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

Motivation

- 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/

Considerations (1)

- 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
  - Dysarthia: 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[1,2]

[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.

[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

Considerations (2)

The objective of the present work will be to find new ways of characterizing ND dysfunction in voicing (hypertension, dispersion, unbalance, tremor)[3]

Mel-cepstrum parameters classically used in organic voice pathology detection can also be used in detecting ND[4]

Studies have been published on the detection of Parkinson from the degradation of the vowel triangle[5]

[3]Pantazis, Y., Koutsogiannaki, M., Stylianou, Y., "A novel method for the extraction of tremor", *Proc. of MAVEBA07*, Florence University Press, pp. 107-110, 2009.

[4]Tsanas, A., Little, M. A., McSharry, P. E., Ramig, L. O., "Accurate Telemonitoring of Parkinson's Disease Progression by Noninvasive Speech Tests", IEEE Trans. on Biomed. Eng., 57 (4), 884-893, 2010.

[5]Sapir, S., Ramig, L. O., Spielman, J., Fox, C., "Acoustic Metrics of Vowel Articulation in Parkinson's Disease: Vowel Space Area (VSA) vs. Vowel Articulation Index (VAI), Proc. of MAVEBA 2011, Florence, August 26-28, pp. 173-175, 2011.

PD Monitoring from Voice Quality Analysis

- 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, higer f0 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

Stem Question

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

A: Yes, we do!

Q: How?

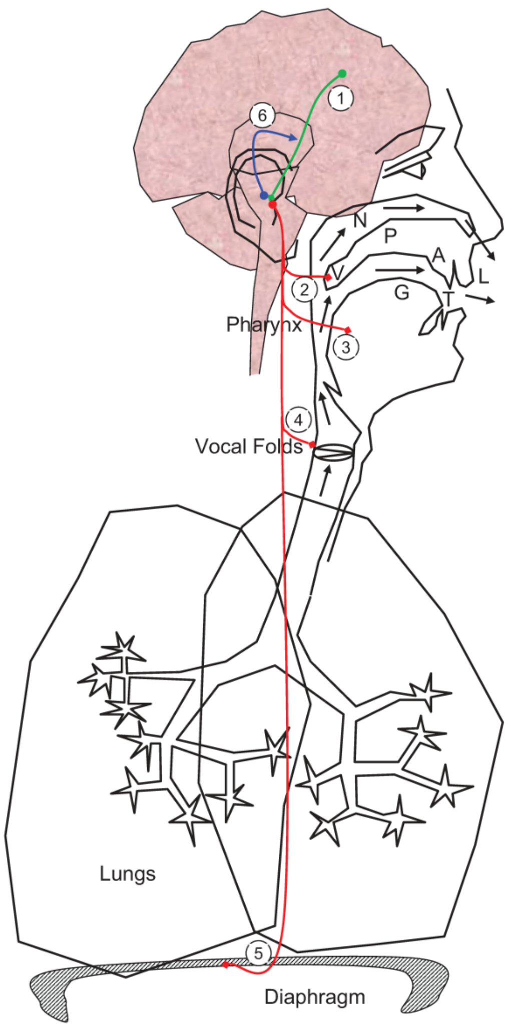
Main Hypothesis: Meaningful correlates of PD disorder in voice may be derived from vocal fold biomechanics (stiffness)

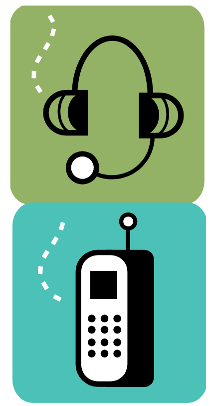
Secondary Hypothesis: Vocal fold stiffness is affected differently by Organic Larynx Diseases than by Neuromotor Disorders

The Challenge

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?





WhatCanWeDo?

ParametersTellingUs

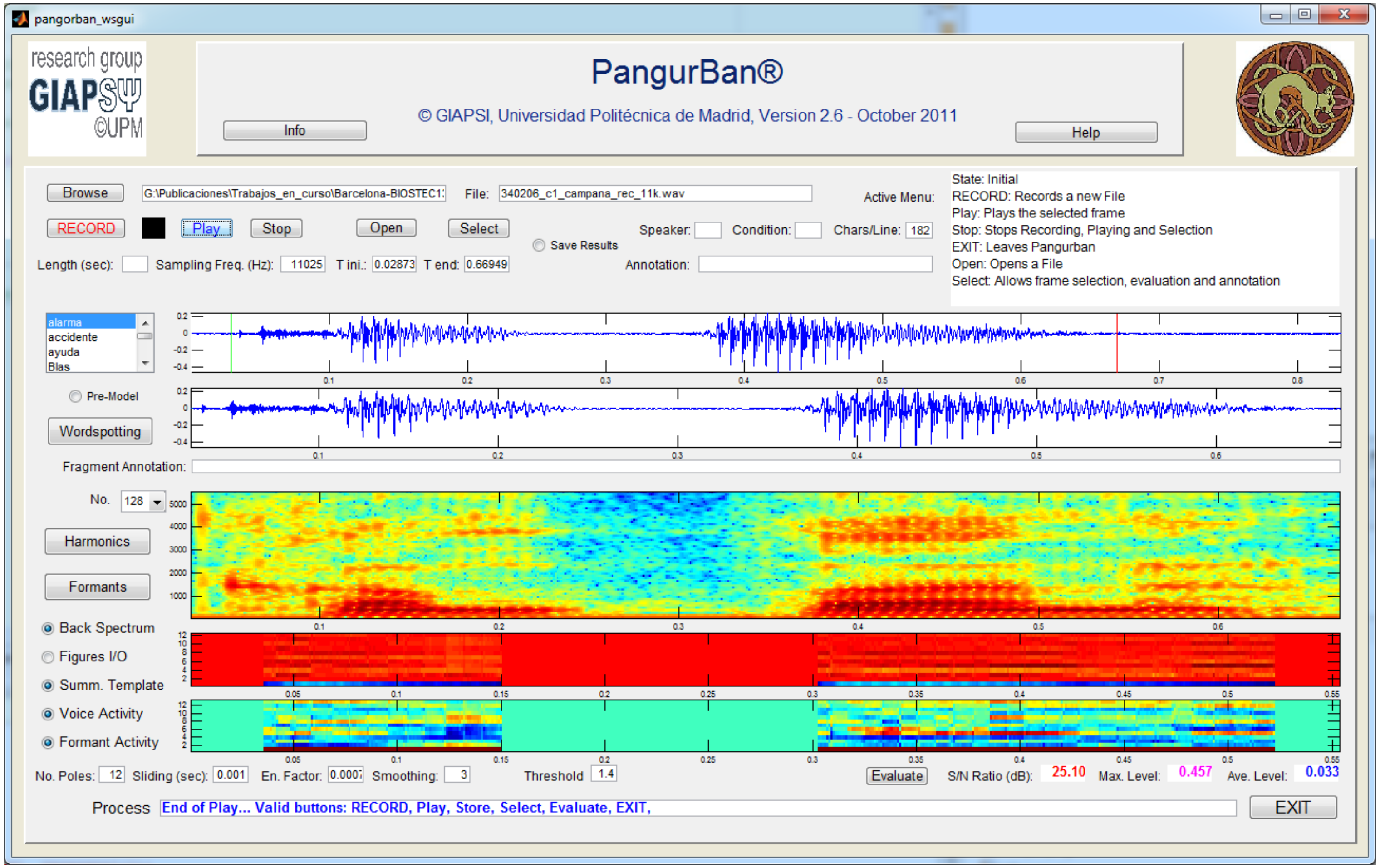
Gain in Knowledge

But WHAT IS 'Gain in Knowledge'?

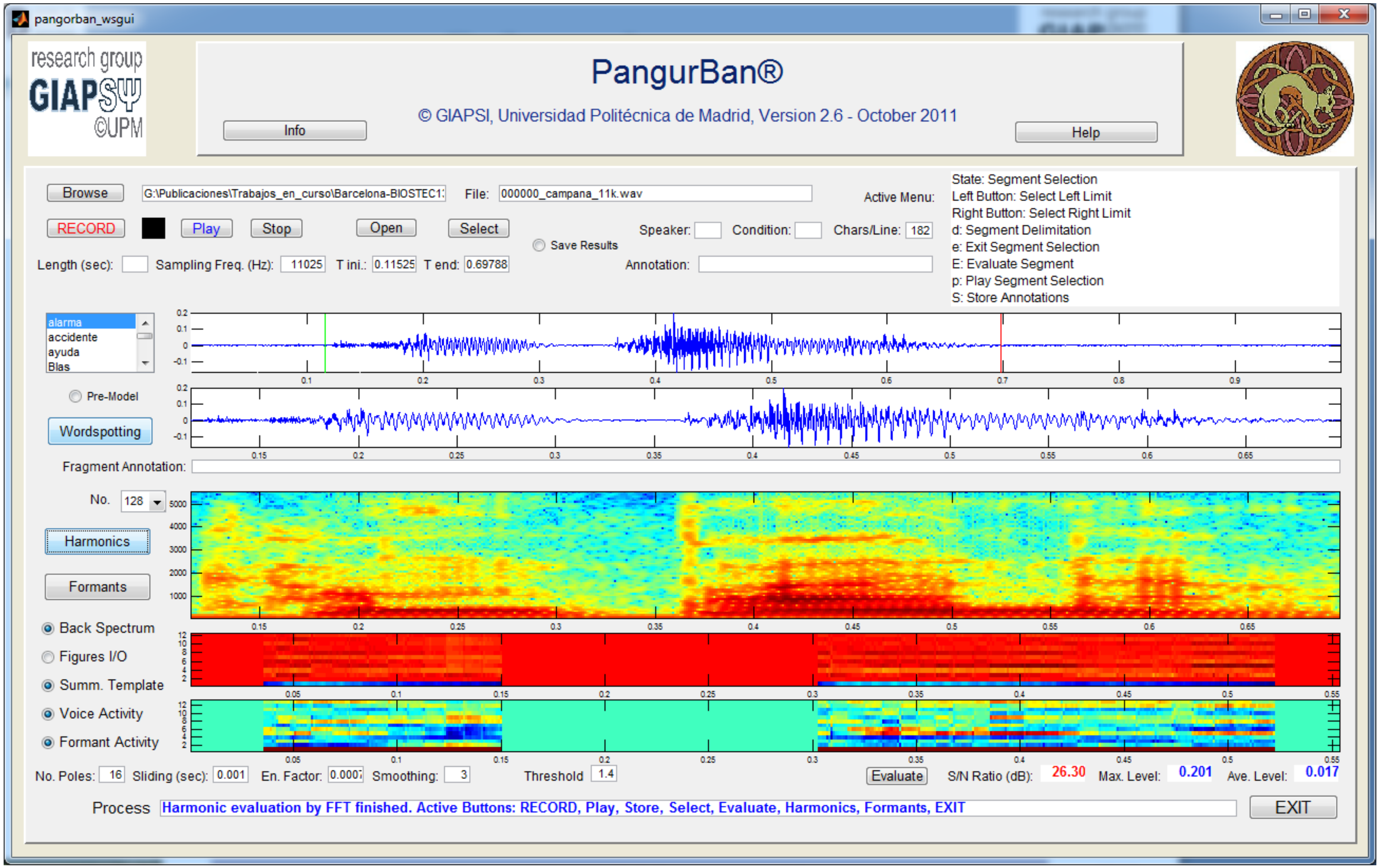
- Formulating a model to understand how voice is produced and **HOW V.F. STIFFNESS INFLUENCE VOICE**, by:
  - Analyzing top-down the neurophysiological pathways
    - how are built and connected
    - what do they do
    - building small models explaining each step (divide and conquer)
  - Synthesizing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

research group GIAPSI ©UPM



Referencia


pangorban_wsgui
research group GIAPSΨ ©UPM
PangurBan®
© GIAPSI, Universidad Politécnica de Madrid, Version 2.6 - October 2011
Info
Help
Browse
G:\Publicaciones\Trabajos_en_curso\Barcelona-BIOSTEC1:
File: 000000_campana_11k.wav
Active Menu:
State: Segment Selection
Left Button: Select Left Limit
Right Button: Select Right Limit
d: Segment Delimitation
e: Exit Segment Selection
E: Evaluate Segment
p: Play Segment Selection
S: Store Annotations
RECORD
Play
Stop
Open
Select
Save Results
Speaker:
Condition:
Chars/Line: 182
Length (sec):
Sampling Freq. (Hz): 11025
T ini.: 0.11525
T end: 0.69788
Annotation:
alarma
accidente
ayuda
Blas
Pre-Model
Wordspotting
Fragment Annotation:
No. 128
Harmonics
Formants
Back Spectrum
Figures I/O
Summ. Template
Voice Activity
Formant Activity
No. Poles: 16
Sliding (sec): 0.001
En. Factor: 0.000;
Smoothing: 3
Threshold 1.4
Evaluate
S/N Ratio (dB): 26.30
Max. Level: 0.201
Ave. Level: 0.017
Process
Harmonic evaluation by FFT finished. Active Buttons: RECORD, Play, Store, Select, Evaluate, Harmonics, Formants, EXIT
EXIT

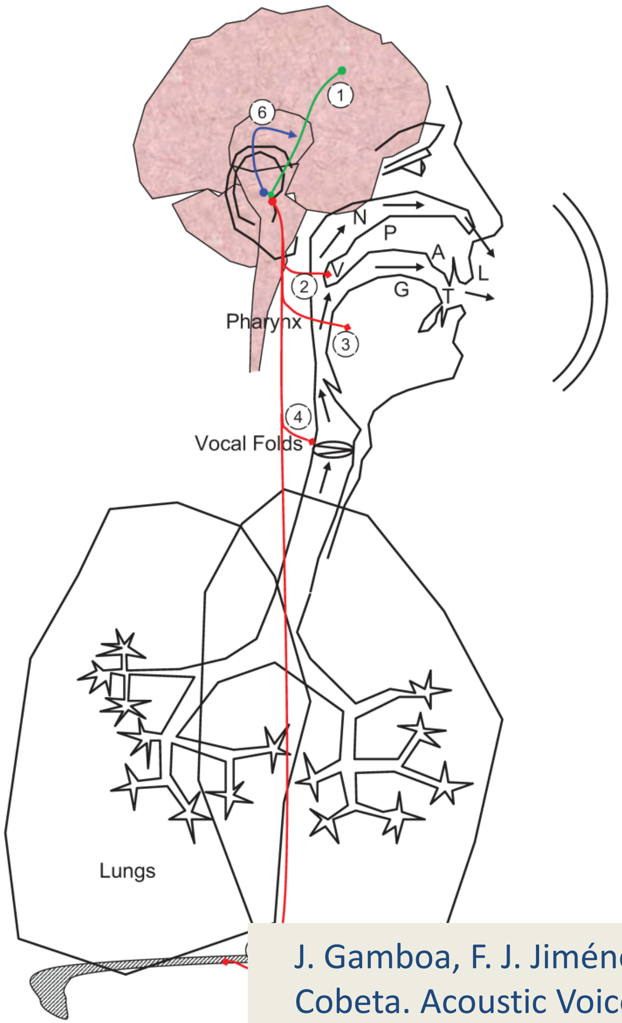
Visual & Auditive Tests: Diferences?

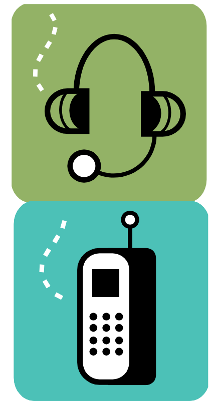
- 1^{st} Utterance:
  - The 1^{st} stop consonant /k/ is long and affricate
  - The 1^{st} /a/ is velarized
  - The interval previous to the 2^{nd} stop consonant is 0.38-0.24=0.14 s
  - The 2^{nd} stop consonant /p/ lacks the vertical plosive burst
  - The 2^{nd} /a/ has a long and slow vowel onset
  - As a result the 2^{nd} stop consonant is configured as a /b/
  - The 2^{nd} /a/ is nasalized, as well as the last one
  - The utterance may be transcribed as /k^{h}am.bãnã/
- 2^{nd} 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 3^{rd} /a/
  - The utterance may be transcribed as /k^{h}ampanã/

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

Early work from Gamboa et al. 97





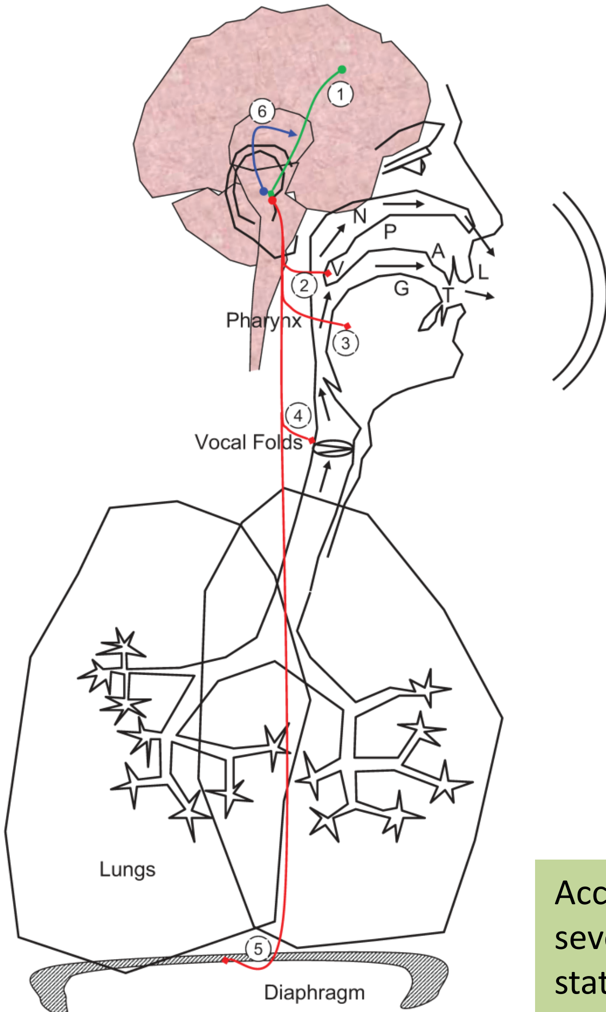
Comp Sp Lab Kay Elemetrics

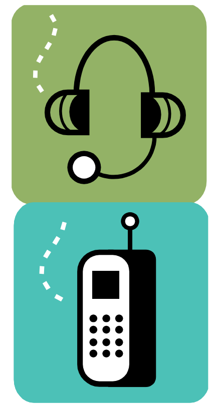
Jitter, shimmer, HNR and other organic pathol clues

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, 1997

Tsanas' Approach





MFCC extraction, GMM, SVM Clustering

↓

MFCC's

↓

Accurate, 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

Tsanas (2012) results

- 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

Table 7.6: Selected dysphonia measures using seven feature selection algorithms and classification performance for motor-UPDRS and total-UPDRS for males.

LASSO	mRMR	mRMR Spearman	GSO	RELIEF	LLBFS	RRCT
6th MFCC	VFER-NSR$_{TKEO}$	6th MFCC	6th MFCC	9th level app.coef$_{entropy}$ of the log-F0	9th level app.coef$_{entropy}$ of the log-F0	6th MFCC
8th MFCC	6th MFCC	2nd MFCC	VFER-SNR$_{TKEO}$	DFA	DFA	2nd MFCC
8th delta MFCC	7th MFCC	8th delta MFCC	8th MFCC	6th MFCC	8th MFCC	8th delta MFCC
VFER-SNR$_{TKEO}$	8th MFCC	12th delta MFCC	8th delta MFCC	3rd MFCC	7th MFCC	Std $F_{0,NDF} - F_{0,exp}$
0th MFCC	3rd level detail wav.coef$_{entropy}$	8th MFCC	HNR$_{std}$	5th MFCC	6th MFCC	10th level detail wav.coef$_{entropy}$
2nd MFCC	Log energy	10th level detail wav.coef$_{entropy}$	3rd MFCC	8th MFCC	VFER-NSR$_{TKEO}$	3rd delta MFCC
IMF-SNR$_{TKEO}$	8th delta MFCC	Shimmer$_{TKEO,std}$	2nd MFCC	7th MFCC	5th MFCC	12th delta MFCC
12th delta MFCC	3rd MFCC	3rd delta MFCC	12th delta MFCC	4th MFCC	1st MFCC	3rd MFCC
HNR$_{std}$	1st MFCC	Std $(F_{0,NDF} - F_{0,exp})$	11th MFCC	10th MFCC	3rd MFCC	8th delta MFCC
Shimmer$_{PQ5}$	VFER-SNR$_{TKEO}$	11th delta MFCC	VFER-NSR$_{TKEO}$	11th MFCC	4th MFCC	Std $F_{0,SHRP} - F_{0,exp}$
3rd MFCC	HNR$_{std}$	0th MFCC	RPDE	9th MFCC	9th MFCC	12th MFCC
4th level detail wav.coef$_{entropy}$	5th MFCC	12th MFCC	Shimmer$_{PQ5}$	Log energy	Log energy	11th delta MFCC
10th level detail wav.coef$_{entropy}$	4th MFCC	GNE-NSR$_{TKEO}$	10th level detail wav.coef$_{entropy}$	2nd MFCC	12th MFCC	Log energy
GNE-SNR$_{SEO}$	8th level detail wav.coef$_{entropy}$	Std $F_{0,SHRP}$	2nd delta MFCC	VFER-NSR$_{TKEO}$	10th MFCC	GNE-SNR$_{SEO}$
2nd delta MFCC	11th MFCC	VFER-NSR$_{TKEO}$	9th delta MFCC	12th MFCC	$(F_{0,NDF} - F_{0,exp})_{prc5-95}$	GNE-SNR$_{TKEO}$
2.25 ± 0.19	1.60 ± 0.17	2.32 ± 0.19	2.21 ± 0.20	1.33 ± 0.14	1.31 ± 0.13	2.15 ± 0.18
2.84 ± 0.24	2.02 ± 0.19	2.95 ± 0.28	2.85 ± 0.27	1.59 ± 0.17	1.70 ± 0.18	2.79 ± 0.25

Table 7.7: Selected dysphonia measures using seven feature selection algorithms and classification performance for motor-UPDRS and total-UPDRS for females.

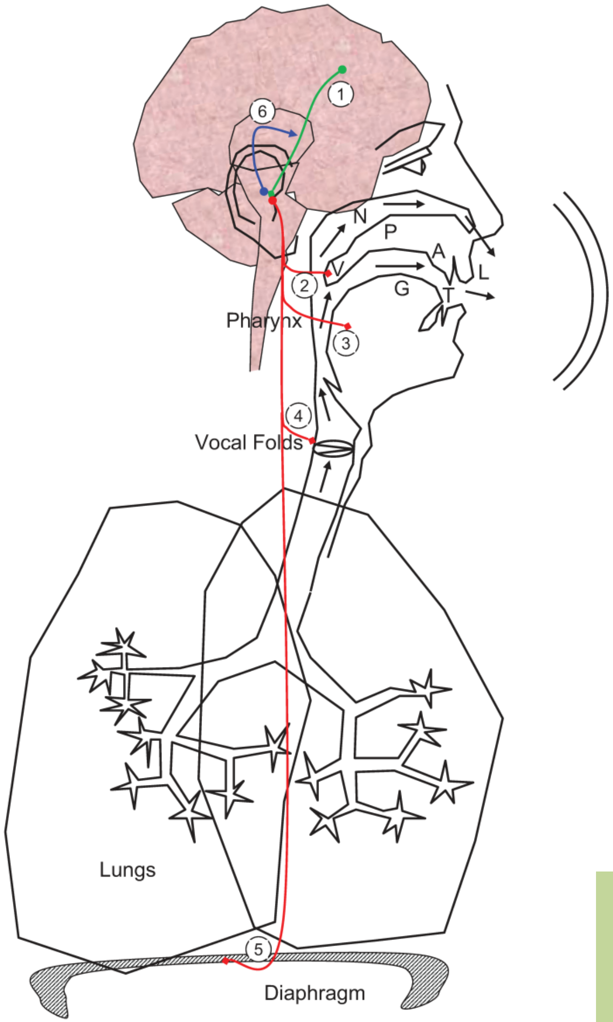
LASSO	mRMR	mRMR Spearman	GSO	RELIEF	LLBFS	RRCT
Log energy	Std $F_{0,RAPT}$	Log energy	Log energy	4th MFCC	Log energy	Log energy
Std $F_{0,RAPT}$	Log energy	NHR$_{std}$	10th MFCC	Log energy	Std $F_{0,RAPT}$	NHR$_{mean}$
HNR$_{mean}$	$(F_{0,ensemble} - F_{0,exp})_{mean}$	PPE	PPE	2nd MFCC	4th MFCC	$(F_{0,ensemble} - F_{0,exp})_{mean}$
4th level detail wav.coef$_{energy}$ log-F0	10th MFCC	4th level detail wav.coef$_{TKEO,mean}$ log-F0	0th MFCC	0th MFCC	11th MFCC	12th MFCC
10th MFCC	4th level detail wav.coef$_{energy}$	12th MFCC	8th MFCC	1st MFCC	1st MFCC	10th MFCC
PPE	VFER-SNR$_{TKEO}$	0th MFCC	12th MFCC	5th MFCC	2nd MFCC	4th level app. wav.coef$_{energy}$
12th MFCC	12th MFCC	3rd level detail wav.coef$_{TKEO,mean}$ log-F0	HNR$_{std}$	Std $F_{0,RAPT}$	10th MFCC	3rd delta MFCC
8th MFCC	4th level detail wav.coef$_{TKEO,std}$ log-F0	6th MFCC	11th MFCC	6th MFCC	PPE	PPE
11th MFCC	1st MFCC	2nd delta MFCC	Jitter$_{F0,TKEO,mean}$	PPE	0th MFCC	4th MFCC
5th level detail wav.coef$_{energy}$ log-F0	3rd MFCC	Jitter$_{F0,TKEO,mean}$	4th MFCC	8th MFCC	3rd MFCC	3rd level detail wav.coef$_{TKEO,std}$ log-F0
5th level detail wav.coef$_{TKEO,mean}$ log-F0	0th MFCC	5th level detail wav.coef$_{TKEO,mean}$ log-F0	3rd delta MFCC	3rd MFCC	4th level detail wav.coef$_{entropy}$	12th delta MFCC
HNR$_{std}$	3rd level detail wav.coef$_{entropy}$	10th MFCC	7th MFCC	7th MFCC	5th MFCC	3rd level detail wav.coef$_{entropy}$ log-F0
4th level detail wav.coef$_{TKEO,std}$ log-F0	4th level detail wav.coef$_{entropy}$ log-F0	4th level detail wav.coef$_{energy}$ log-F0	10th delta MFCC	11th MFCC	6th level detail wav.coef$_{TKEO,std}$	5th level detail wav.coef$_{TKEO,std}$ log-F0
2nd delta MFCC	6th level detail wav.coef$_{TKEO,mean}$ log-F0	HNR$_{mean}$	4th level detail wav.coef$_{energy}$ log-F0	10th MFCC	3rd level detail wav.coef$_{entropy}$	Shimmer$_{AM}$
3rd level detail wav.coef$_{entropy}$ log-F0	11th MFCC	12th delta MFCC	$(F_{0,ensemble} - F_{0,exp})_{mean}$	3rd level detail wav.coef$_{entropy}$ log-F0	HNR$_{mean}$	2nd delta MFCC
2.83 ± 0.29	1.99 ± 0.20	2.42 ± 0.25	2.02 ± 0.23	1.58 ± 0.19	1.72 ± 0.18	2.37 ± 0.22
3.26 ± 0.31	2.42 ± 0.24	2.99 ± 0.30	2.43 ± 0.20	1.95 ± 0.19	2.13 ± 0.20	2.91 ± 0.26

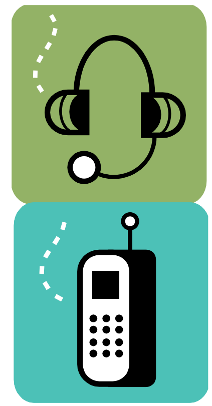
UPDRS

- Unified Parkinson's Disease Rating Scale: http://www.mdvu.org/library/ratingscales/pd/updrs.pdf
- Quantifies subjectively 5 levels of intensity from each behavioral feature:
  - I. Mentation, Behavior and Mood
  - II. Activities of daily living
  - III. Motor Examination

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

Sapir's Approach





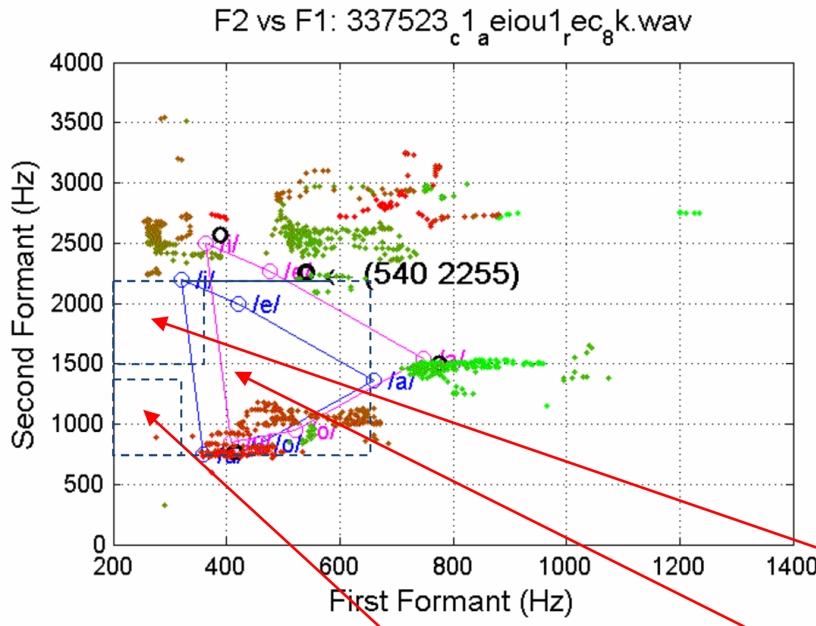
Vowel Triangle Estimation

Formant-based Parameters

VT Shrinkage, may serve in AD

Sapir, S., Ramig, L. O., Spielman, J., Fox, C., "Acoustic Metrics of Vowel Articulation in Parkinson's Disease: Vowel Space Area (VSA) vs. Vowel Articulation Index (VAI), Proc. of MAVEBA 2011, Florence, August 26-28, pp. 173-175, 2011.

More on Sapir's Approach



Vowel Space Area:

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

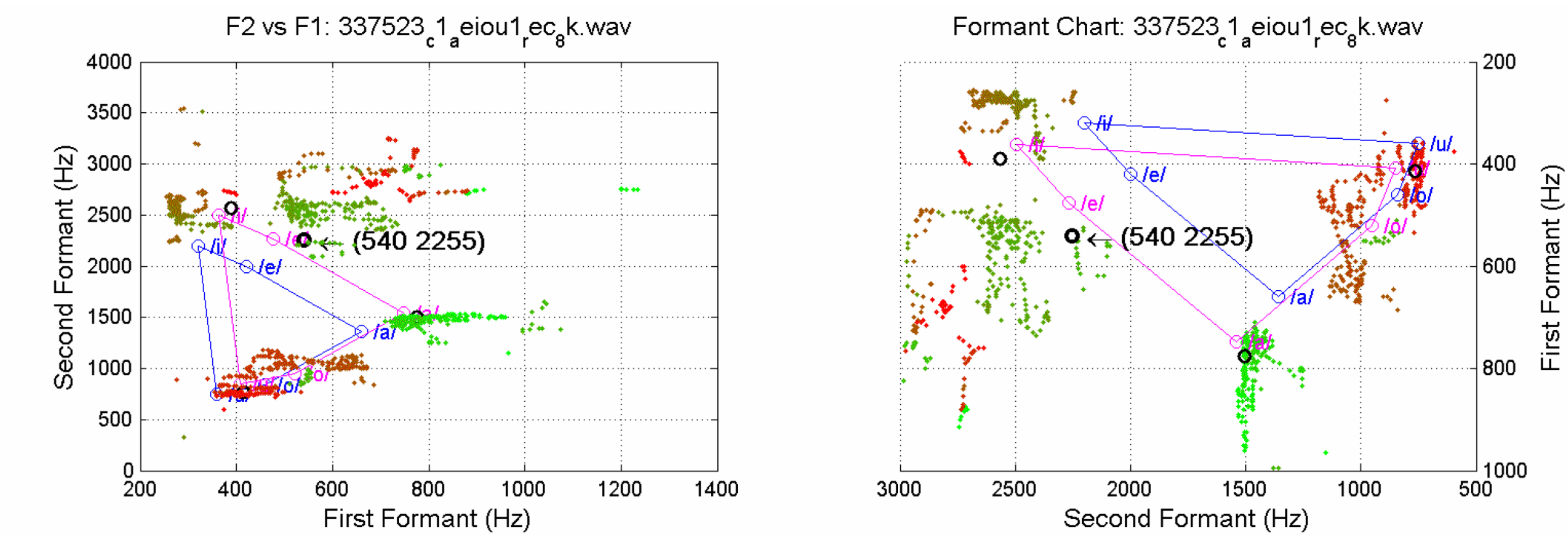
Vowel Articulation Index:

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

Formant Centralization Ratio:

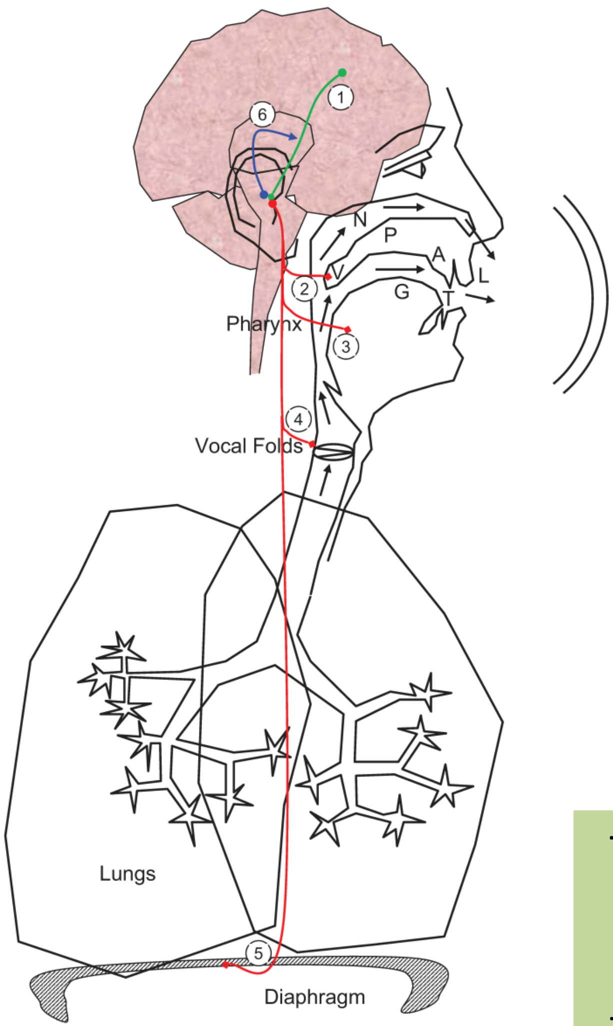
$$FCR = \frac{f_{2u} + f_{2a} + f_{1u} + f_{1i}}{f_{2i} + f_{1a}}$$

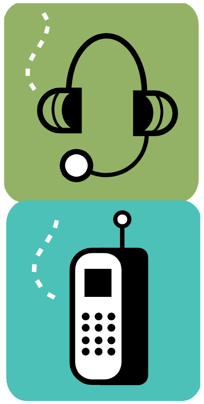
& More on Sapir's Approach



	f1u	f1a	f1i	f2u	f2a	f2i	VSA	VAI	FCR
Av. Male	380	650	340	800	1400	2200	709000	0.9760	1.0246
Av. Female	400	780	380	800	1500	2600	1055000	1.0974	0.9112
F 75y PD	420	850	280	800	1500	2600	1094000	1.1500	0.8696

Present Approach





Vocal Tract Removal & Fold Tension Estimation

Hyper-Tension and Cyclicity Modelling

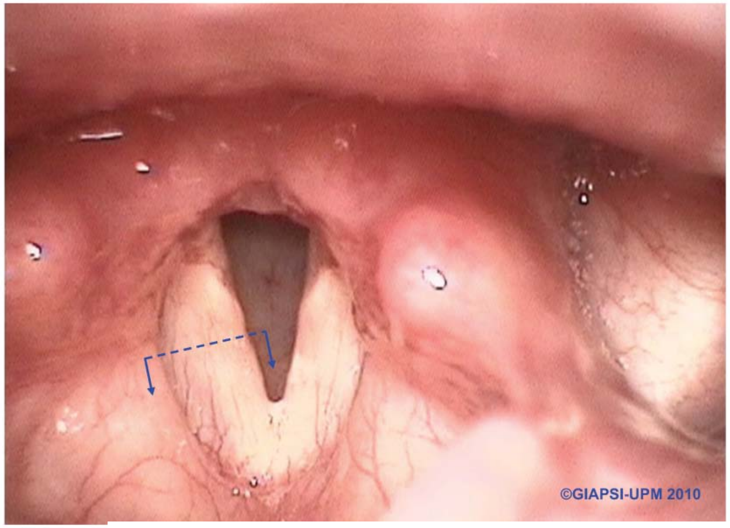
Hyper-Tension, Asymmetry, Dispersion, Tremor

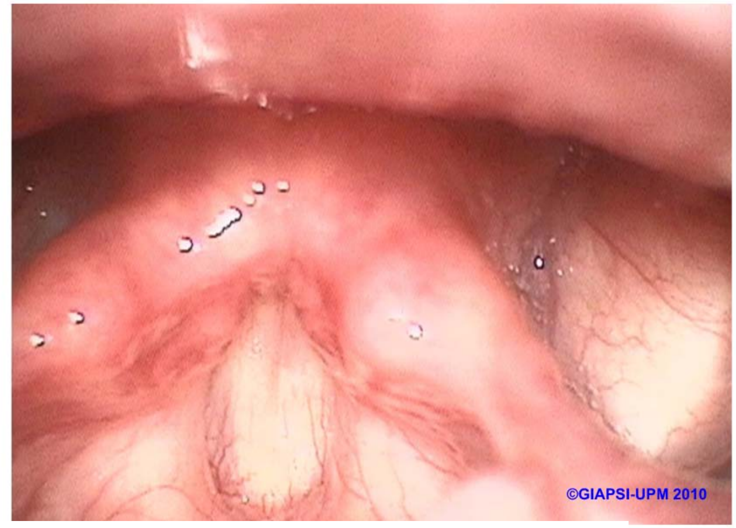
These correlates are ambiguous, but toghether may produce robust markers in ND (PD, AD, LAS, HC, ET, etc.) Even emotional and intentional tremor may be treated center for biomedical technology

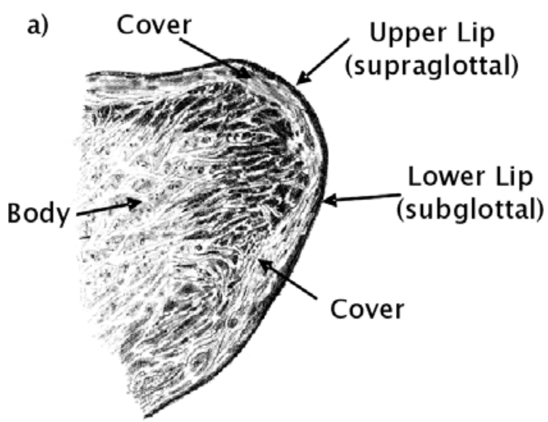
CAMPUS DE EXCELENCIA INTERNACIONAL

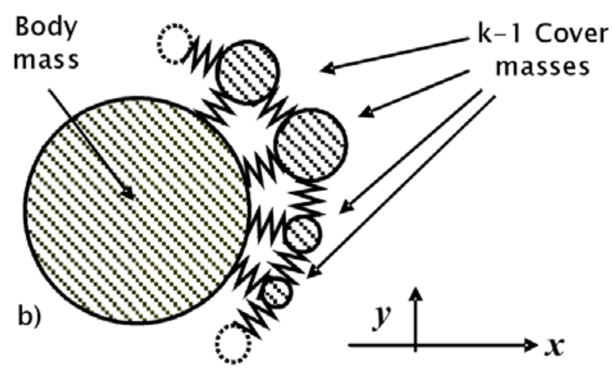
research group GIAPSΨ ©UPM

Voicing is produced by the vocal folds

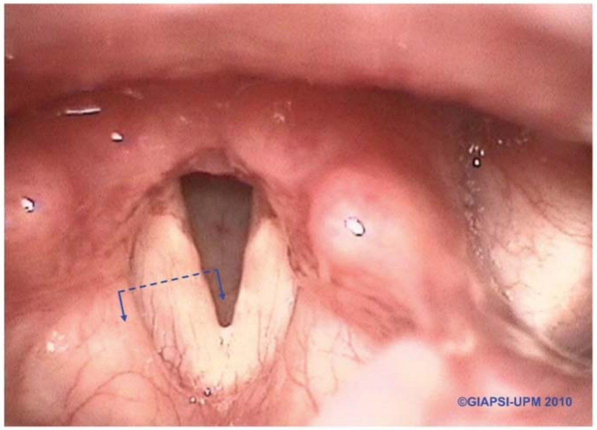








The phonation cycle



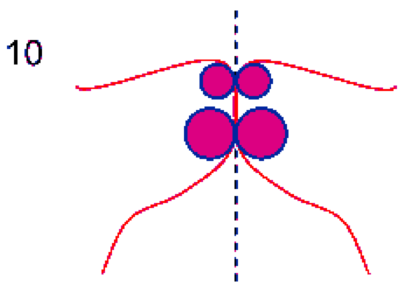
Data taken from Story (2002)

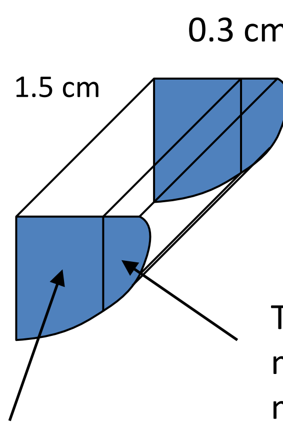
Total Volume of the fold:

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

Volume of right section:

$0.011\ cm^3$

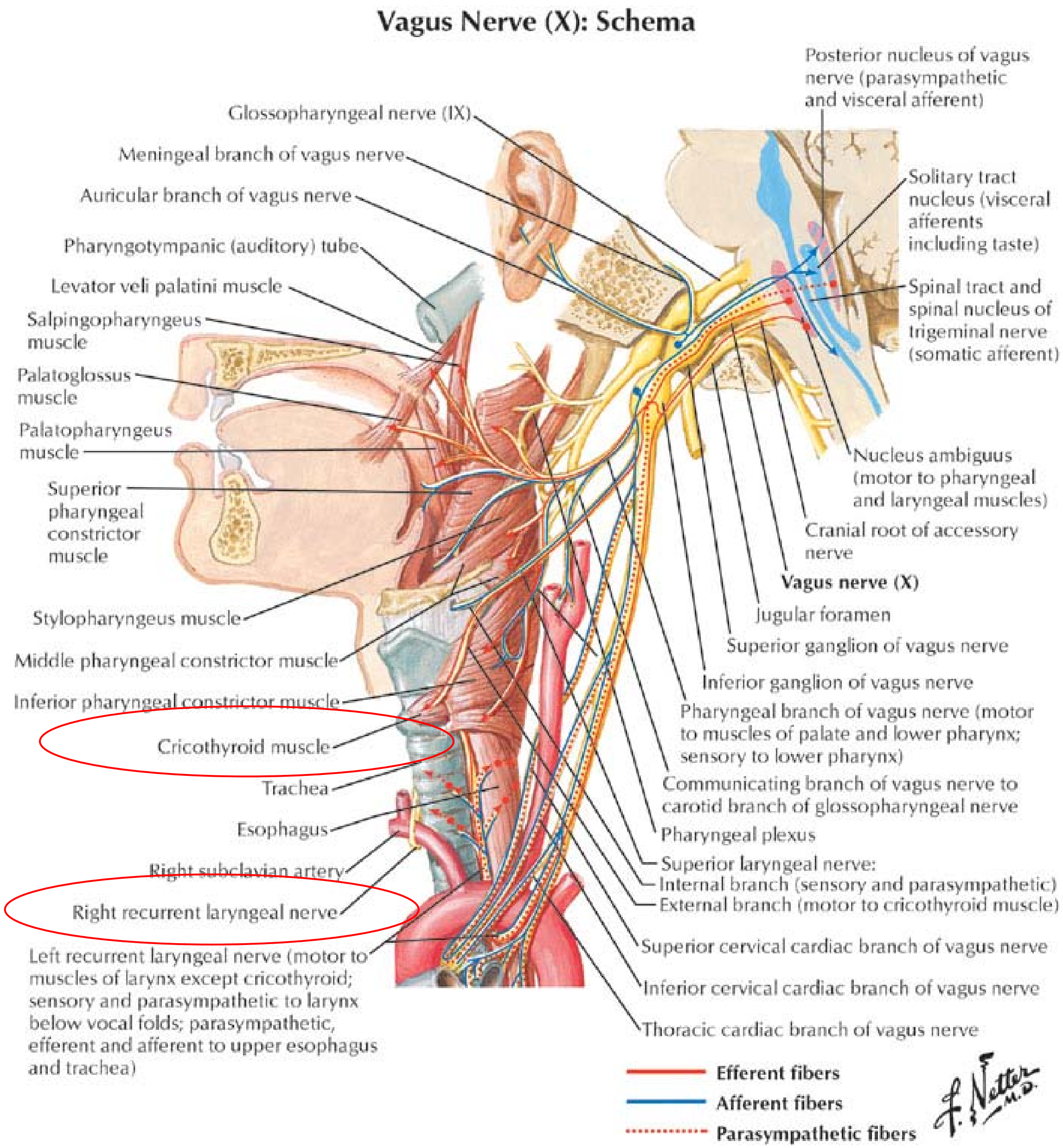




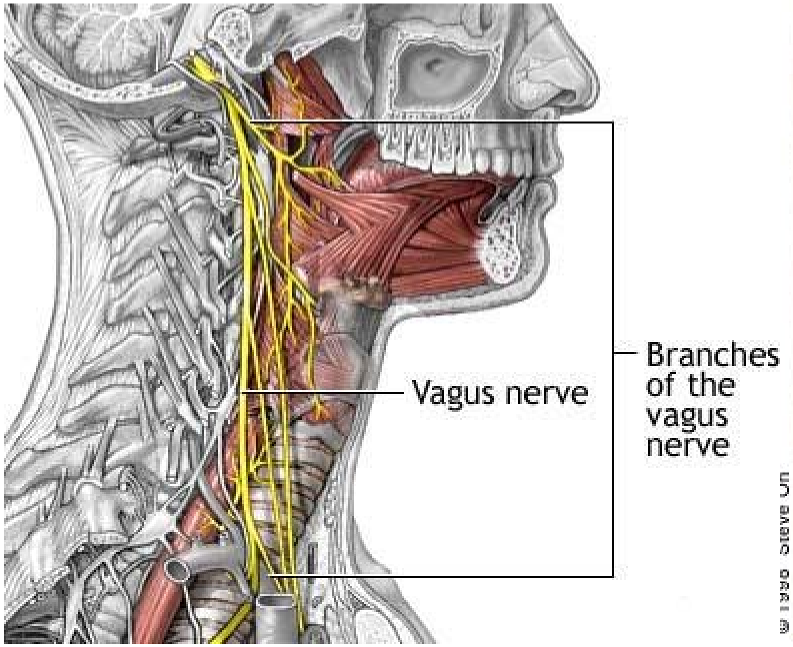
This section behaves more as an inertial mass

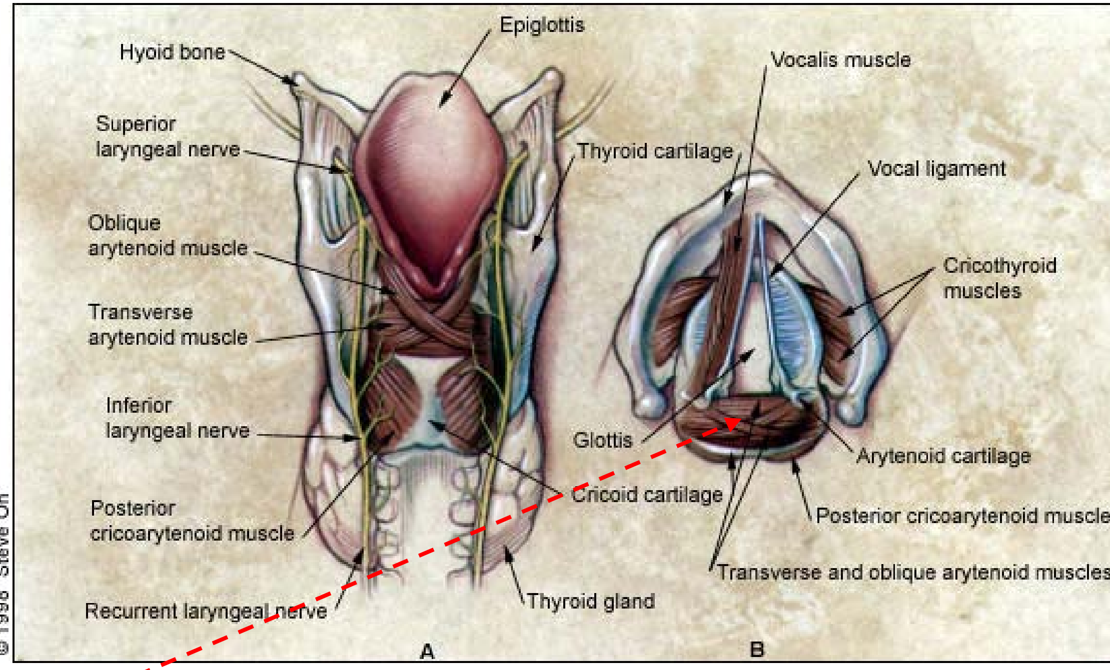
This section behaves more as a reactive spring

Descending Neural Pathways to Larynx



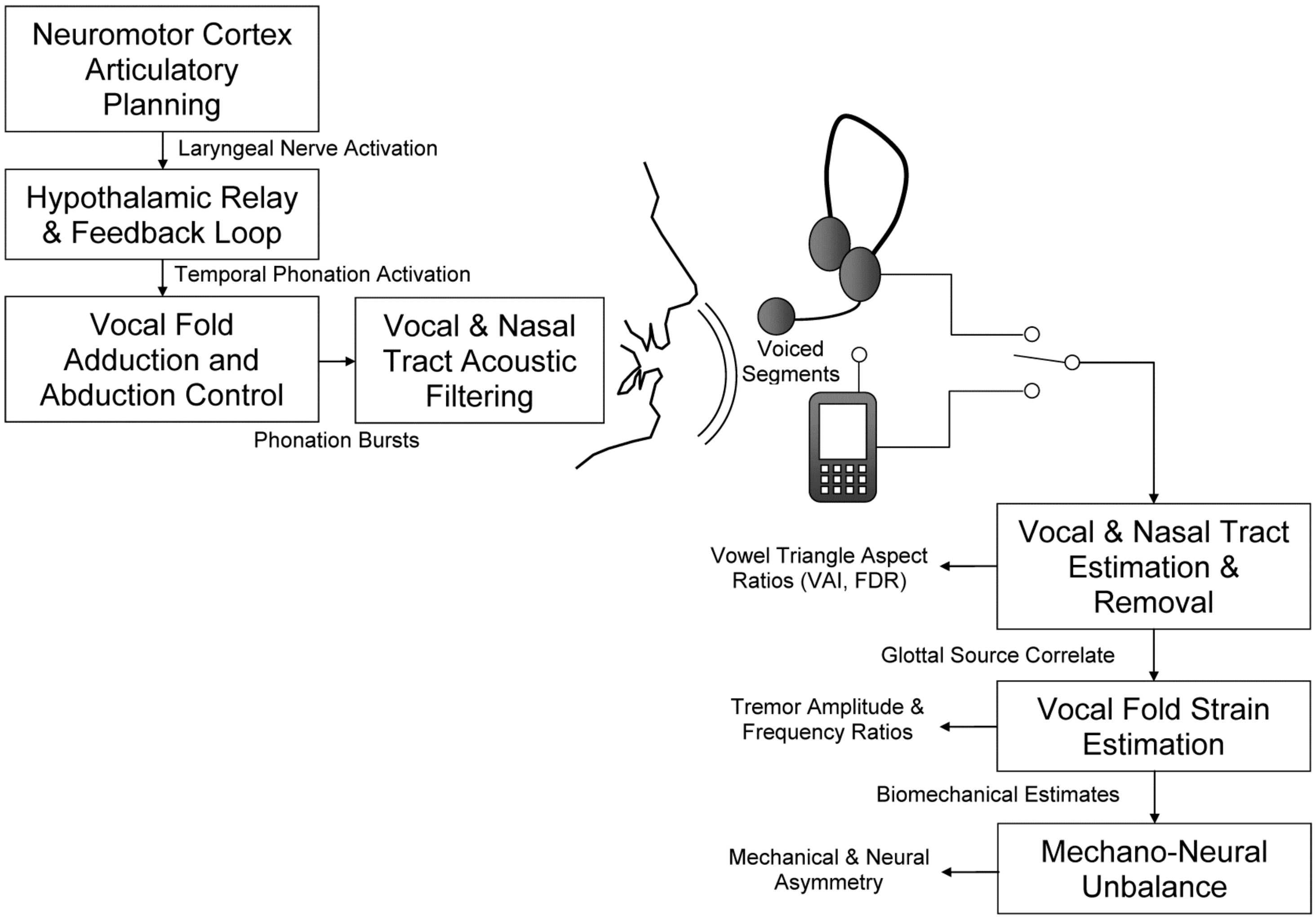
Larynx Innervation



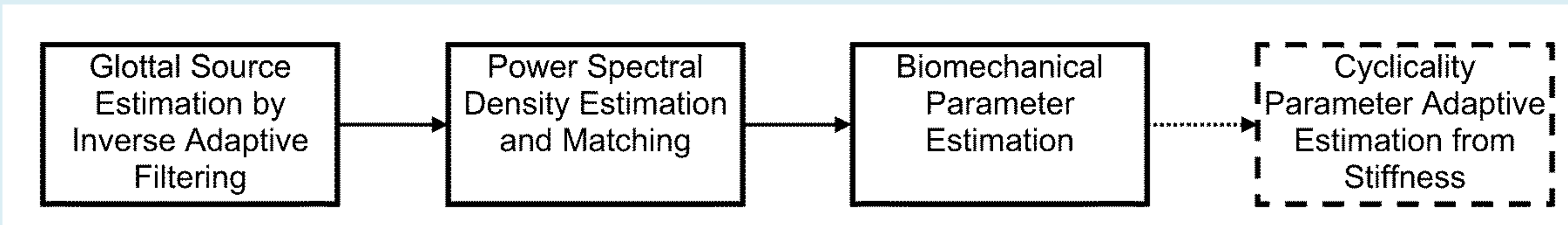


The innervation of the **transverse and oblique arytenoid muscles** by the **superior, inferior and transversal** laryngeal nerves is responsible of the vocal fold tension by enlargement or shortening of the musculus vocalis, see: Luschei, E. S., Ramig, L. O., Baker, K. L., Smith, M. E., "Discharge characteristics 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



Model Inversion: Three inverse problems

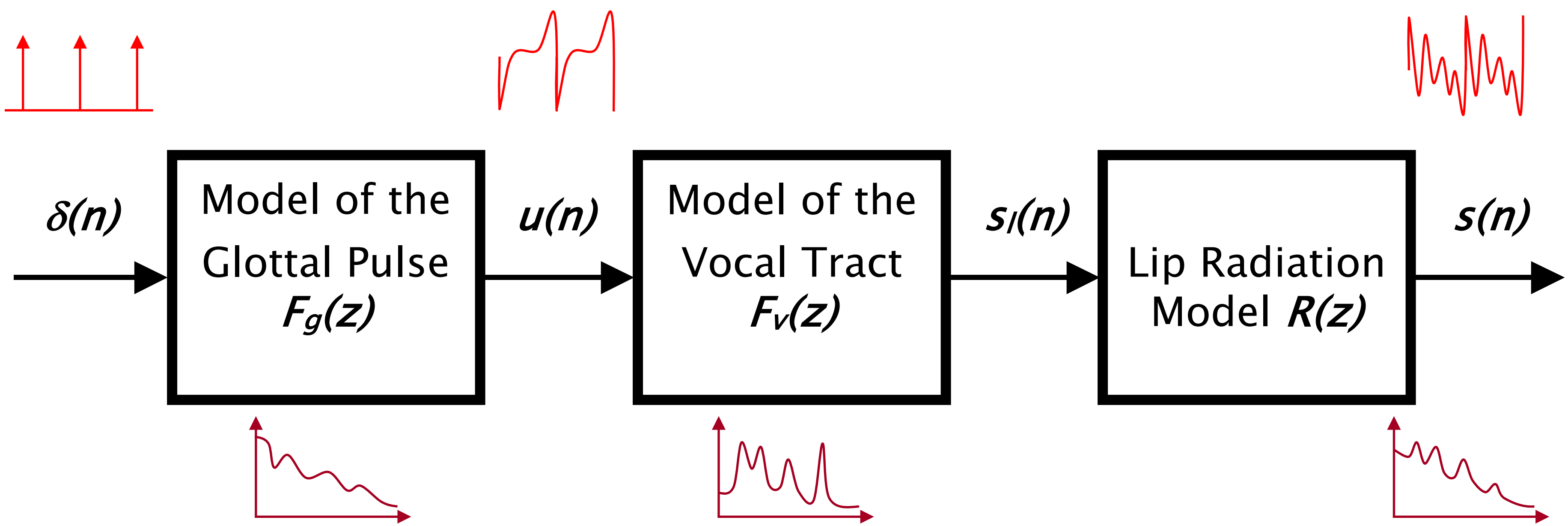


- **First level inversion**
  - Vocal Tract Model: Voice→Glottal Source
- **Second level inversion**
  - 2-mass vocal fold biomechanics: Glottal Source PSD→mass, viscoelasticity
- **Third level inversion**
  - Tremor, over-tension: Vocal Fold Body Stiffness→Cyclicality

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)

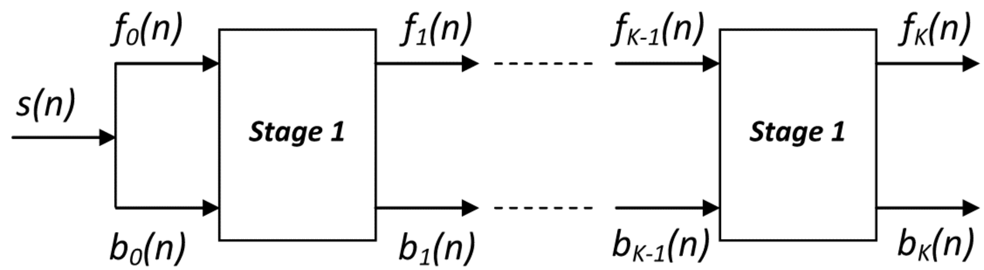
First Model Inversion: Glottal Source Reconstruction

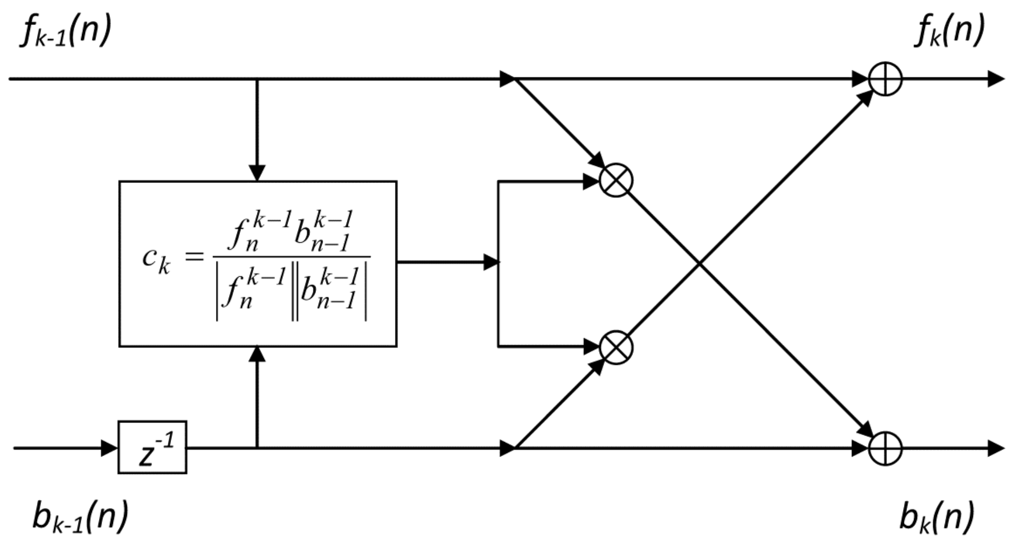
In the voice production model of Gunnar Fant it is assumed that the glottal source is produced by a train of delta pulses δ(n) which are modeled by a Glottal Function $F_g(z)$ to reproduce the glottal source u(n). This signal, when injected in the vocal tract composed by a chain of tubes $F_v(z)$ produces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



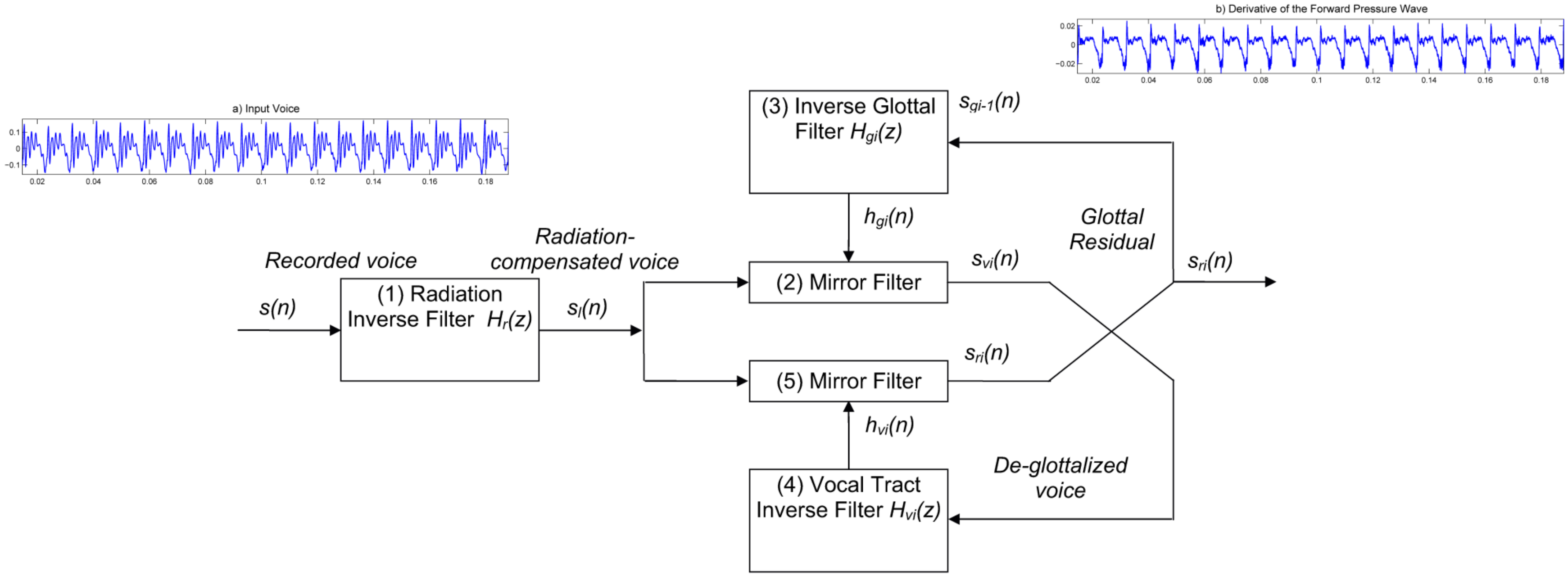
Lattice Prediction-Error Filters

- 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)

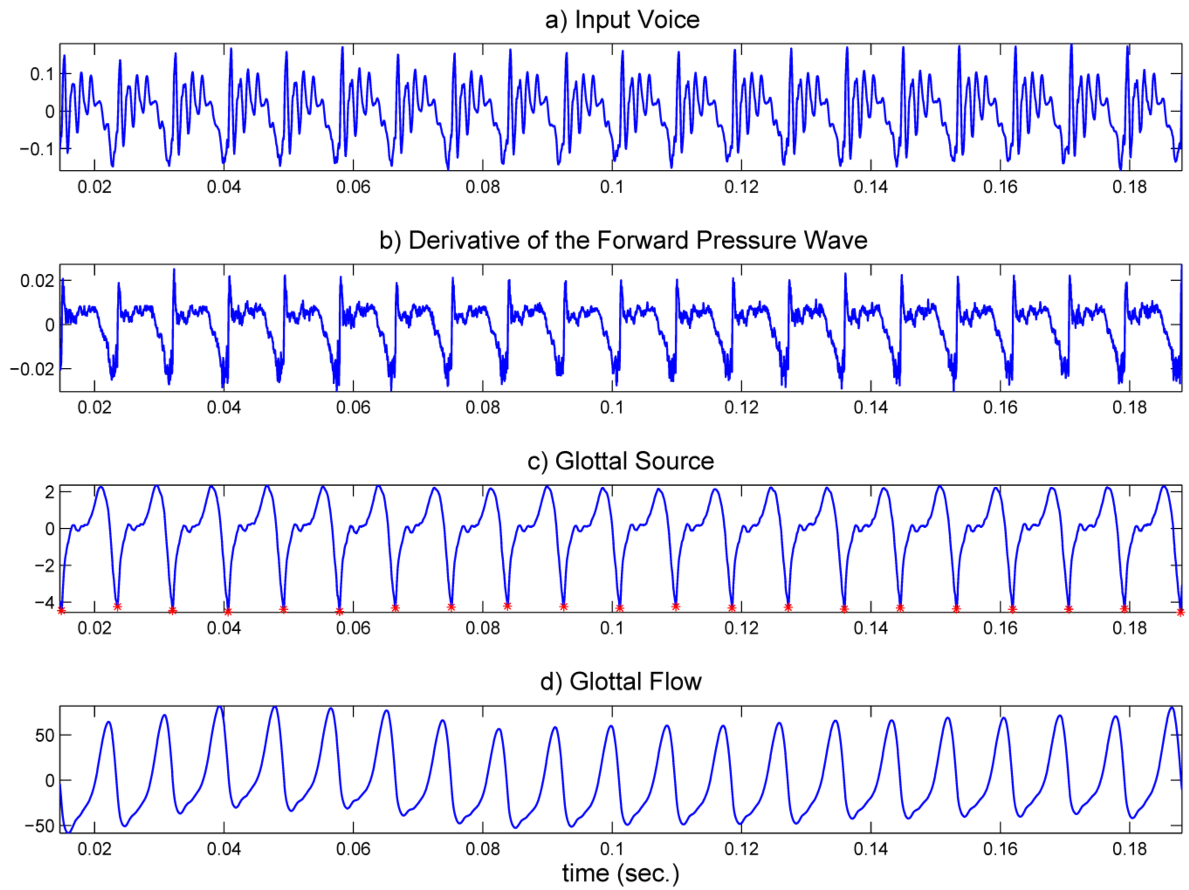




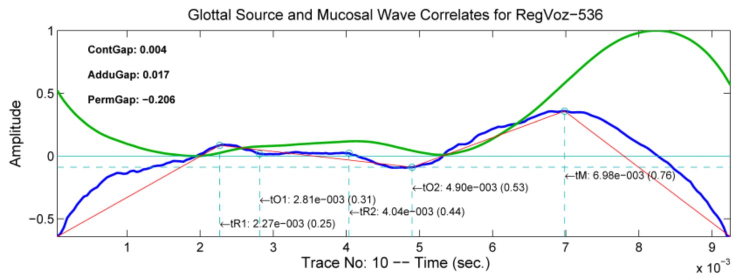
Estimation of Glottal Source by Inverse Filtering

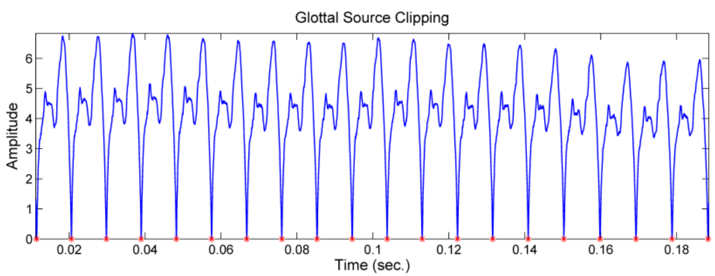


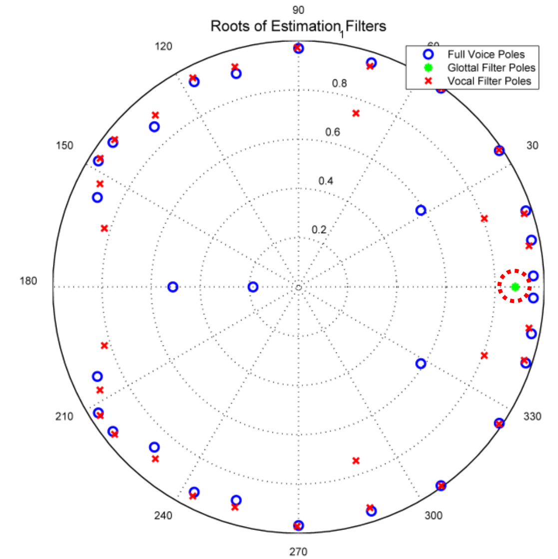
Glottal Correlates

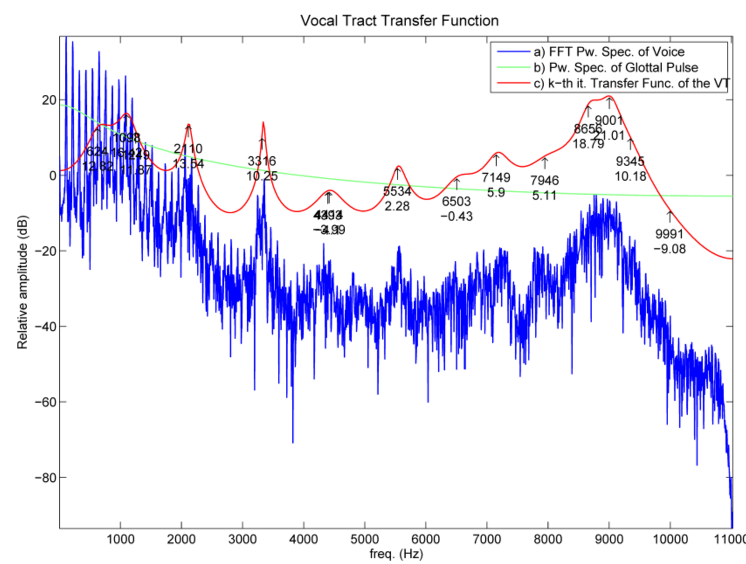


Prototype Male Speaker (1)



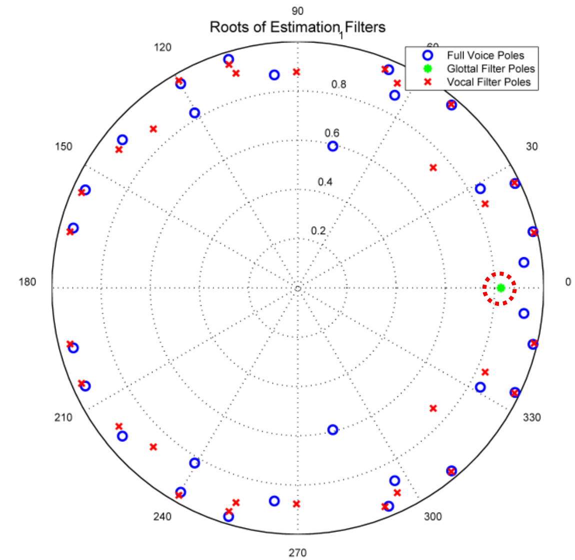


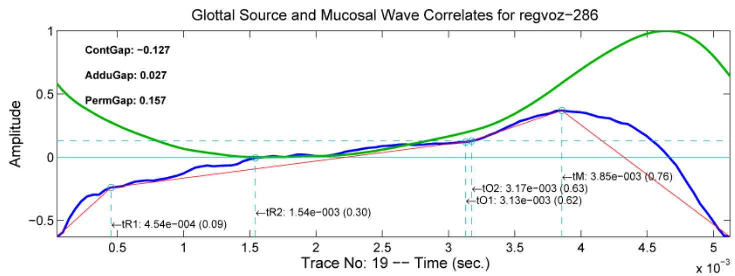


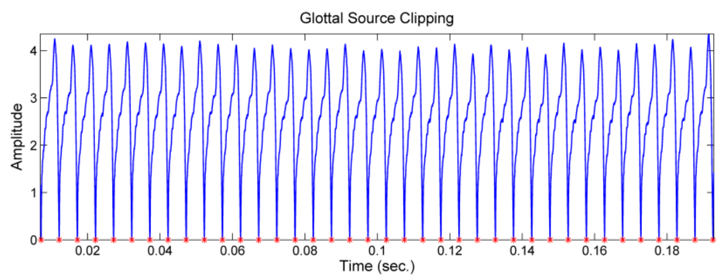


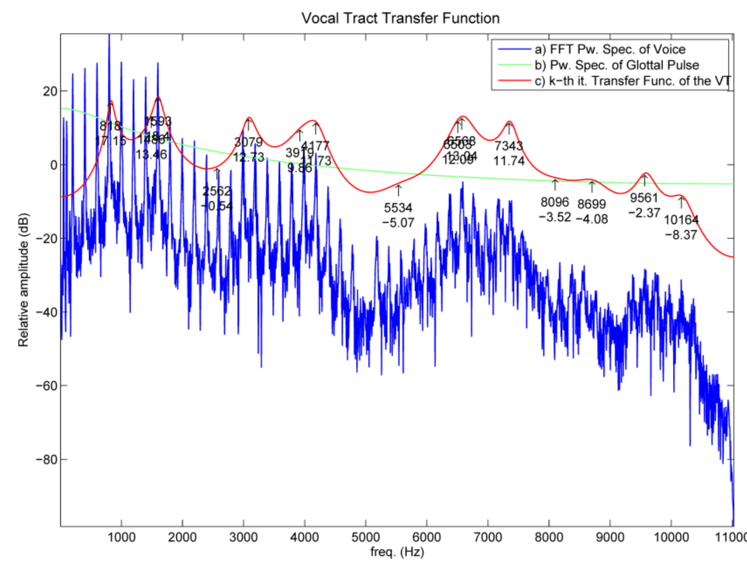
This example shows an order-1 estimation with good profiling

Prototype Female Speaker (1)



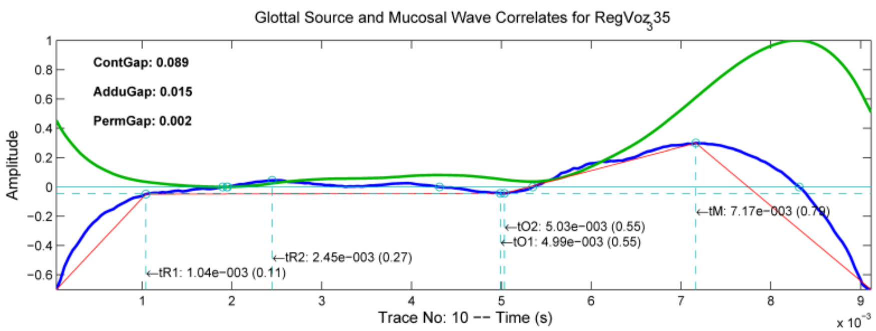


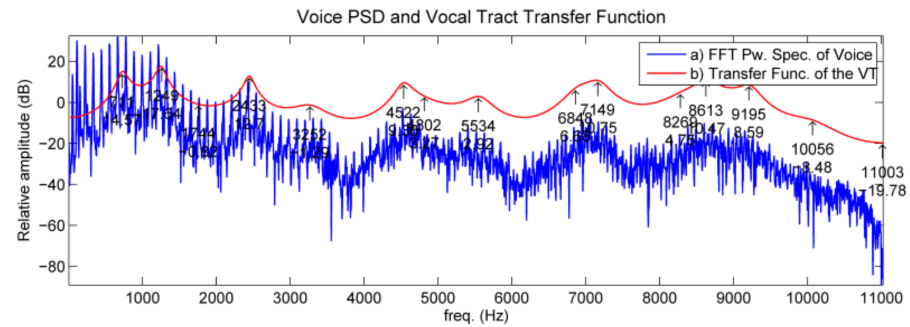


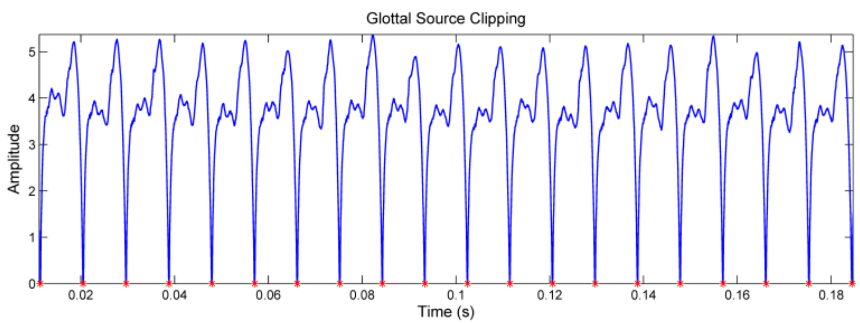


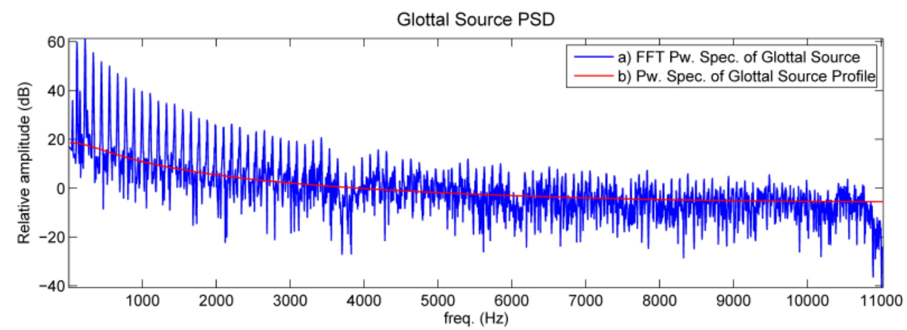
This example shows an order-1 estimation with good profiling

First Inversion Validation Results



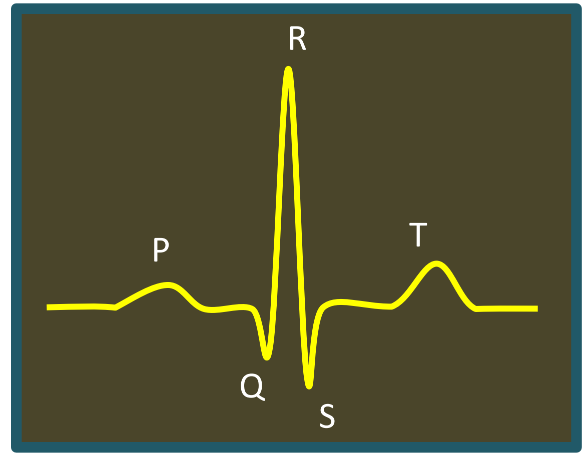


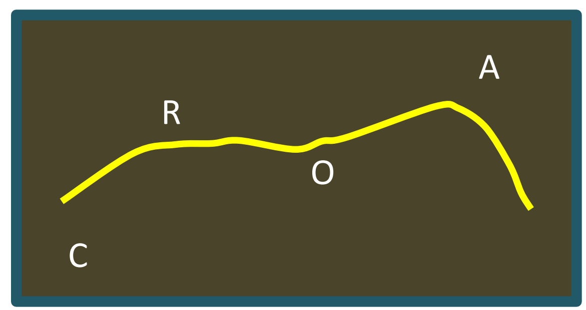




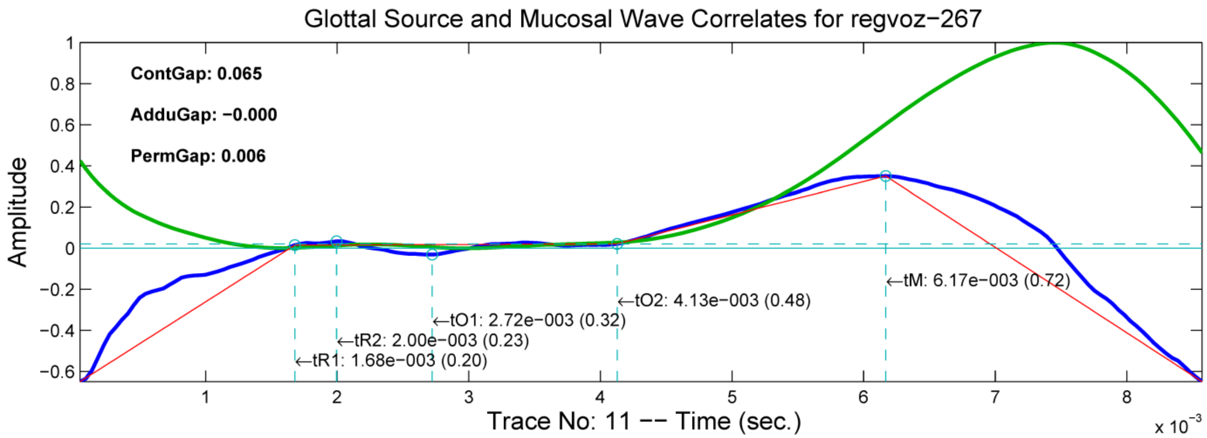
The Glottal Source

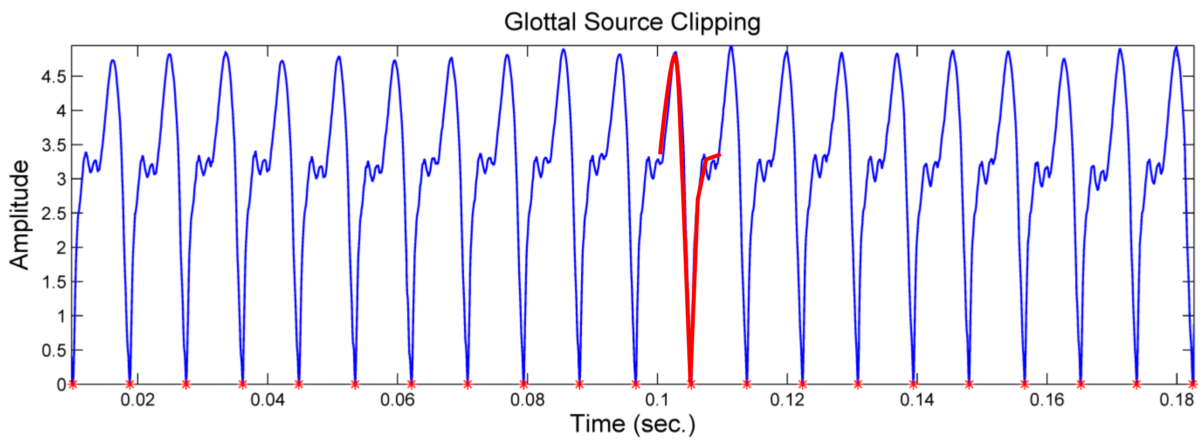
- It is a biometrical signature which can be compared in semantic power with the ElectroCardioGram
- Easy to obtain, store and match
- As the electrocardiogram describes semantic correlates in its singularities (both in time and amplitude), the glottogram points to a similar semantics in the biomechanics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



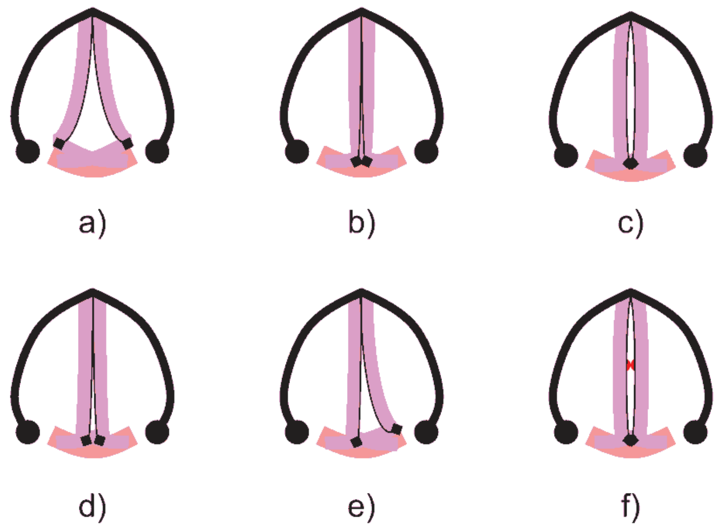


Glottal Source





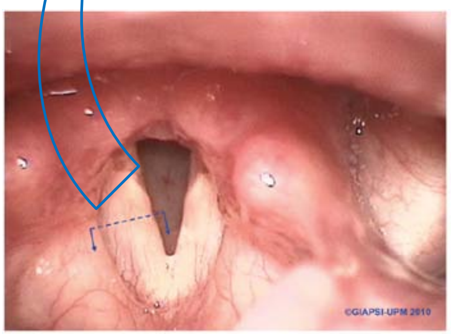
Defective Contact and Permanent Gaps

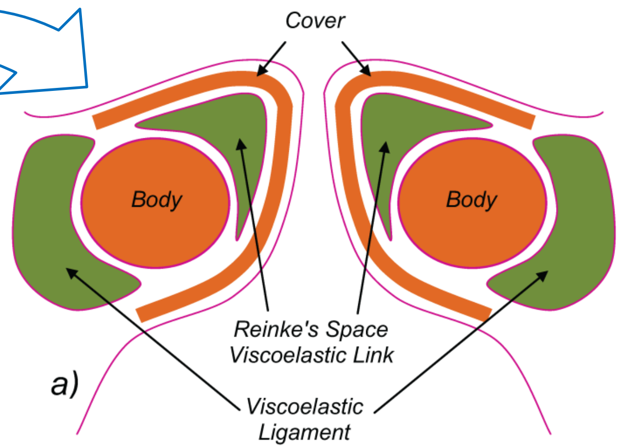


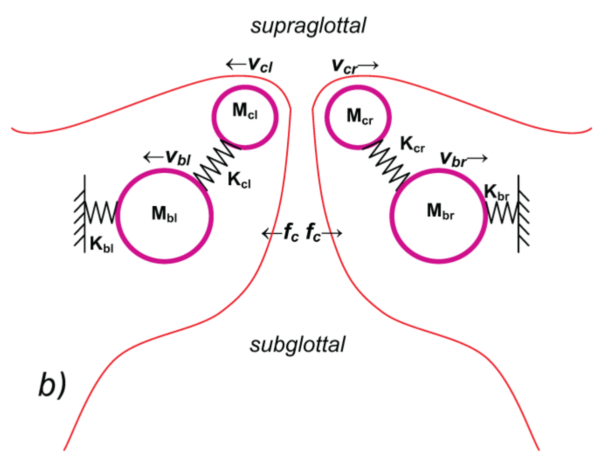
a) Inspiration/Espiration
b) Adduction
c) Abduction
d) Defective Permanent Gap
e) Defective Asymmetric Gap
f) Defective Contact Gap

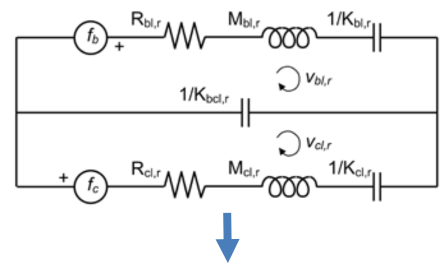
Second Problem Inversion

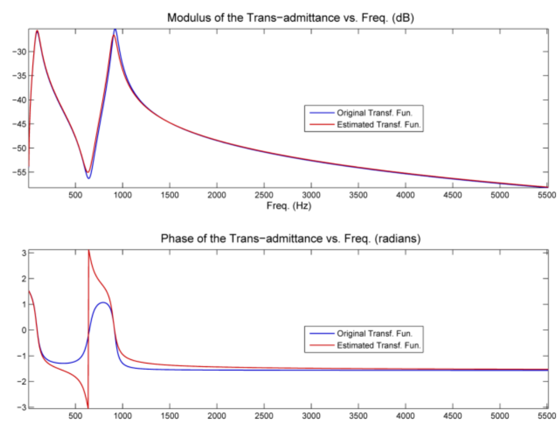
David A. Berry, Modal and nonmodal phonation, J. Phonetics, (29) 431-450, 2001



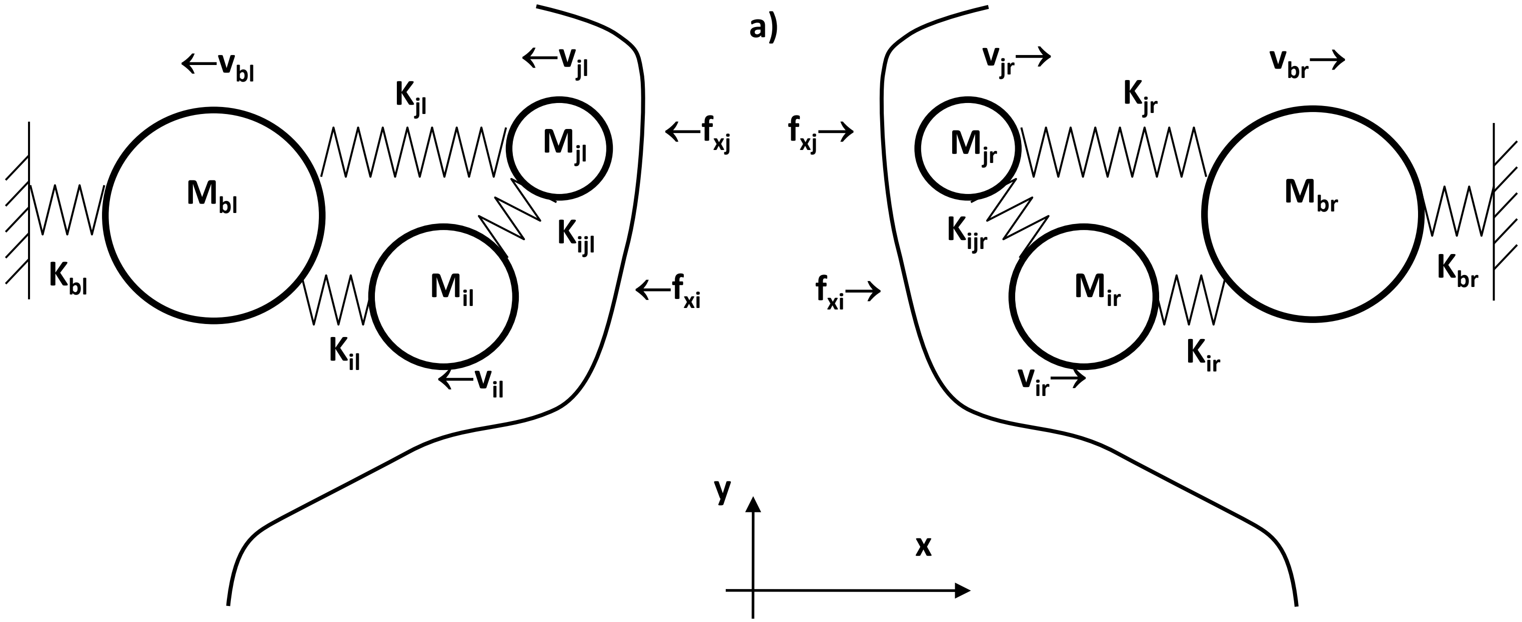






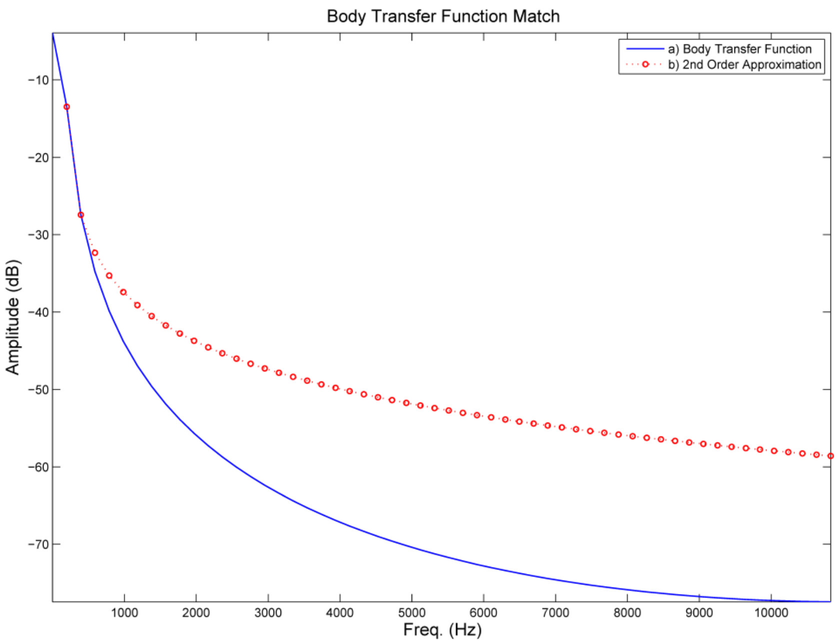


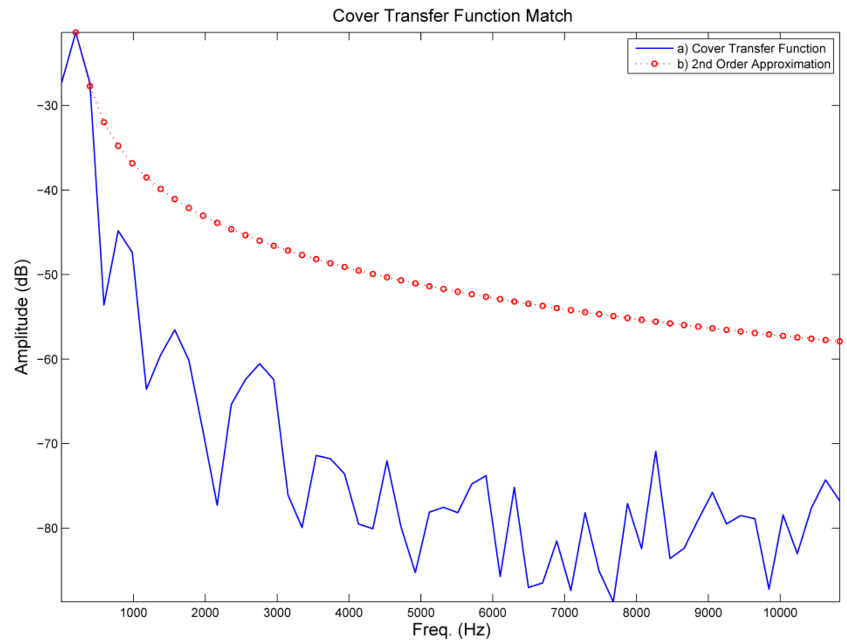
Glottal Source Biomechanical Parameterization



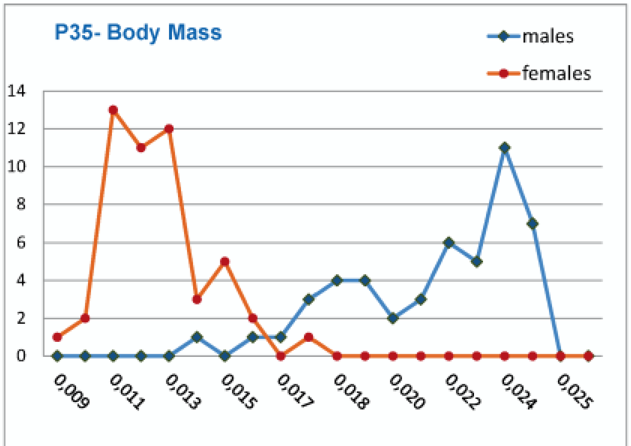
Giovanni, A., Ouaknine, M., Guelfucci, B., Yu, P., Zanaret, M. and Triglia, J. M.,a "Nonlinear Behavior of Vocal Fold Vibration: The Role of Coupling Between the Vocal Folds", *Journal of Voice*, Vol. 13, No. 4, 1999, pp. 465-476.

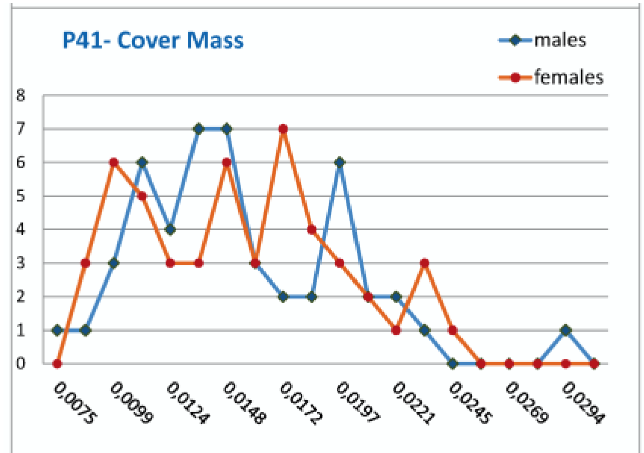
Results for a typical adjustment (body and cover): Case 346 F 34y N

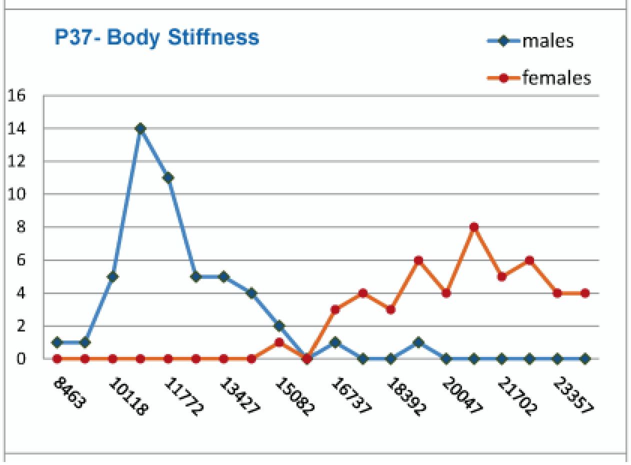


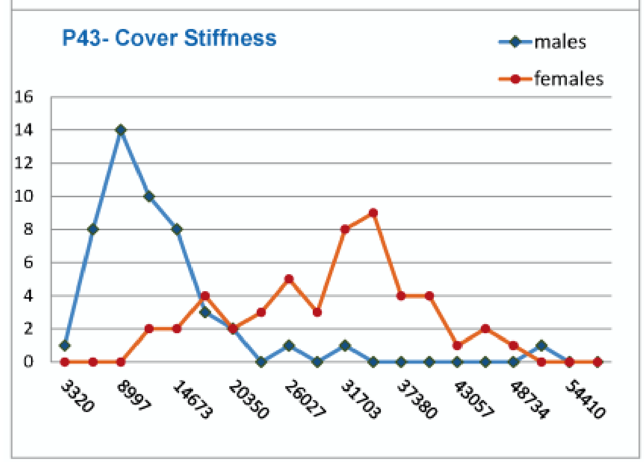


Second Inversion Validation Results









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)

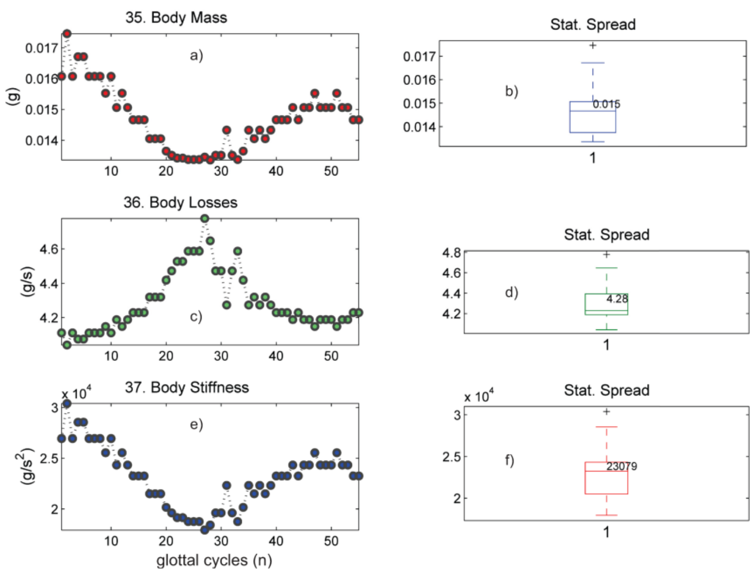
Third Model Inversion

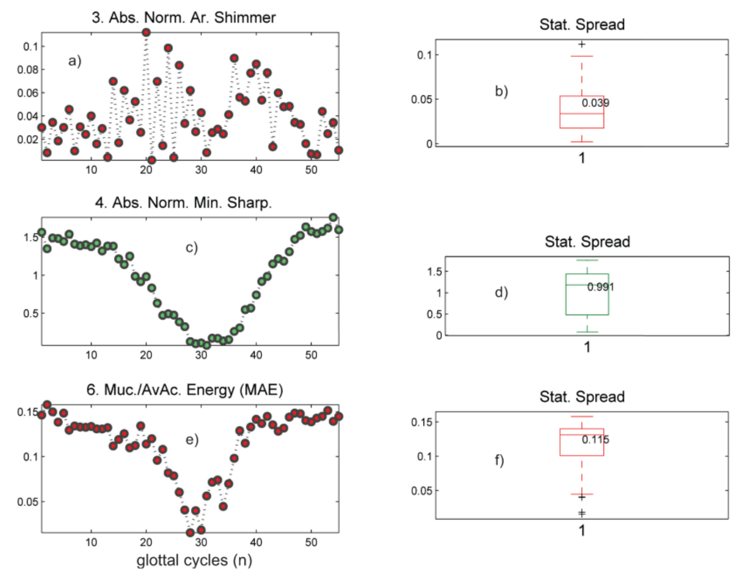
Some parameters show clear cyclicality

Q: How may we measure cyclicality?

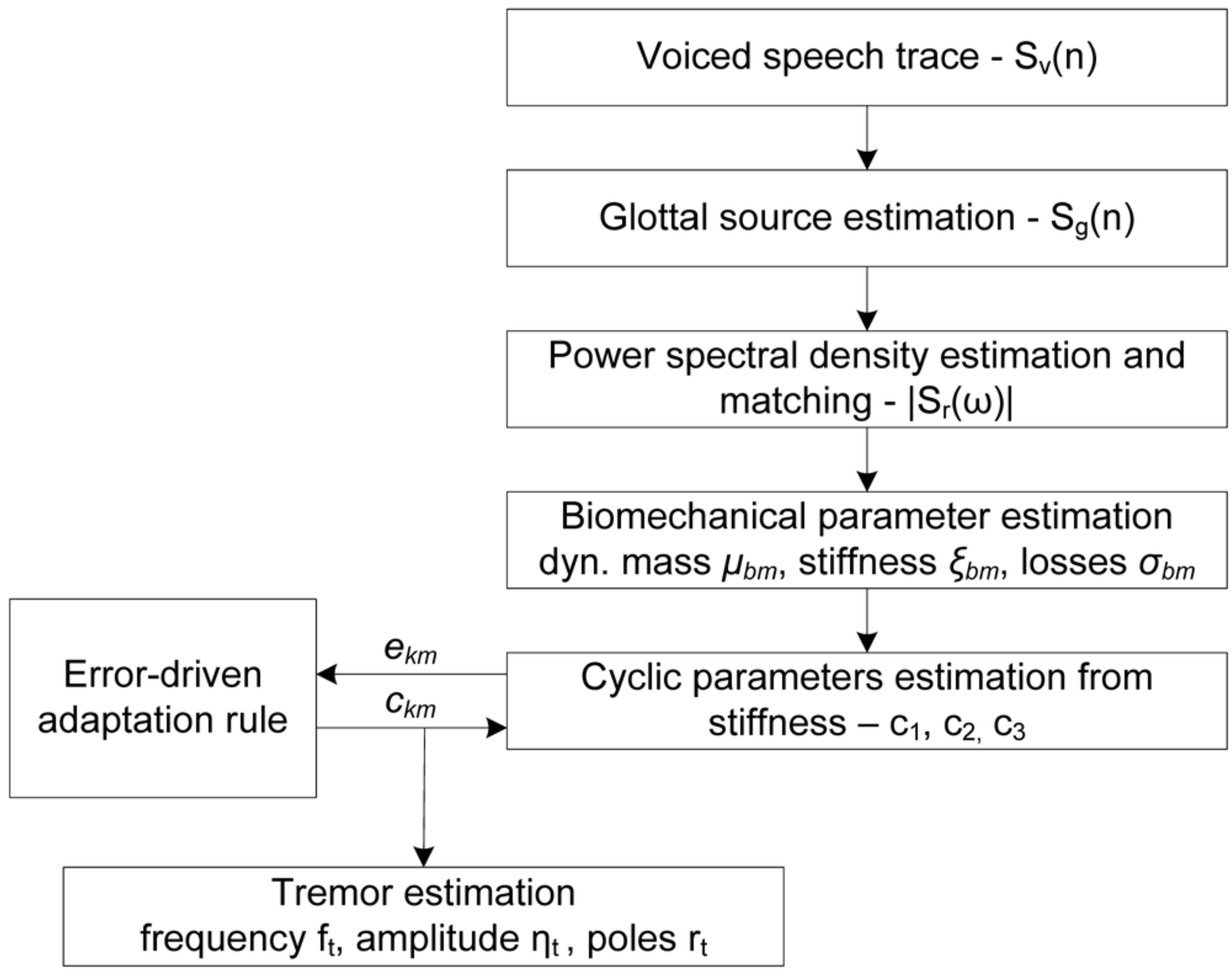
A: How about adaptive AR modelling?

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

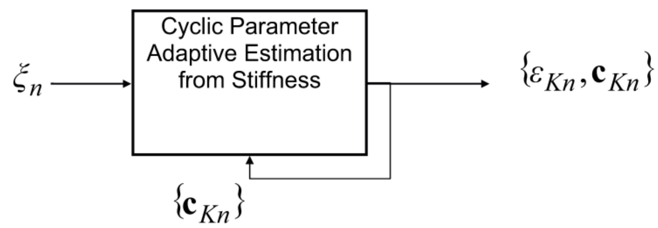




Third Inversion Process



Third Inversion Process: AR Model



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:

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

Third Inversion Process: Properties

Frequency estimate:

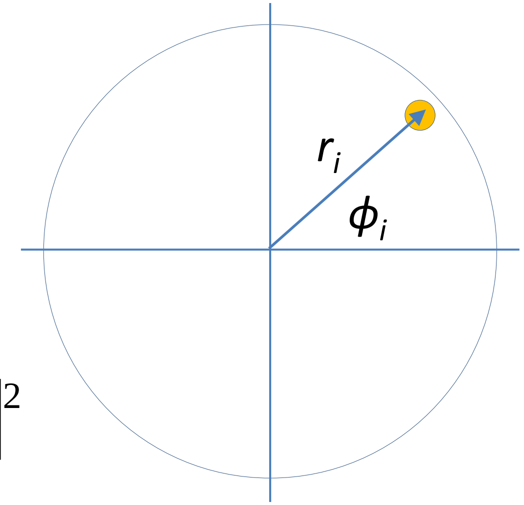
$$f_{ti} = \frac{\varphi_i}{2\pi} f_s;$$

Amplitude estimate:

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

Amplitude relevance:

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



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

$$c_1 = \frac{a_1 - a_2 a_3}{1 + a_2 - a_1 a_3 - a_3^2}; \qquad c_2 = \frac{a_2 - a_1 a_3}{1 - a_3^2}; \qquad c_3 = a_3$$

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

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

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

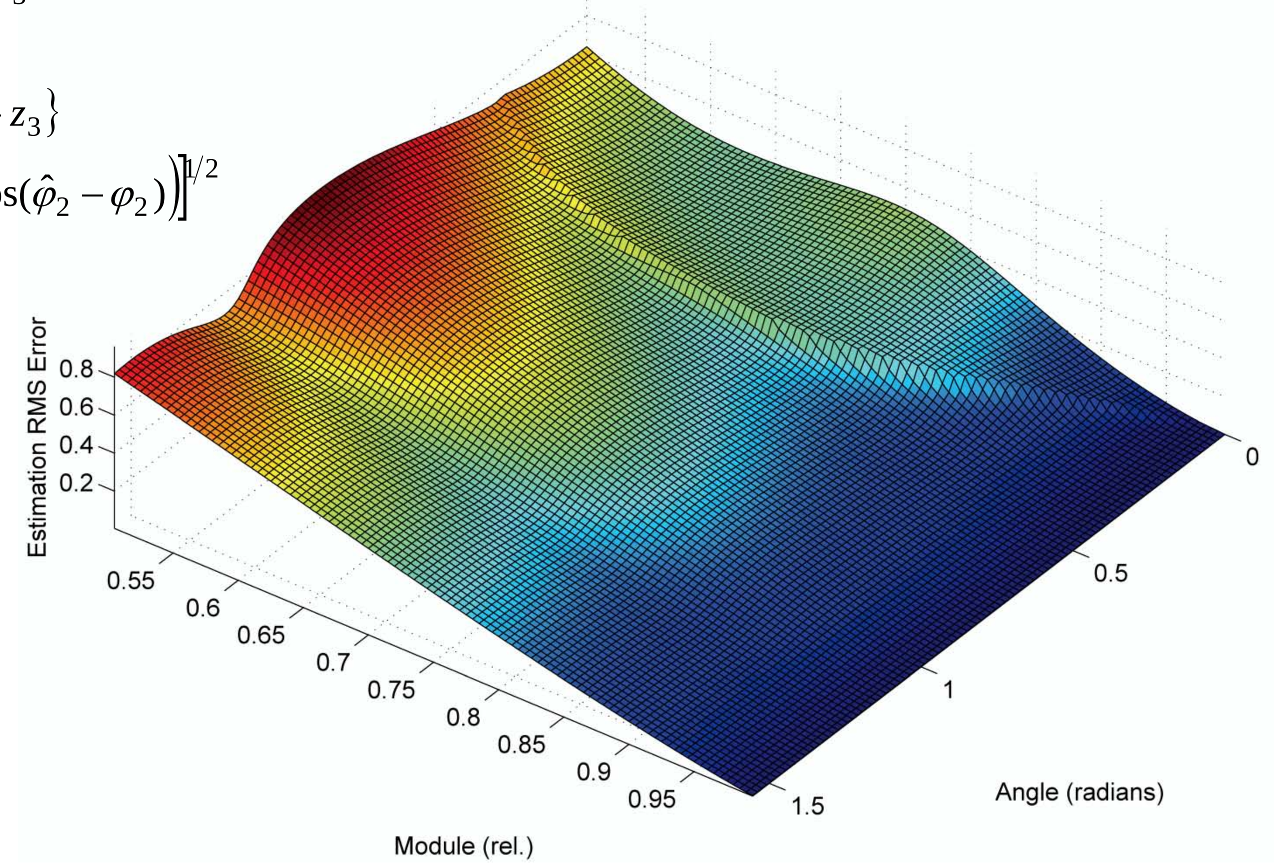
$$r_1 \in \Re; \quad r_2 = r_3 \in \Re; \quad r_2, r_3 > 0$$

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

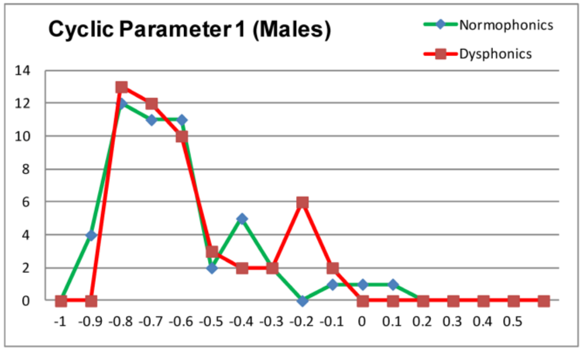
$$\varepsilon_z = \hat{\zeta} - \zeta = \{\hat{z}_2 - z_2, \hat{z}_3 - z_3\}$$

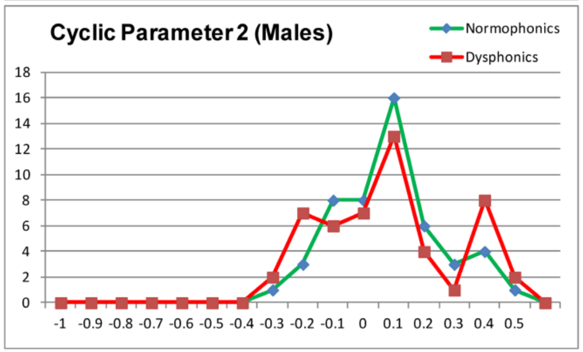
$$\|\varepsilon_z\| = \left[2\left(\hat{r}_2^2 + r_2^2 - 2\hat{r}_2 r_2 \cos(\hat{\varphi}_2 - \varphi_2)\right)\right]^{1/2}$$

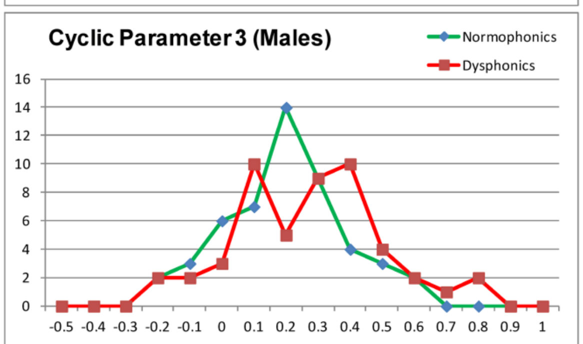
$$\varepsilon_r = \frac{\|\varepsilon_z\|}{\|\zeta\|}$$



Normalizing for a large population





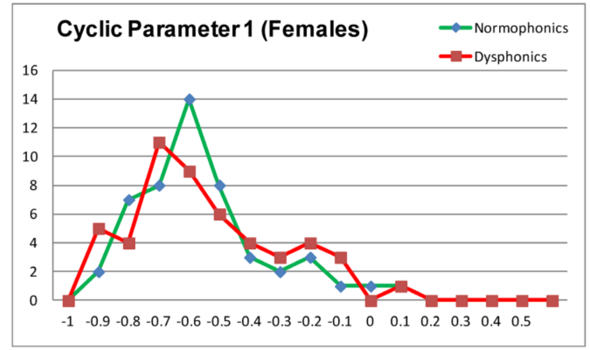


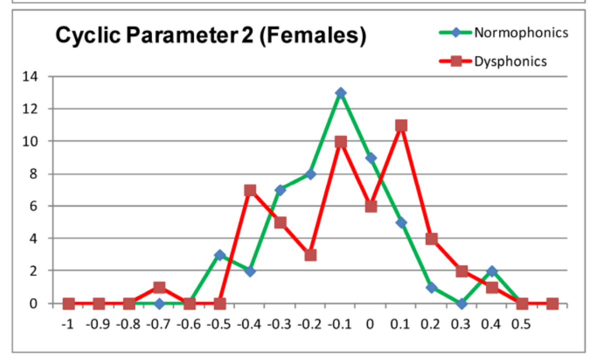
Normophonic

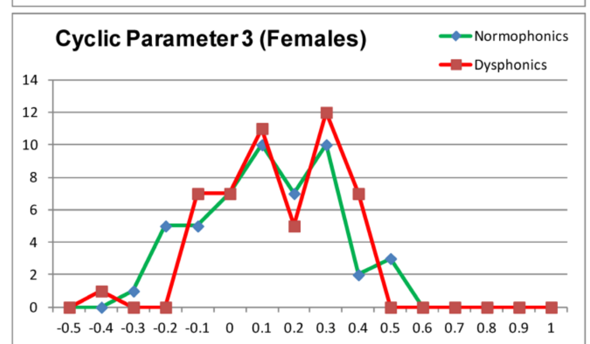
50 Males

50 Females

200 ms of vowel /a/

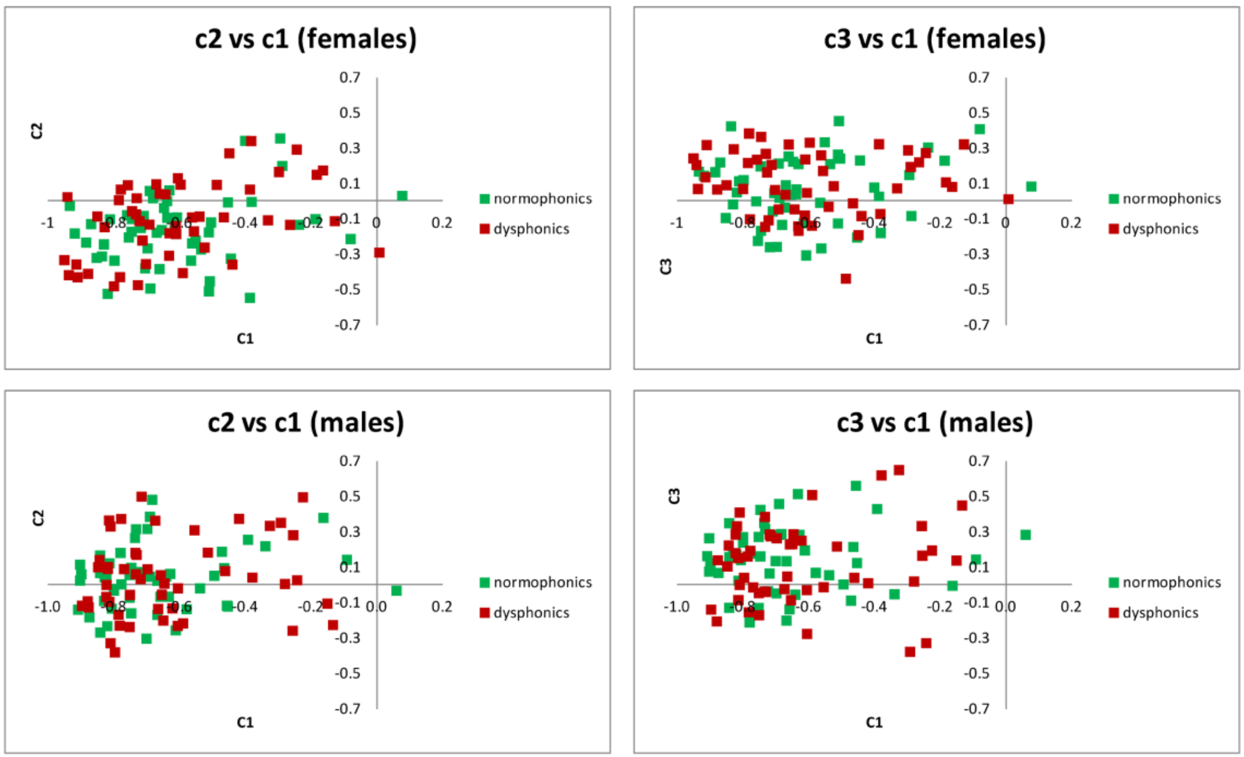






Statistical Analysis (1st level): Scatter Plots

- 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?



research group GIAPSΨ ©UPM

Statistical Analysis (1st level): Dist. Quartiles

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

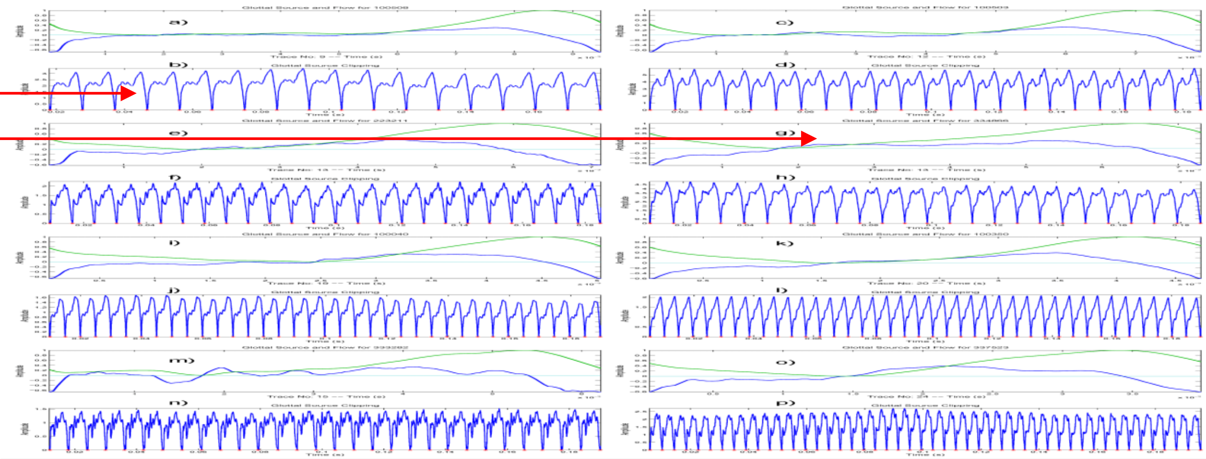
DISTRIBUTION STATISTICAL DESCRIPTION. Q: QUARTILES 1-3; (M/F): RESPECTIVE GENDER.

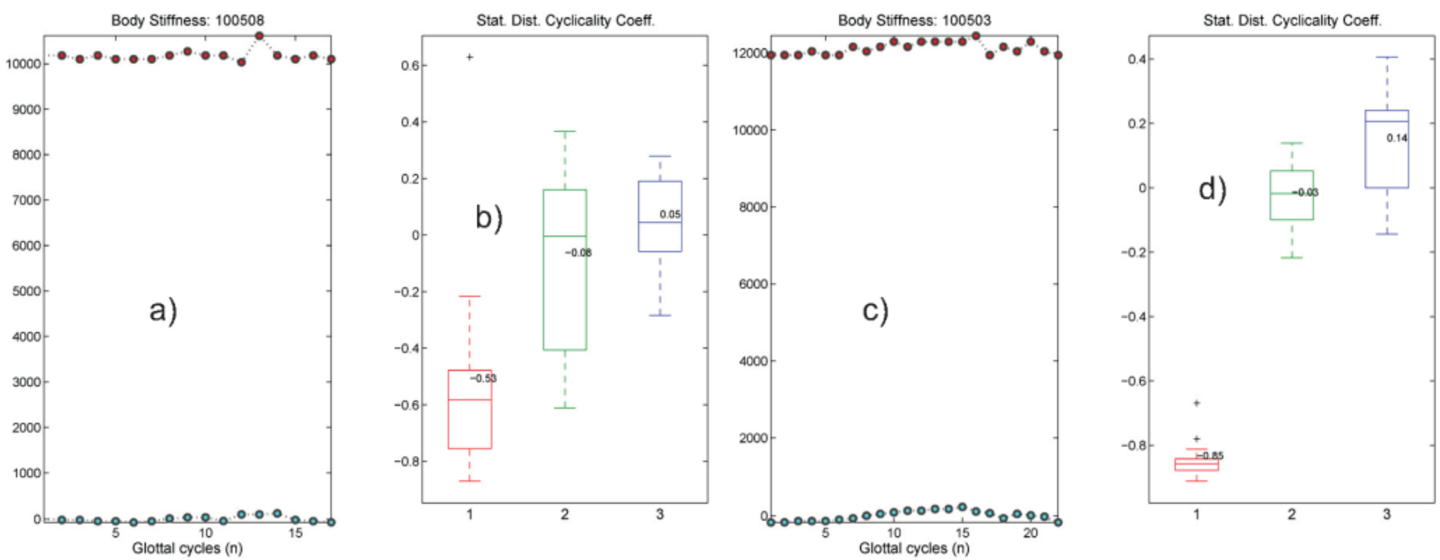
Qart./Par.	Q_1 (M)	Q_2 (M)	Q_3 (M)	Q_1 (F)	Q_2 (F)	Q_3 (F)
c_1 (rel)	-0.977	-0.962	-0.930	-0.984	-0.970	-0.934
c_2 (rel)	-0.016	0.303	0.463	-0.238	-0.007	0.149
c_3 (rel)	0.145	0.274	0.349	0.031	0.220	0.310
f_t (Hz)	2.716	3.819	4.902	4.492	6.002	8.909
r_t (rel)	0.898	0.935	0.953	0.884	0.914	0.947
η_t (rel)	0.006	0.010	0.013	0.007	0.010	0.016
ξ_b (N/m)	10.428	11.162	12.861	18.739	20.580	22.425
ξ_c (N/m)	6.856	9.651	13.922	20.490	29.582	35.043
$\Delta\xi_b$ (rel)	0.009	0.012	0.017	0.012	0.018	0.023
$\Delta\xi_c$ (rel)	0.028	0.040	0.060	0.037	0.050	0.072
Σ_b (N/m)	0.154	0.196	0.279	0.265	0.386	0.641
Σ_c (N/m)	0.330	0.564	1.248	1.355	1.891	3.171

- 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)

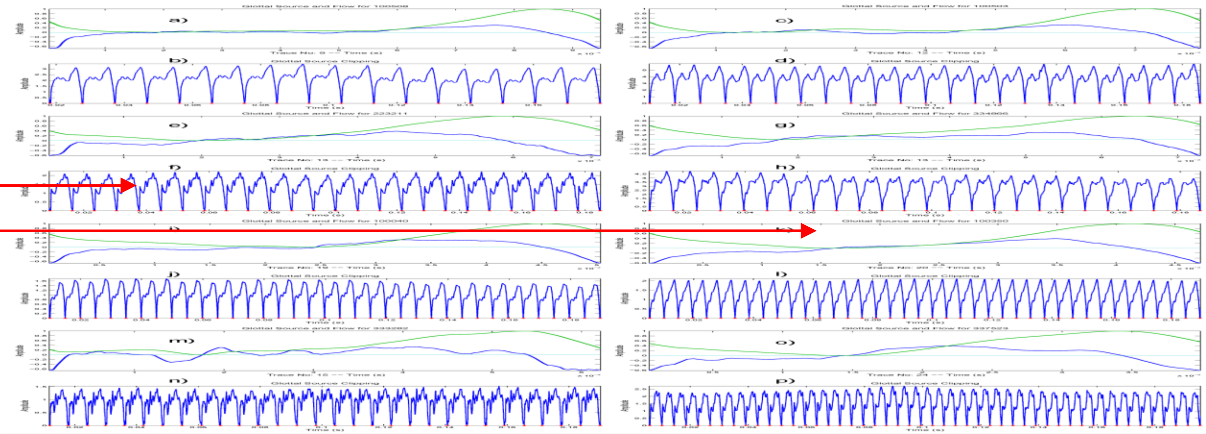
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

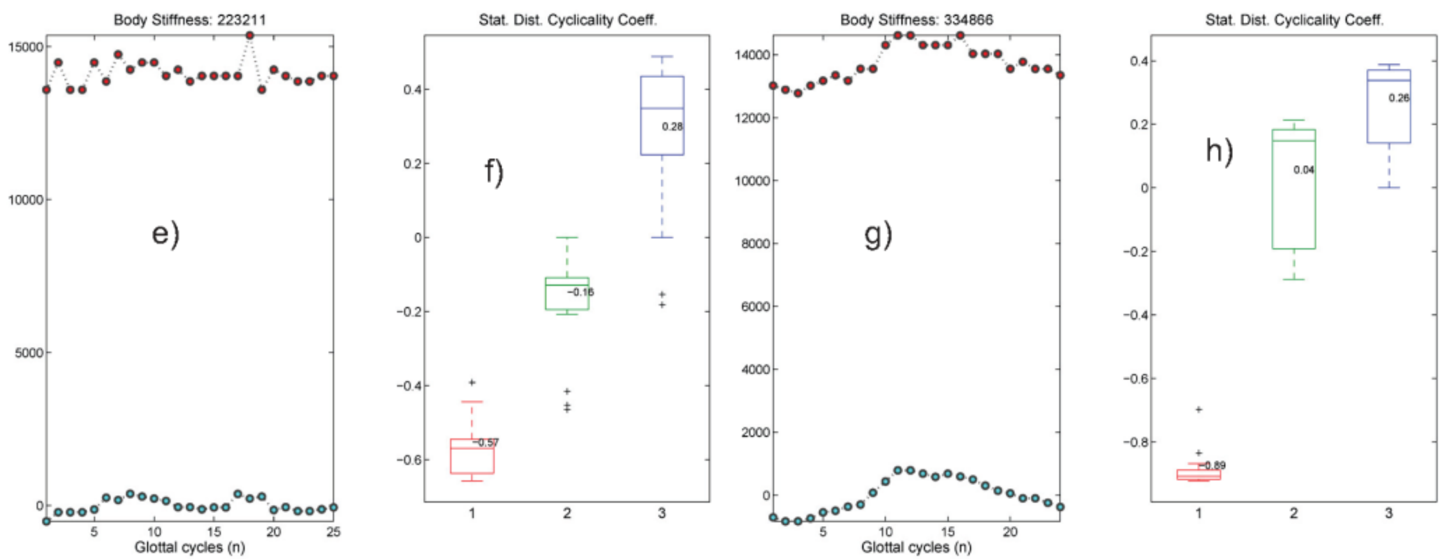




Male PD (no tremor and tremor)

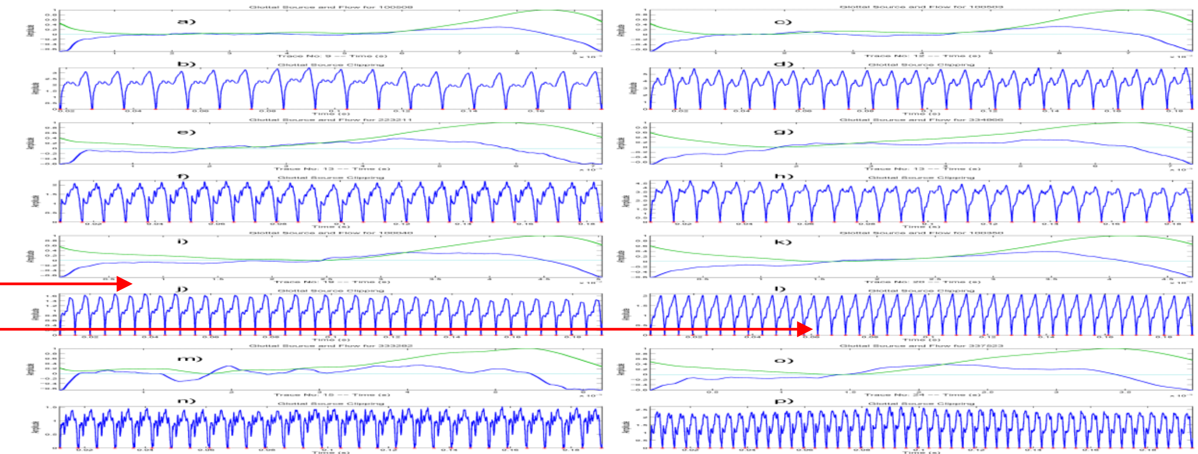
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

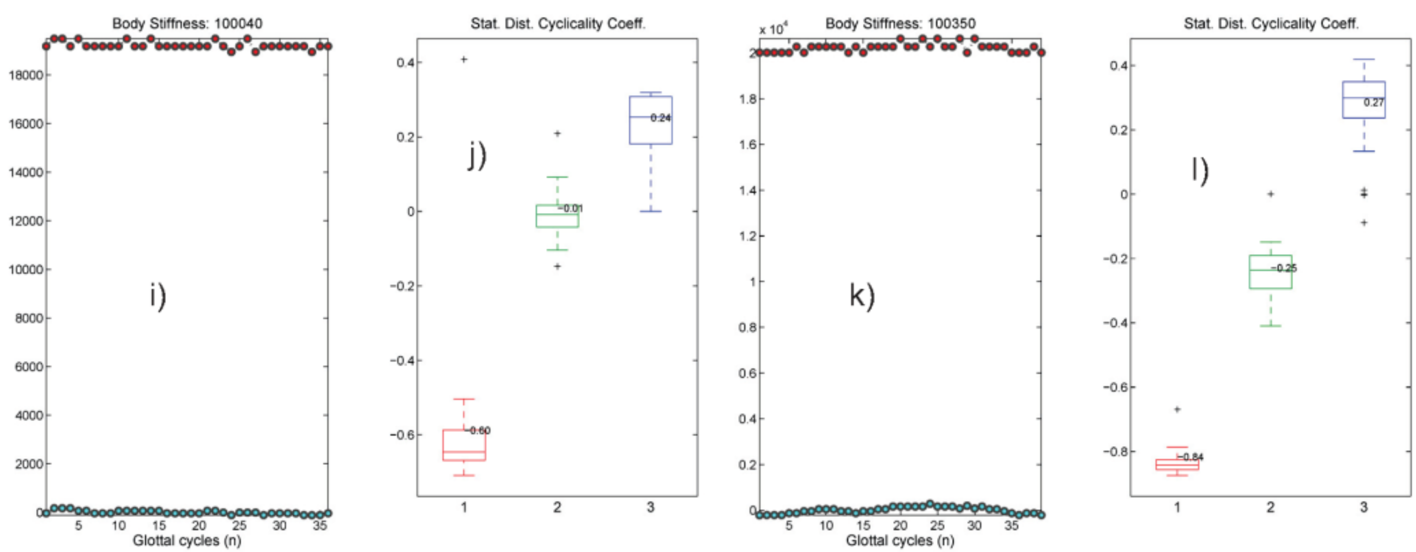




Female Normals (no tremor and tremor)

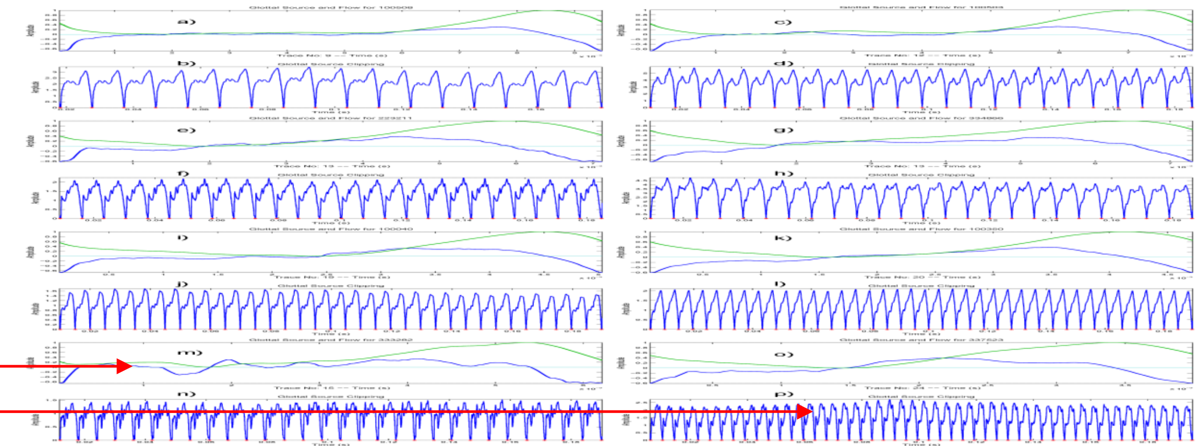
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

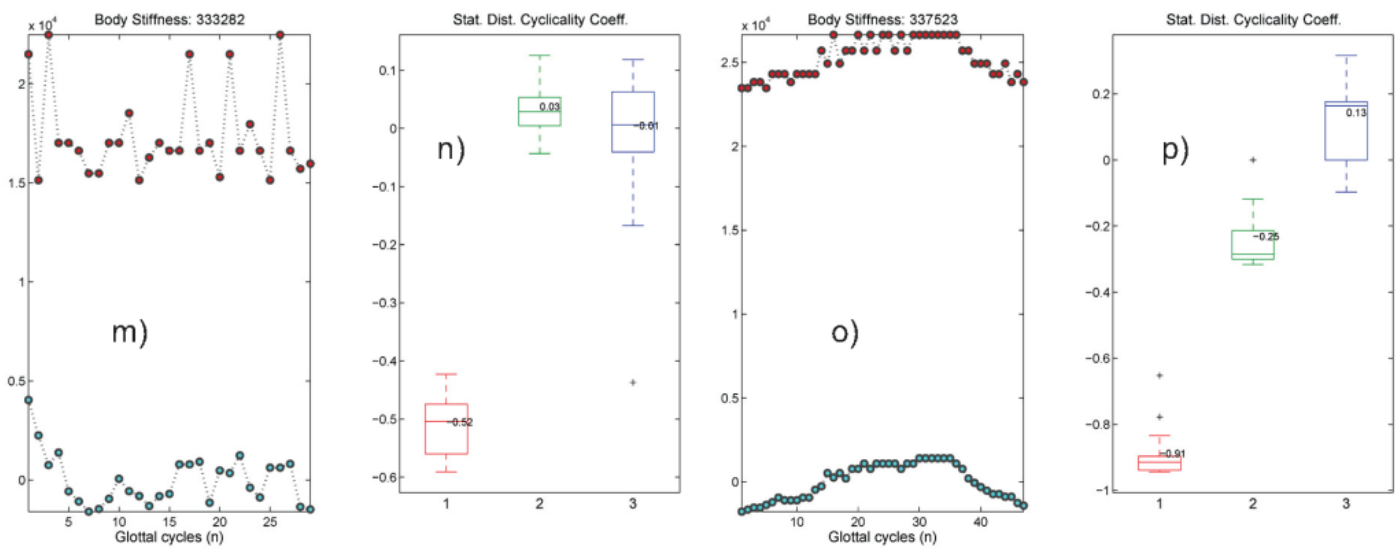




Female PD (no tremor and tremor)

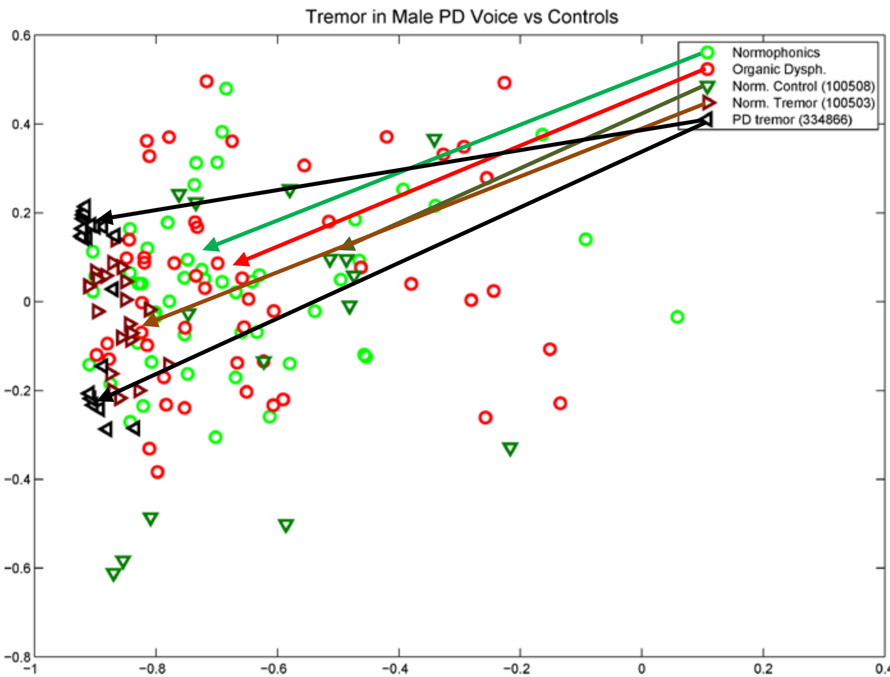
Case	Gender	Cond.	Tremor	Figs.
100508	Male	Norm.	No	a-b
100503	Male	Norm.	Yes	c-d
223211	Male	Par. Dis.	No	e-f
334866	Male	Par. Dis.	Yes	g-h
100040	Female	Norm.	No	i-j
100350	Female	Norm.	Yes	k-l
333282	Female	Par. Dis.	No	m-n
337523	Female	Par. Dis.	Yes	o-p

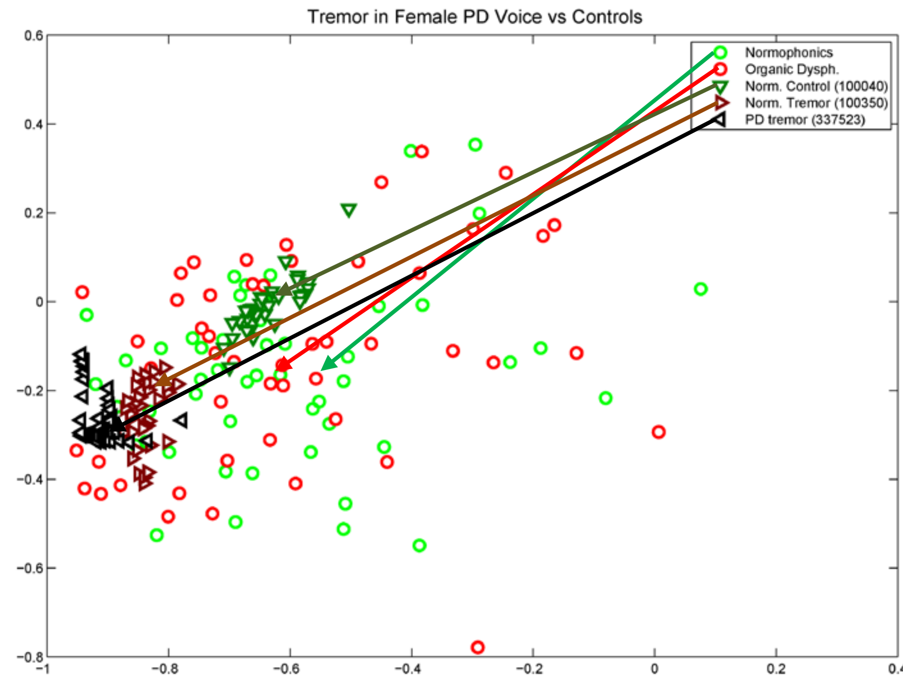




Comparing Results

Intra- and inter-subject scatter plots ($c_2$vsc_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:

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

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

$$\mathbf{C}_{\mathbf{m}} = \mathbf{X}_{\mathbf{sm}}^{\mathbf{T}} \mathbf{X}_{\mathbf{sm}} ; \quad \mathbf{C}_{\mathbf{f}} = \mathbf{X}_{\mathbf{sf}}^{\mathbf{T}} \mathbf{X}_{\mathbf{sf}}$$

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

$$\chi_{\mathbf{m}} = \mathbf{E} \langle \mathbf{X}_{\mathbf{sm}} \rangle_{\mathbf{s}} ; \quad \chi_{\mathbf{f}} = \mathbf{E} \langle \mathbf{X}_{\mathbf{sf}} \rangle_{\mathbf{s}}$$

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

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

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

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:

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

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

Detection Results combining stiffness and tremor

Cyclicality parameters and likelihood ratios															
Case	G	Cond.	Trem.	f_t	η_t	μ_b	μ_c	c_1	σ_{c1}	c_2	σ_{c2}	c_3	σ_{c3}	λ_{c1}	λ_{T1}
100508	Male	Norm.	No	17,55	0,0049	10,180	5,393	-0,6	0,19	-0,08	0,33	0,05	0,15	**0.48**	**-0,44**
100503	Male	Norm.	Yes	5,59	0,009	12,134	7,010	-0,85	0,03	-0,03	0,1	0,14	0,16	**-1.53**	**-0,28**
223211	Male	Par. Dis.	No	10,78	0,0136	14,145	19,498	-0,57	0,05	-0,16	0,11	0,28	0,19	**0.70**	**-13,9**
334866	Male	Par. Dis.	Yes	5,39	0,0342	13,777	14,498	-0,89	0,02	0,04	0,19	0,26	0,13	**-1.88**	**-6,03**
100040	Female	Norm.	No	10,66	0,0034	19,227	14,023	-0,63	0,04	0	0,06	0,24	0,07	**-0.21**	**-0,03**
100350	Female	Norm.	Yes	7,49	0,006	20,247	20,373	-0,84	0,02	-0,25	0,07	0,27	0,12	**-1.82**	**-0,45**
333282	Female	Par. Dis.	No	16,34	0,0597	17,276	12,815	-0,52	0,04	0,03	0,04	-0,01	0,11	**1.19**	**-11,4**
337523	Female	Par. Dis.	Yes	5,15	0,0383	25,314	30,274	-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

Other lines

- Other ND alterations in voice (AD, LAS, HC, etc.)
- Combination of larynx biomechanical, diaprhagmal, nasopharyngeal, and mandibular, having in mind that evidence shows that:
  - Larynx biomechanics alters f_t and rMSA (tremor frequency and amp.)
  - Diaprhagm 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 esential 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

How far may we go?

- **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

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