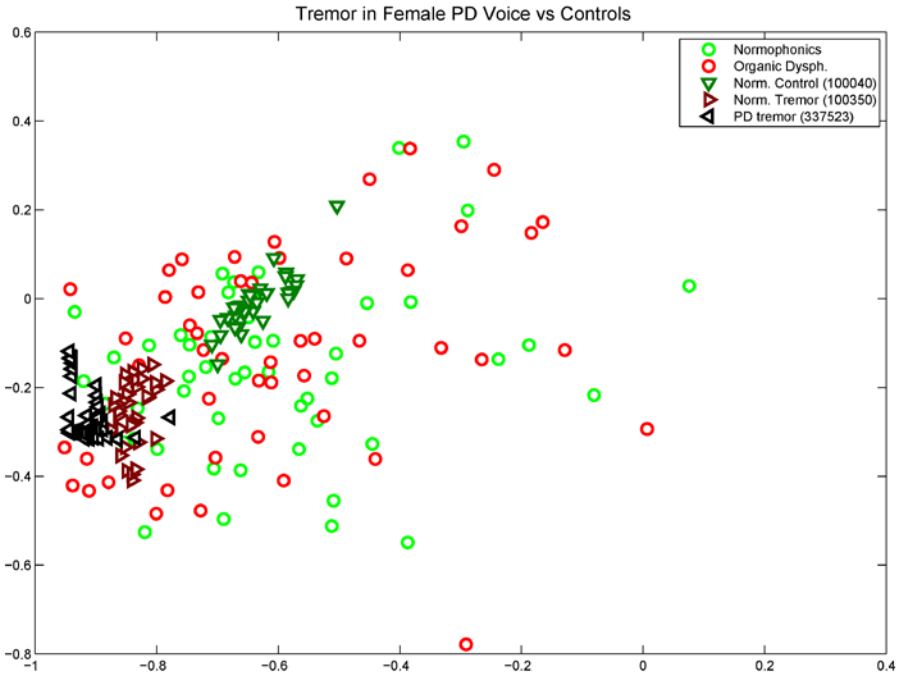




Comparative Studies

Intra- and inter-subject scatter plots (c_2vsc_1) for female control and tremor affected PD Patients





Statistical Classification Methods

Observation vectors are based on parameter averages of body and cover stiffness and first cyclicality parameter:

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

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

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

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

$$\mathbf{C_m} = \mathbf{X_{sm}^T X_{sm}}; \quad \mathbf{C_f} = \mathbf{X_{sf}^T X_{sf}}$$

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

$$\boldsymbol{\chi}_\mathbf{m} = \mathbf{E}\langle \mathbf{X_{sm}} \rangle_\mathbf{s}; \quad \boldsymbol{\chi}_\mathbf{f} = \mathbf{E}\langle \mathbf{X_{sf}} \rangle_\mathbf{s}$$

Statistical Classification Methods

Classification is based on conditional probabilities of an observation vector from a new subject (patient) of being produced by the model considered

$$\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m}) = \frac{1}{(2\pi)^{3/2}\left|\mathbf{C_m}\right|^{1/2}} \iiint_{(-\infty,\mathbf{x_q})} e^{-1/2(\zeta-\chi_\mathbf{m})^T \mathbf{C_m^{-1}}(\zeta-\chi_\mathbf{m})} \mathbf{d}\zeta$$

$$\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f}) = \frac{1}{(2\pi)^{3/2}\left|\mathbf{C_f}\right|^{1/2}} \iiint_{(-\infty,\mathbf{x_q})} e^{-1/2(\zeta-\chi_\mathbf{f})^T \mathbf{C_f^{-1}}(\zeta-\chi_\mathbf{f})} \mathbf{d}\zeta$$

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:

$$\lambda_{\mathrm{Nm}}(\mathbf{x_q}) = \log\frac{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})}{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})} = \log\left\{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})\right\} - \log\left\{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_m})\right\}$$

$$\lambda_{\mathrm{Nf}}(\mathbf{x_q}) = \log\frac{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})}{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})} = \log\left\{\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})\right\} - \log\left\{1-\Pr(\mathbf{x_q} \mid \mathbf{\Gamma_f})\right\}$$

Detection results

Cyclicality parameters and likelihood ratios										
Case	**f_t**	**η_t**	**c_1**	**σ_{c1}**	**c_2**	**σ_{c2}**	**c_3**	**σ_{c3}**	**λ_{c1}**	**λ_{T1}**
100508	17,55	0,0049	-0,6	0,19	-0,08	0,33	0,05	0,15	**0.48**	**-0,44**
100040	10,66	0,0034	-0,63	0,04	0	0,06	0,24	0,07	**-0.21**	**-0,03**
100503	5,59	0,009	-0,85	0,03	-0,03	0,1	0,14	0,16	**-1.53**	**-0,28**
100350	7,49	0,006	-0,84	0,02	-0,25	0,07	0,27	0,12	**-1.82**	**-0,45**
223211	10,78	0,0136	-0,57	0,05	-0,16	0,11	0,28	0,19	**0.70**	**-13,9**
333282	16,34	0,0597	-0,52	0,04	0,03	0,04	-0,01	0,11	**1.19**	**-11,4**
334866	5,39	0,0342	-0,89	0,02	0,04	0,19	0,26	0,13	**-1.88**	**-6,03**
337523	5,15	0,0383	-0,91	0,03	-0,25	0,06	0,13	0,13	**-2.91**	**-5,36**

- 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

Summary

- Problem definition
- Voice Production
- Neural Pathways controlling Voice
- First Inverse Model
- Second Inverse Model
- Third Inverse Model
- Statistical validation
- Study Cases
- Emotional Biomechanical Distortion
- Conclusions

Emotion Detection in Voice

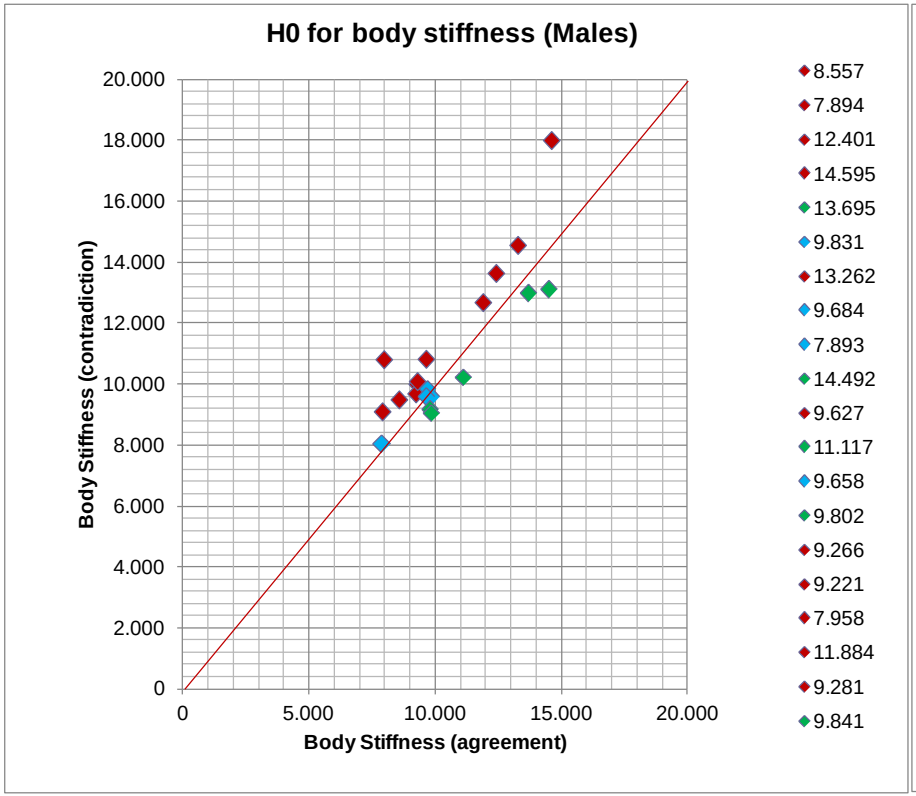
Emotion detection in voice: Materials and Methods

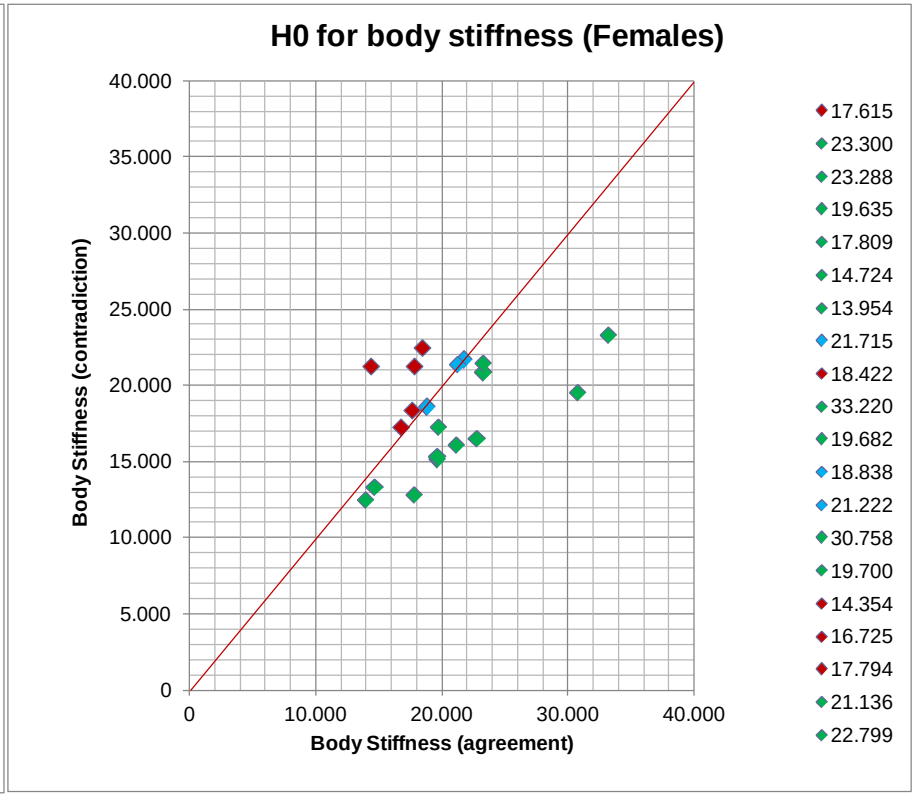
From the Master's Thesis of Elena Bartolomé Morala (2012): "Contribución al estudio de las alteraciones de la fonación en habla contradictoria frente a espontánea"

- 40 speakers balanced by gender were interviewed
- In a first recording samples of mainteined vowels /a, e, o/ in modal phonations were obtained
- In a first interview they were asked to produce opinions freely on "hot issues" (economic crisis, unemployment, politicians, etc.)
- In a second interview they were asked to produce opinions on the same issues but expressing an opinion contradictory with the former one
- Interviews were 3-5 min. long, interviewer was absent
- Fillers of vowel /e/ produced inadvertedly by speakers during natural speech (from words as /de/, /que/…) were isolated and used in the study
- Biomechanical parameters were obtained from 200 ms long segments on these vowels (vocal fold body and cover mass and stiffness)
- Tremor cyclicality parameters were obtained from 400 ms long segments
- Comparisons between pre- (self-conform) and post- (contradictory) estimates were used in the study

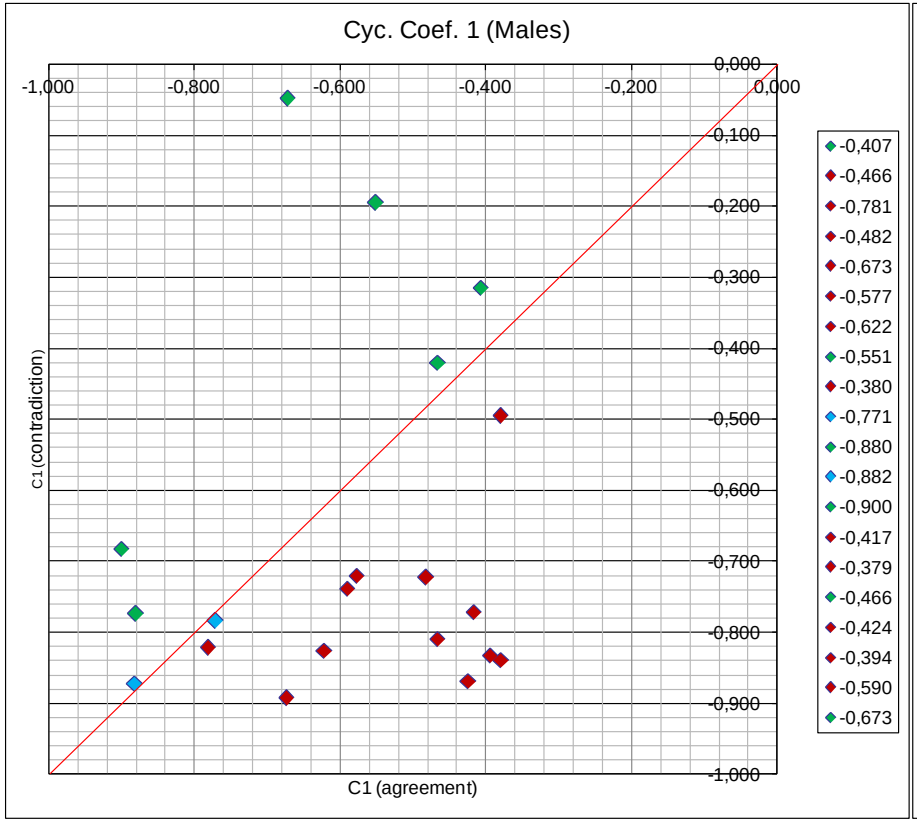
Emotional effects (over-stiffness)

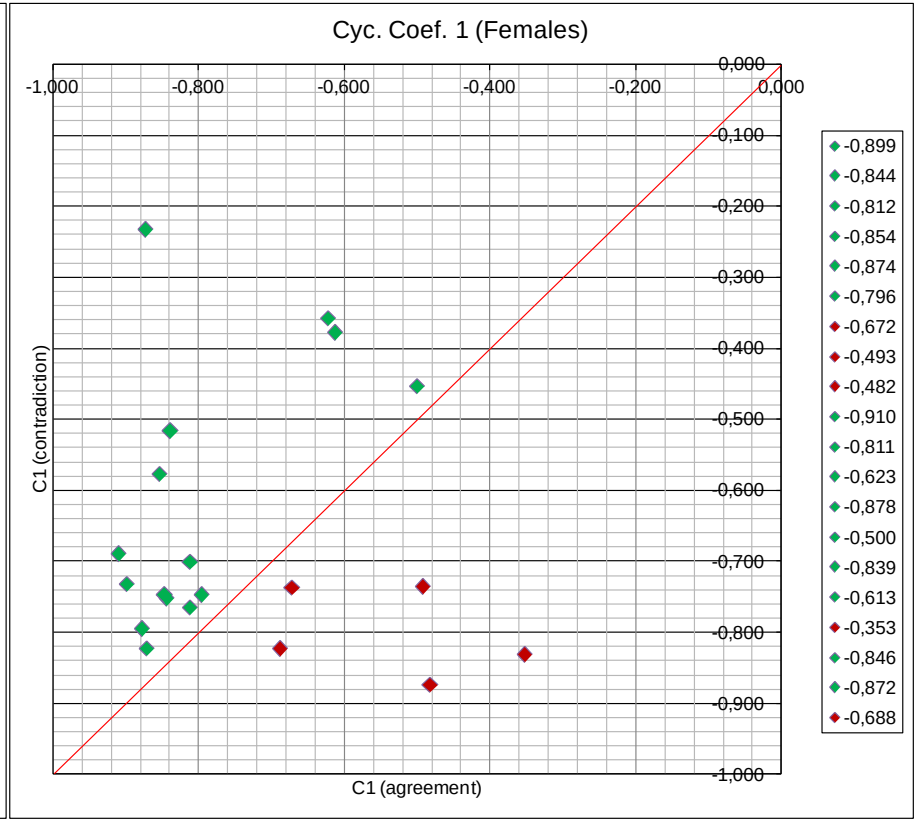




Emotional effects (tremor)





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 under (perceived) or over 10 Hz (not perceived)
- Tremor may be detected 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 esential from emotional, intentional or pathological tremor
- Emotional: Doddington's Zoo again?
- Larger databases are required
- Better modelling of upper neural pathways have to be investigated

How far may we go?

- Techniques:
  - Pitch and energy contours
  - Articulatory: distortions in formant space (Sapir)
  - 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 Sing education and rehabilitation

Contact us

Pedro Gómez-Vilda, PhD
Head of Research Group GIAPSI

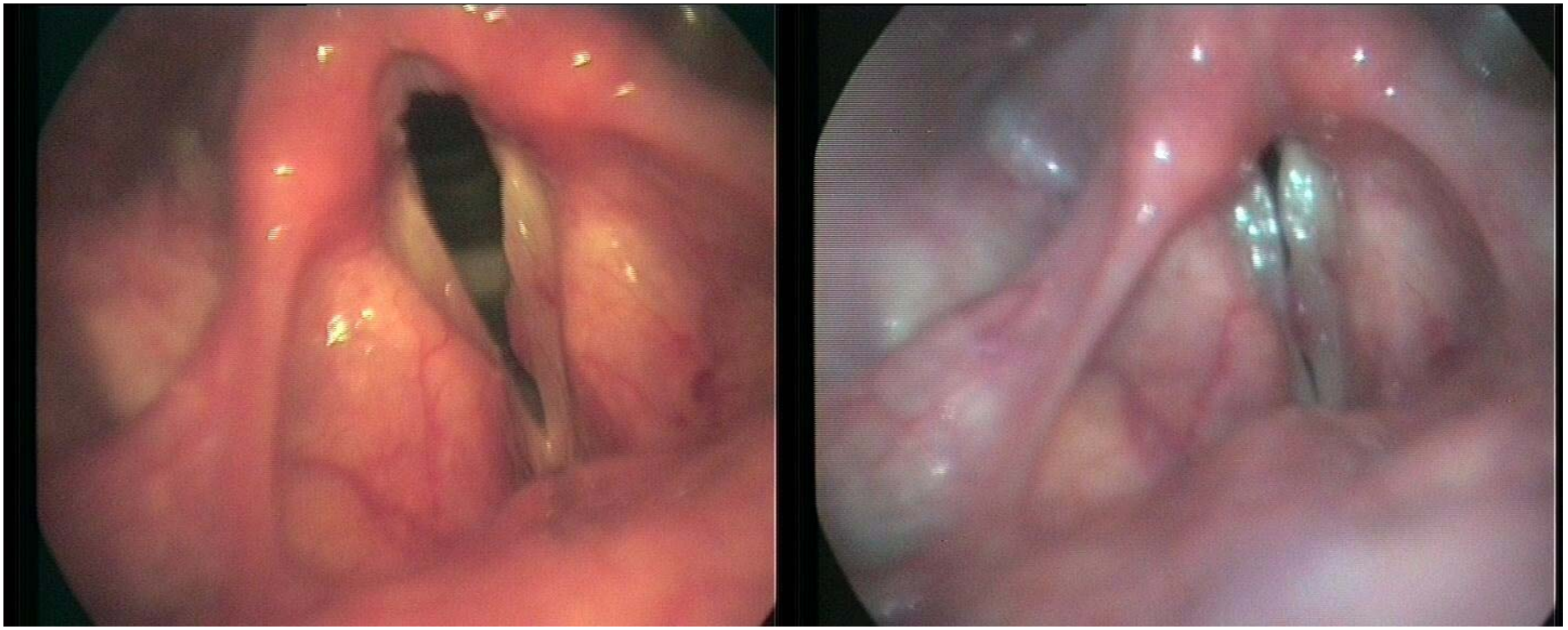
Neuromorphic Speech Processing Lab
Center for Biomedical Technology
Campus de Montegancedo, s/n, 28223 Pozuelo de Alarcón, Madrid, Spain

Tel: +34.913367384; +34.913364661
Fax: +34.913366601

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

URL: www.ctb.upm.es

How transmissive glottis may be



F:\Taller\Voz_Patologica\bases-datos\HUGM-estudio-casos-patologicos\Estudio de Casos\42169 Edema