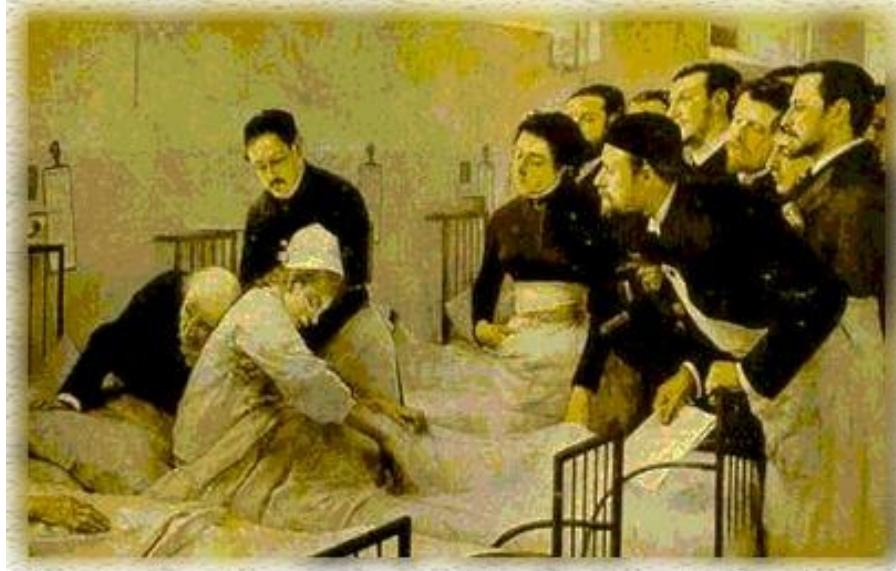




2nd SPLab Workshop 2012

Brno , October 24 to October 26 2012



Acoustic Analysis for the Clinical Evaluation of the phonatory system

Jesús B. Alonso Hernández (Ph.D.)
jalonso@dsc.ulpgc.es



Origin and Motivation

Division of Digital Signal Processing

Institute for Technology Development and Innovation in Communications
University of Las Palmas de Gran Canaria (Spain)

- Teaching experience in the field of **digital signal processing**
- Research Experience in the field of **speakers Recognition Systems**
- Since 1998 we have been collaborating with the Department of Otolaryngology of General Hospital of GC "Dr. Negrín" (Spain):
 - The assessment of patiences with laryngeal pathologies



Objectives of the research line

- ✓ Study of the **manifestations of laryngeal pathologies in the phonatory system.**
- ✓ Study the **qualitative and quantitative assessment of the quality of the voice from a speak recording**, through different domains of representation (temporal, spectral, cepstral, non-linear domains, etc).
- ✓ Design a software tool for the **assessment of the phonatory system** using the speech signal for **clinical use**.





The assessment of the phonatory system



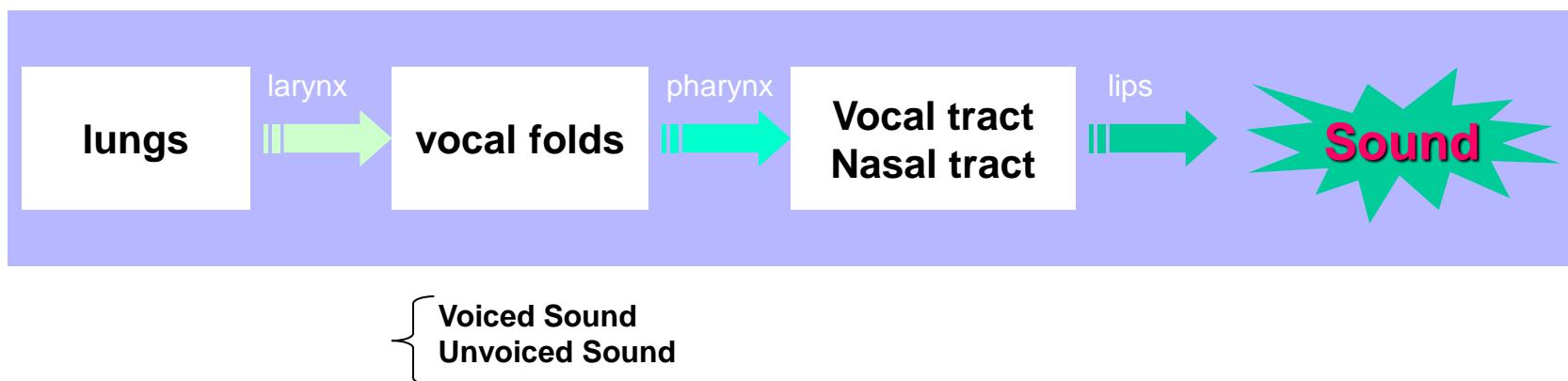
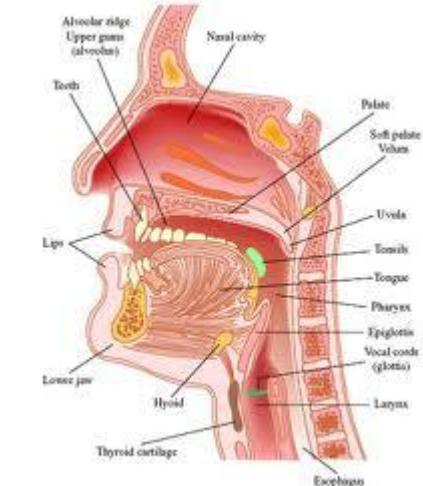
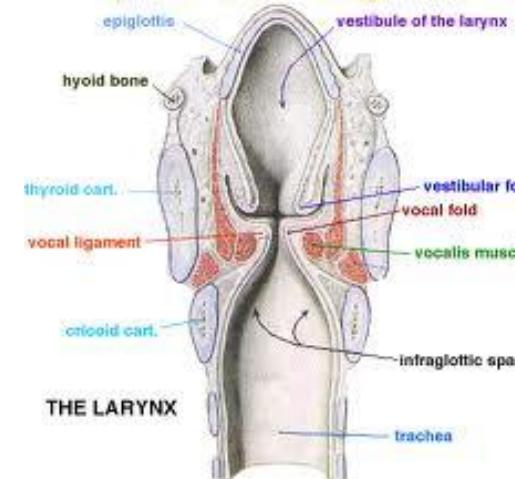
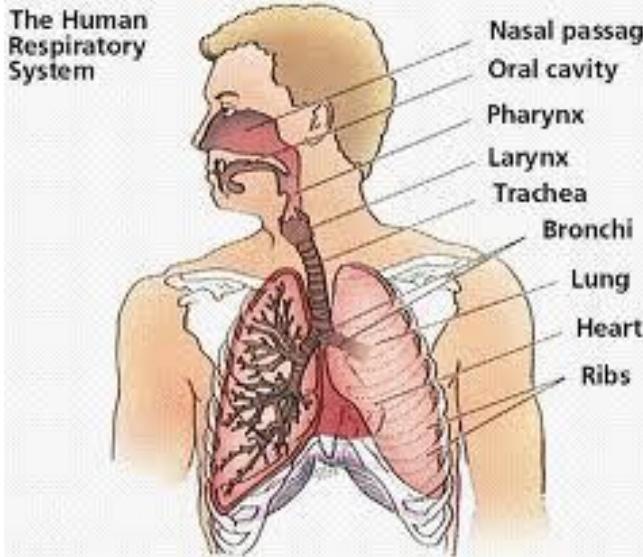
UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

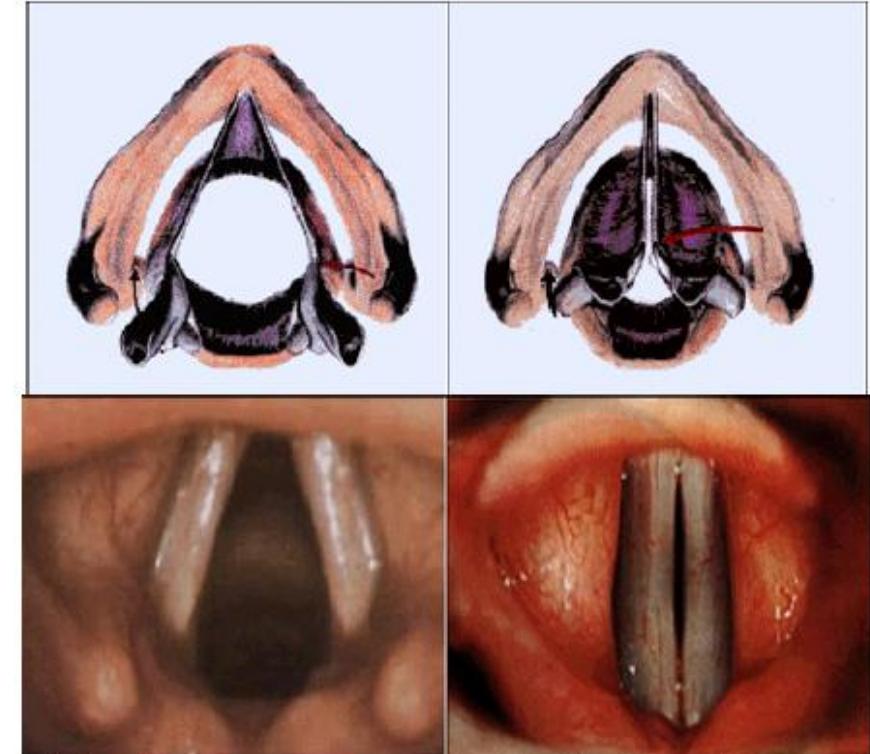
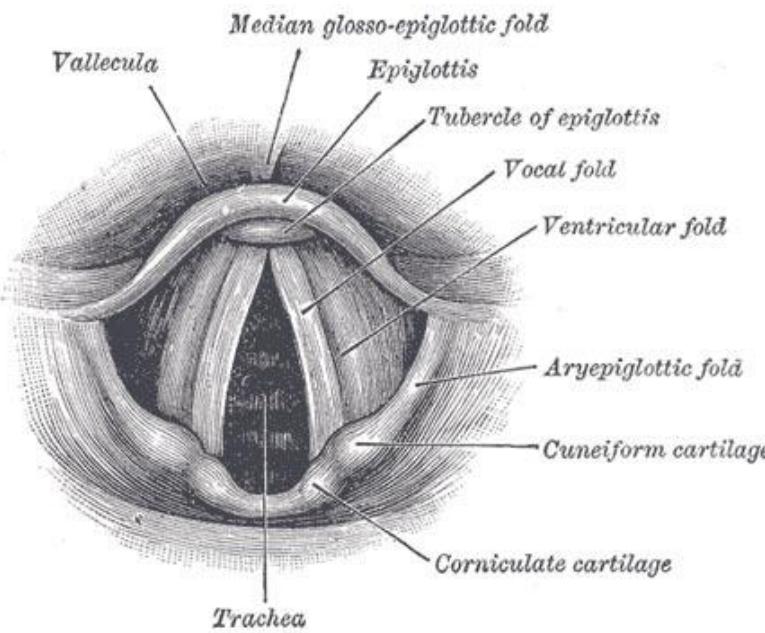
Jesús B. Alonso Hernández



Voice Production System: Physiology



Voice Production System: Physiology



Abduction

Adduction



Voice Production System: Physiology

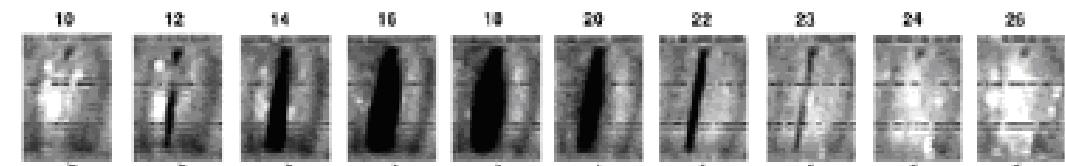
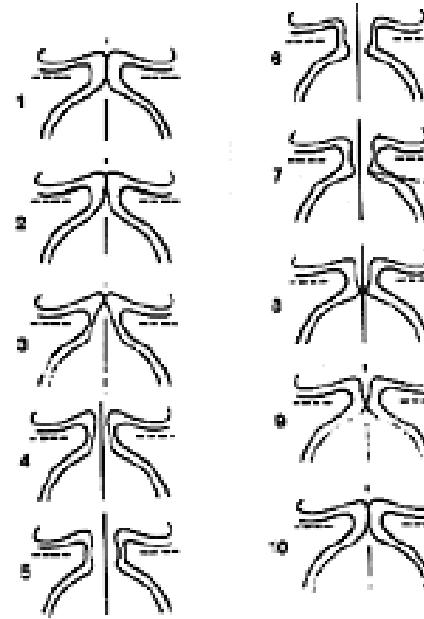
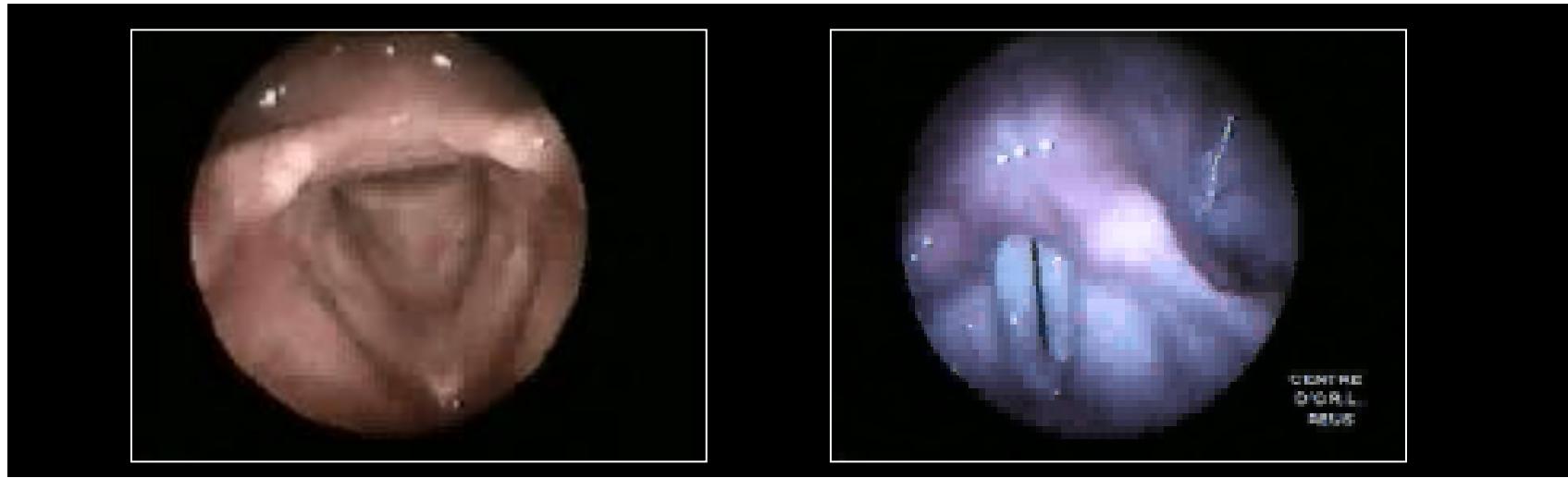
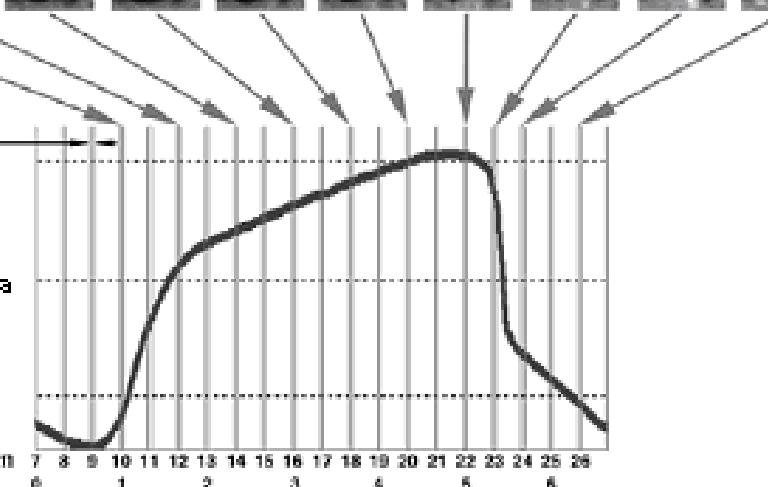


Imagen 9

0.1 ms

Aplitude de la
Abducción

Número de imagen
Tiempo (ms)



Changes in the quality of the voice: dysphonia

dysphonia:

*"Alteration of one or more of the acoustic characteristics of the voice:
intensity, pitch and spectral richness"*

- **Intensity** is defined as the amplitude of the sound pressure level corresponding to a sound phonation.
- The **pitch** is determined by the frequency of vibration of the vocal folds.
- The **spectral richness** is associated with the energy distribution in the spectrum.



Changes in the quality of the voice: Causes

The dysphonia are mainly caused by:

- ✓ **The presence of pathology in the phonatory system :**
 - Organ damage (such as malformations, acquired benign lesions, inflammation, infection, precancerous lesions and cancer, trauma, or neurological, endocrine and audiology)
 - Functional dysphonia (both spoken voice and singing voice)
 - Psychiatric origin
- ✓ **Changes in mood:**
 - *Emotional States*
 - *Neurodegenerative diseases such as Alzheimer ??*



Changes in the quality of the voice

Examples of pathological vocal folds

Kissing nodules

NODULES

DE

CORDES VOCALS

CENTRE
GYRAL
SEUS

Hyperfunction



CENTRE
GYRAL
SEUS

Polyp



Vocal cord paralysis



CENTRE
GYRAL
SEUS



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno ,October 24 to October 26 2012

Jesús B. Alonso Hernández

Methods for assessment the phonatory system

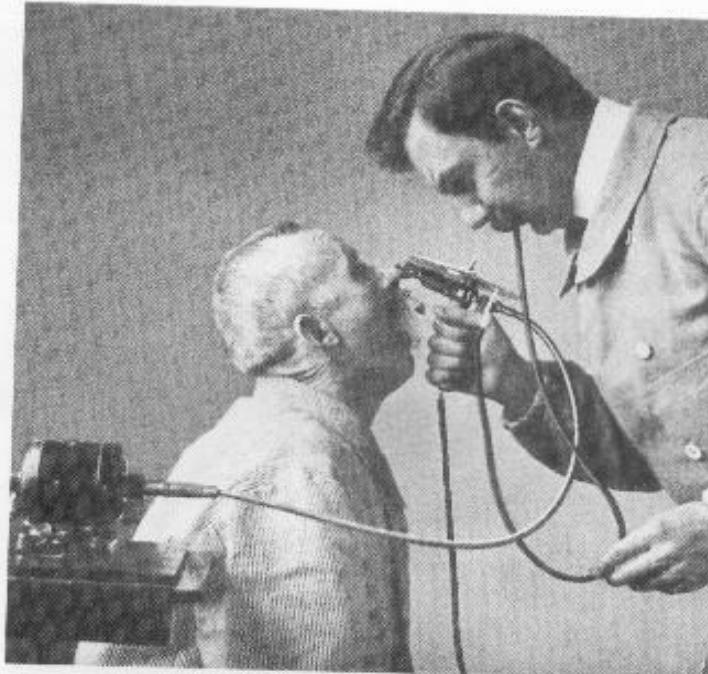


FIG. 9.—Estroboscopia directa bajo anestesia local. Brünings, 1910.



FIG. 16.—Brünings realizando cirugía a dos manos bajo anestesia general y telescopio de magnificación.

Methods for assessment the phonatory system

Classification of the main methods to study the phonatory system

- Direct visual methods
- Indirect visual methods
- Direct methods
- Indirect methods
- Extraction of the glottal flow pressure
- Hearing-based techniques



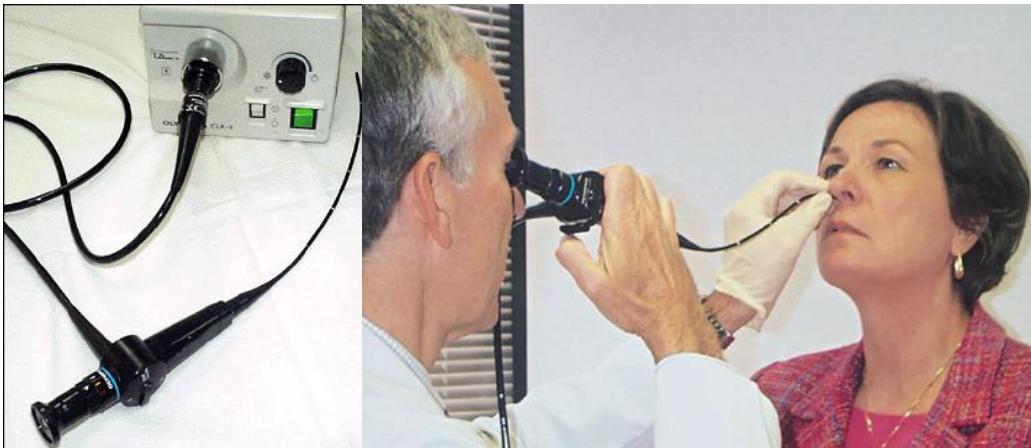
Methods for assessment the phonatory system

Direct visual methods

Rigid laryngoscope (through mouth)



Flexible laryngoscope (through nose)



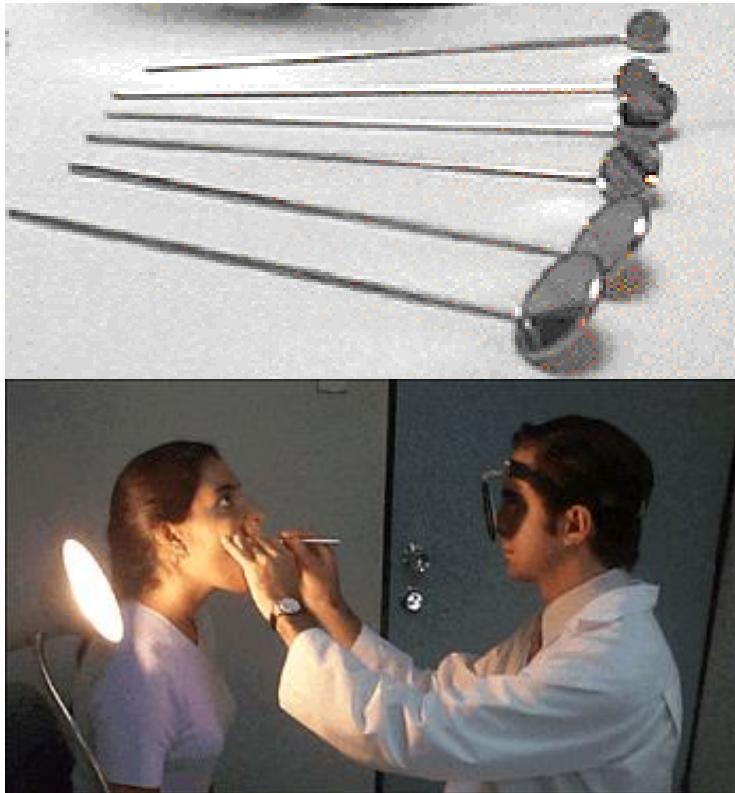
Laryngoscopic exploration



Methods for assessment the phonatory system

Indirect visual methods

Reflects conventional laryngoscopy



Microlaryngoscopy reflects



Methods for assessment the phonatory system

Direct methods and Indirect methods

Direct Methods

Transillumination (located in the throat)

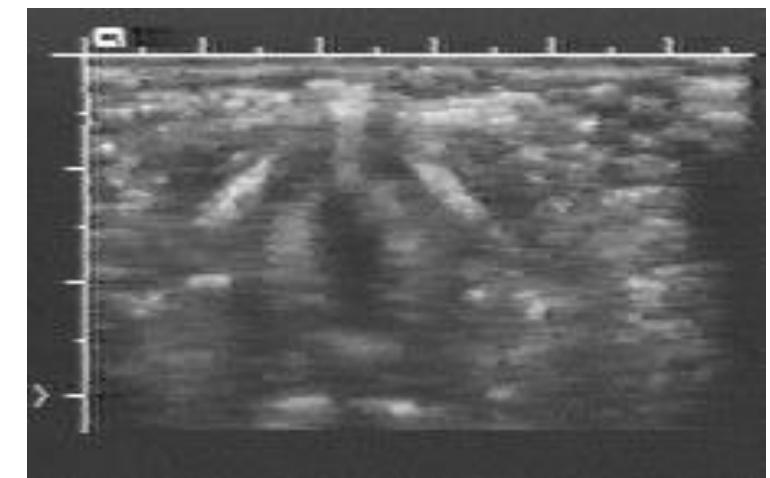


Indirect Methods

Tomography

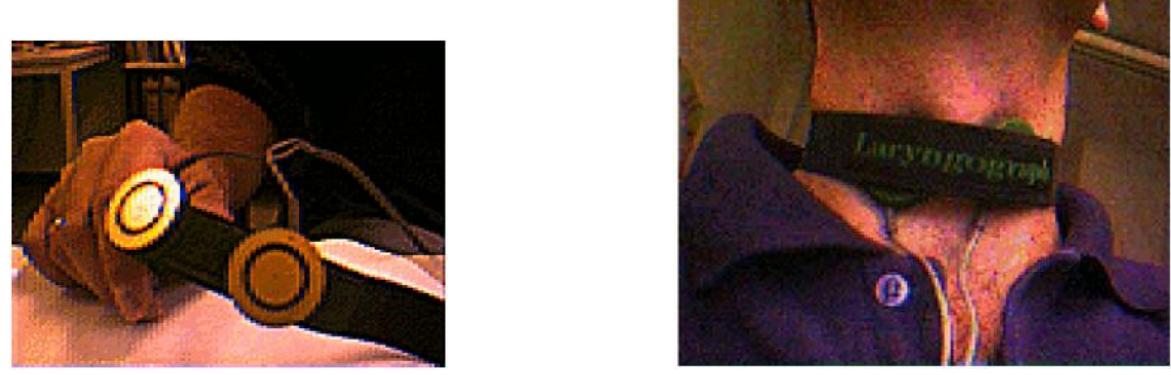


Ultrasound



Methods for assessment the phonatory system

Extraction of the glottal flow pressure

Direct measures	<ul style="list-style-type: none">• Electroglottography 
Indirect measures	<ul style="list-style-type: none">• Microphone• Rothenberg Mask 



Methods for assessment the phonatory system

Hearing-based methods

Subjective measures

Based on measuring the quality of the voice through the subjective perception of clinician

[Hirano, 1981] → scale **GRABS**

[Laver, J., 1991]

[Hammarberg, B. & Gauffin, J., 1995]

[Dejonckere, 1996]

...

Objective measures

Based on automatically measure of the quality of the voice from a speaker recording using a computer system



Hearing-based methods – Objective Techniques

- Commercial Products for the Qualitative Analysis
 - **SoundScope** (*CM Instrument*)
 - **Cspeech**
 - **Dr Speech** (*Tiger Elemetric*)
 - **SSVA** (*System for Sigle Voice Analysis*)
 - **MDVP** (*Multi-Dimensional Voice Program*) *Kay Elemetric*
 - **EVA** (*Evaluation Vocal Assistee*)
 - **CSL** (*Computerized Speech Laboratory – Speech Technology Research (STR)*)
 - **PRAAT**
 - **VISHA**
 - **CSRE** (*Computerized Speech Research Environment*)
 - **SpeechViewer**
 - **MEDIVOZ**
 - ...
- Commercial Products for the Quantitative Analysis
 - **MDVP** (*Multi-Dimensional Voice Program*) dofe *Kay Elemetric* (33 measures)
 - **Dr Speech** (*Tiger Elemetric*)
 - **MEDIVOZ**
 - **Differents research lines**





Laryngeal manifestations of the pathologies in the voice signal



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

Jesús B. Alonso Hernández

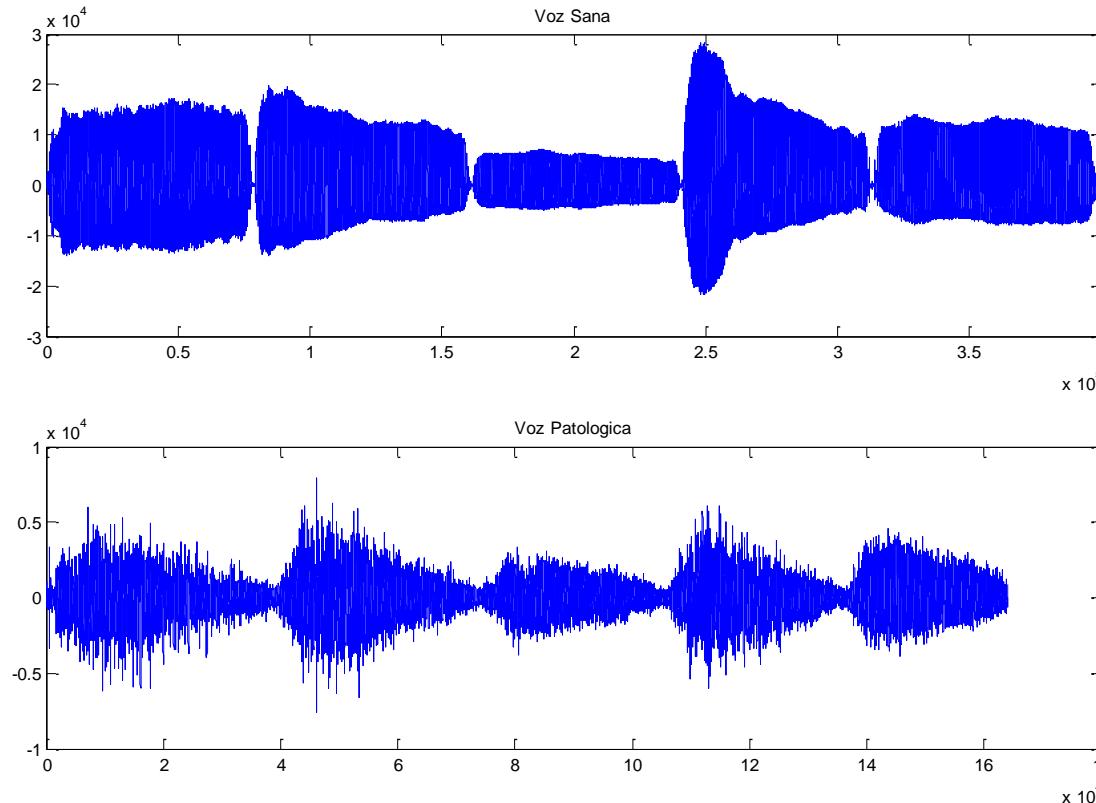


Main events in the speech signal of a pathological voice

- Time Domain
- Spectral Domain
- Cepstral Domain
- Domains which manifest the nonlinear behavior of the voice signal



Time Domain (1/3)

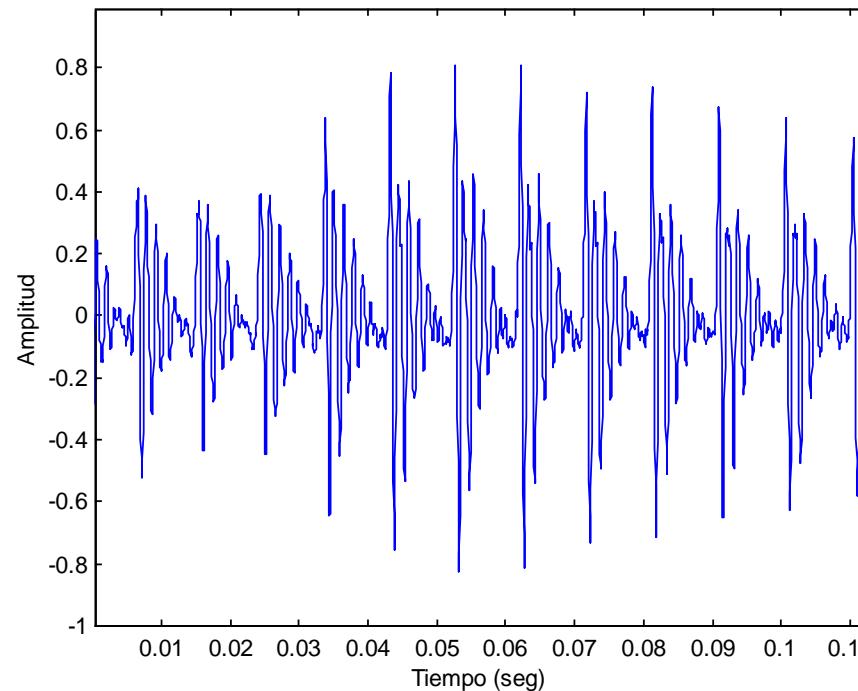


The Pathological voices present:

- The **envelope** is **less regular** than the envelope of a voice with low quality voice.



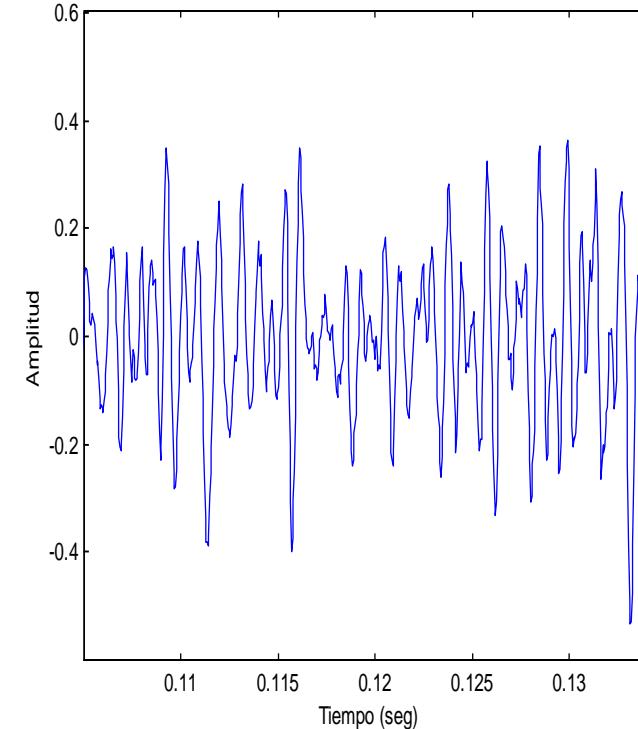
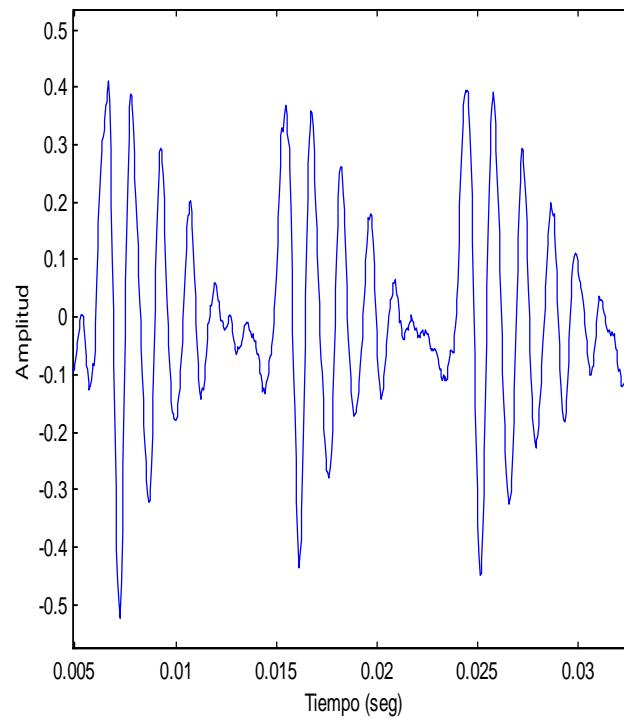
Time Domain (2/3)



- The energy of the voiced signal contained in a small time interval (e.g. 33 ms) varies greatly from one interval to the next in the voice signals of low quality.



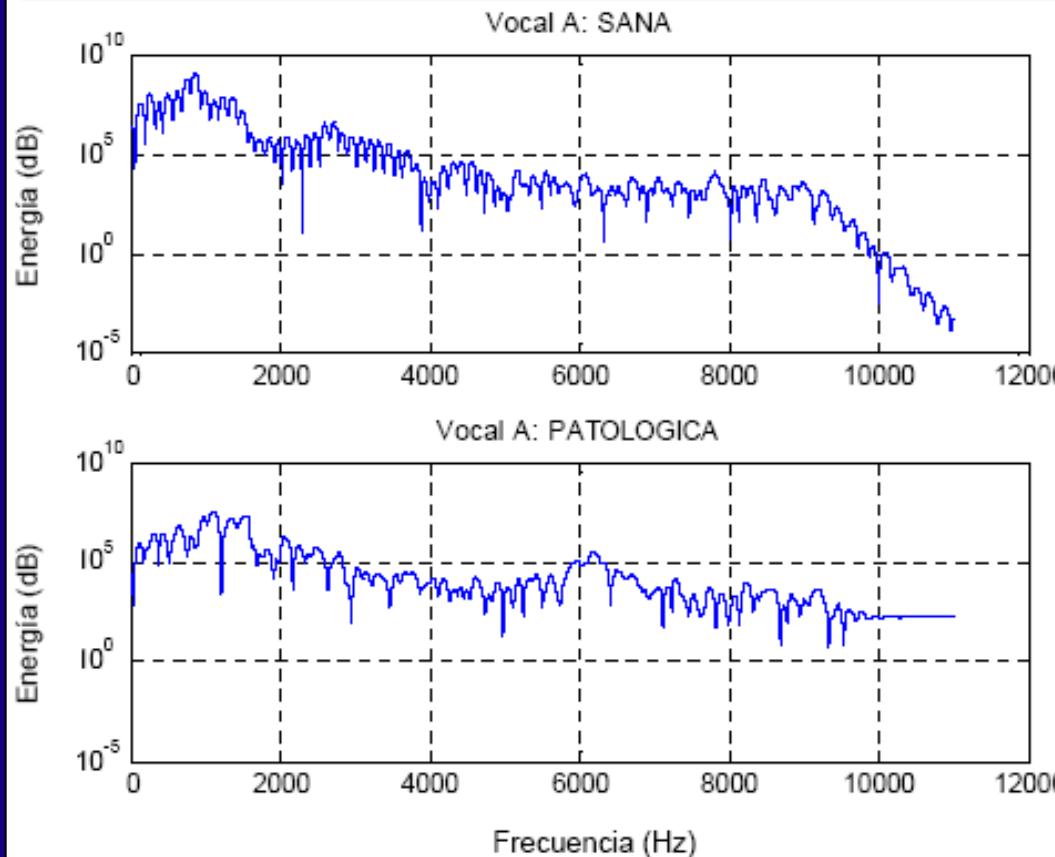
Time Domain (3/3)



- Low quality voices can have time intervals in which there is not vibration of the vocal folds during sustained sound phonations.



Spectral Domain (1/4)

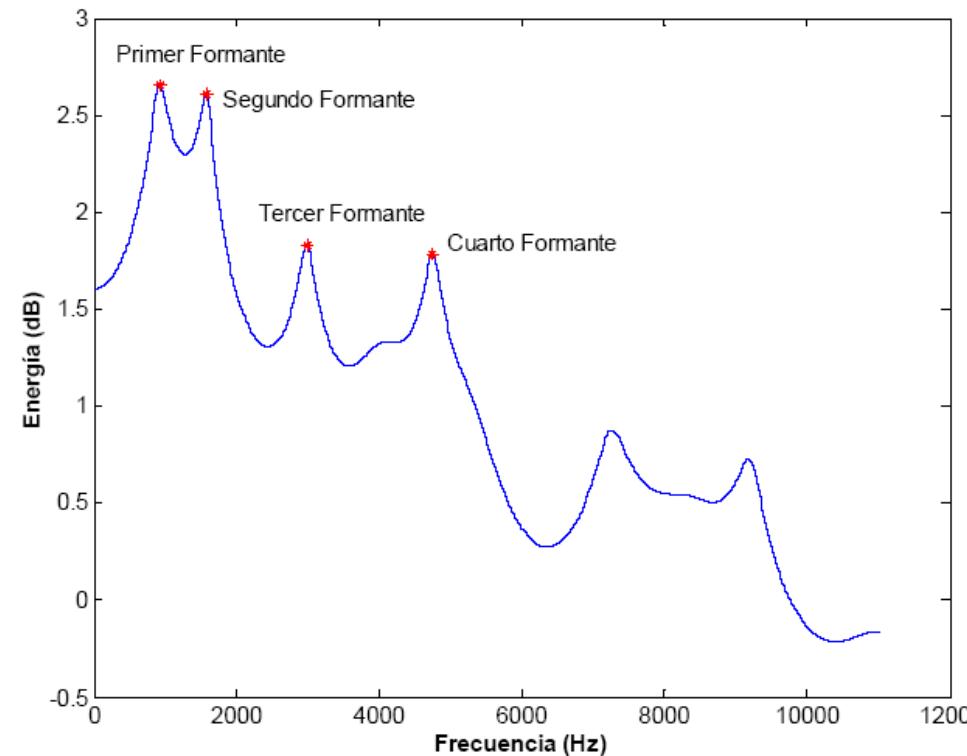


The Pathological voices present:

- Low regularity of the spectral envelope, particularly at low frequencies.
- Bigger percentage of energy in the low frequencies
- Presence of blocks of energy at high frequencies, which are due to the presence of glottal noise.
- Great variation of the spectrum of a time interval (eg 33 ms) with respect to the time interval adjacent



Spectral Domain(2/4)

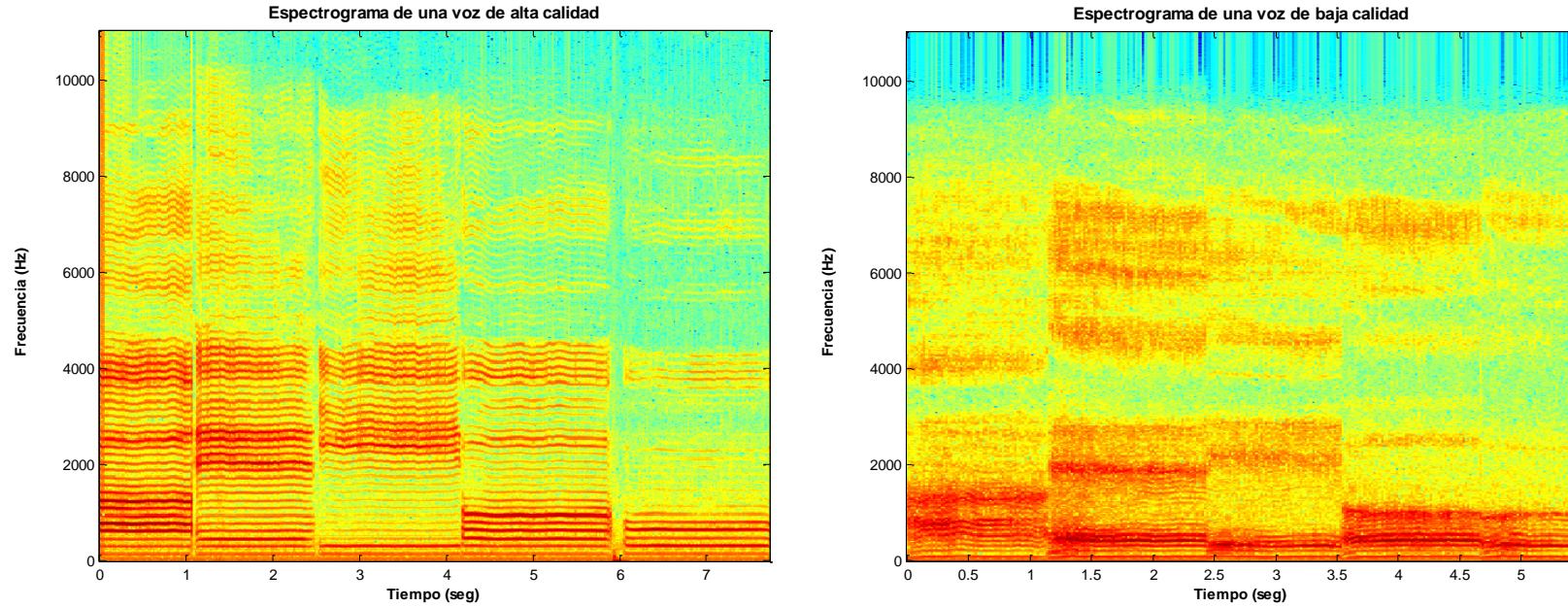


The Pathological voices present:

- large component of noise around the formants.



Spectral Domain(3/4)



The healthy voices:

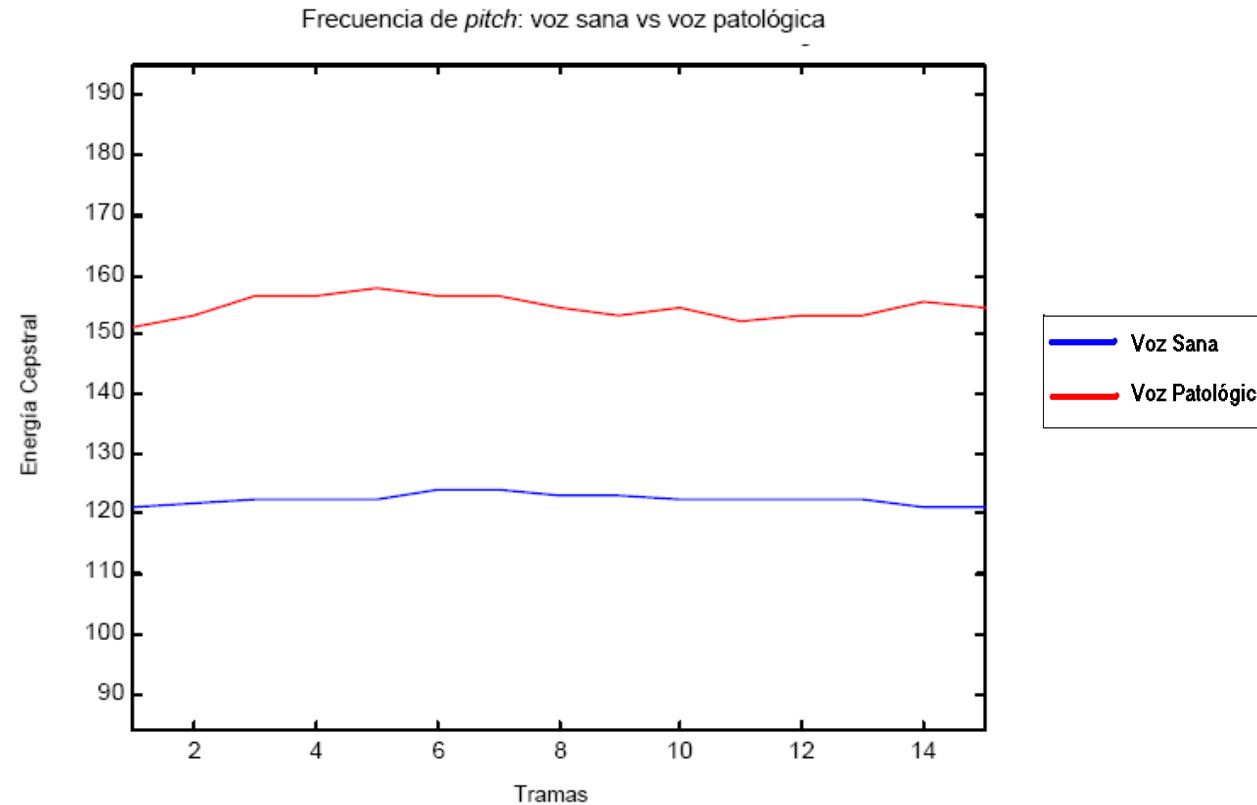
- Existence of harmonic components in both low and high frequencies (**spectral richness**)

The pathological voices:

- Have very **little harmonic component**, being concentrated in the very low frequency components.
- **Have abnormal amounts of noise.** This phenomenon occurs primarily at high frequencies and around the formants, which signifies the presence of glottal noise.



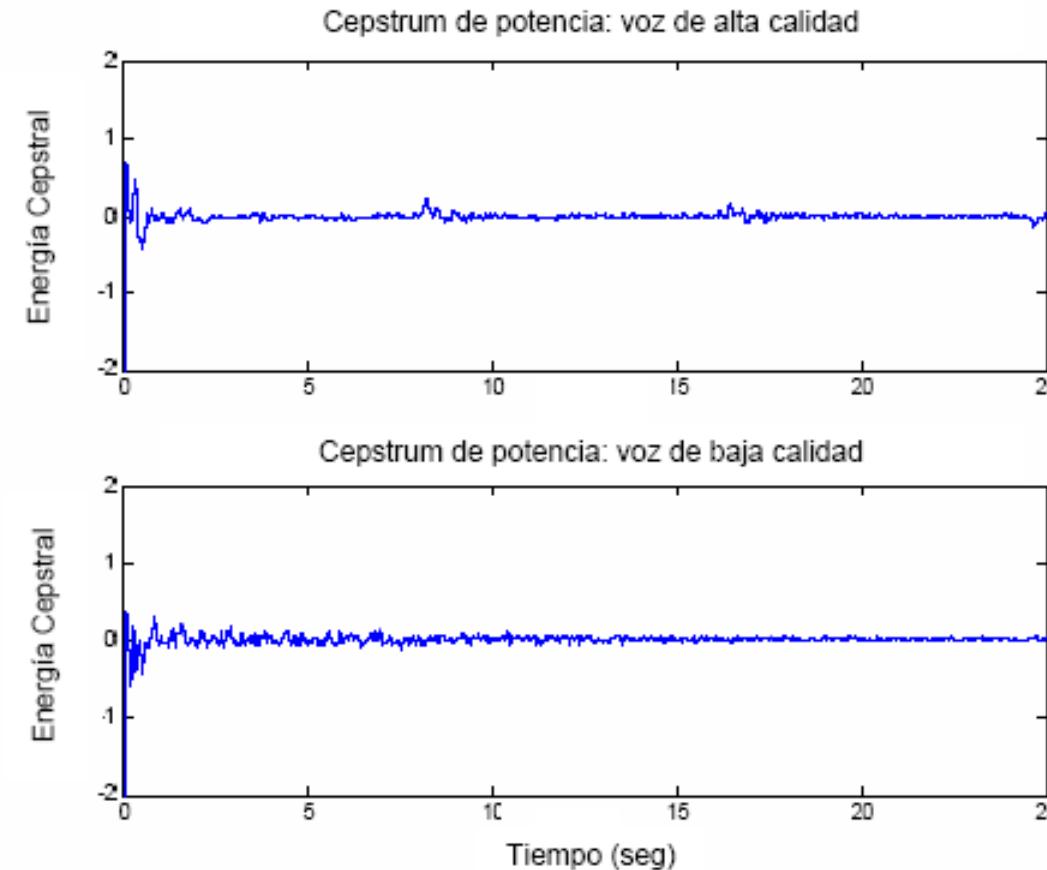
Spectral Domain (4/4)



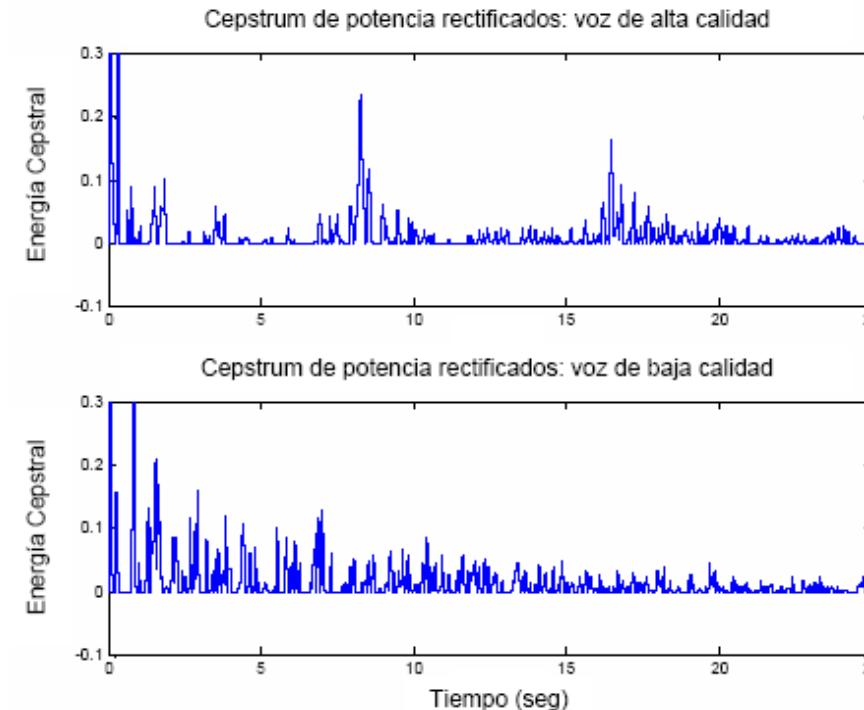
- A pathological voice has **variations in the vibration rate of the vocal folds** during the phonation of a sustained voiced sound.



Cepstral Domain (1/3)



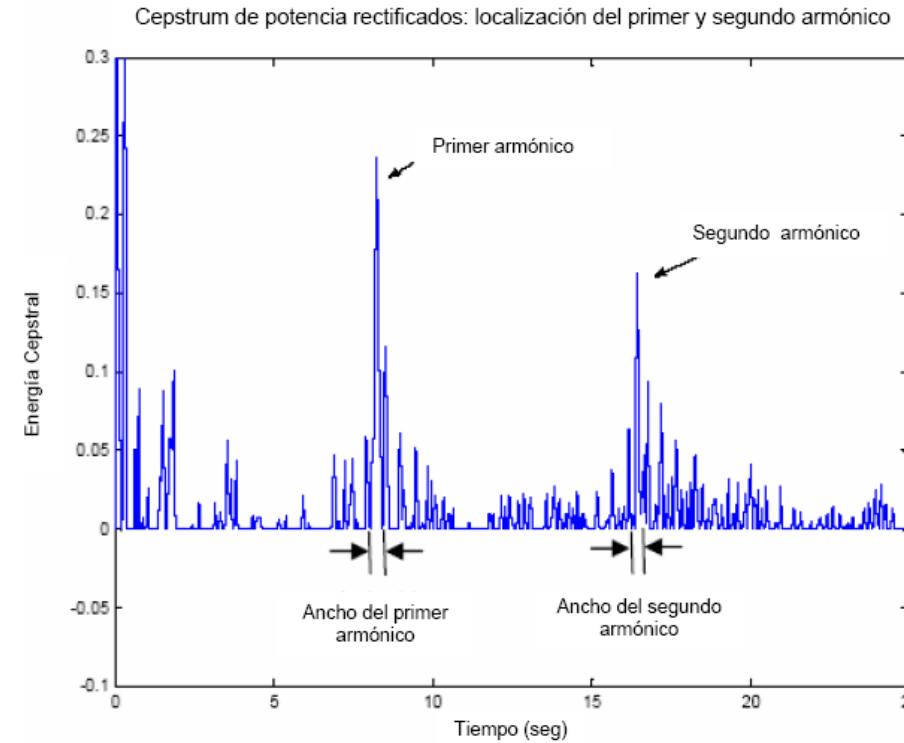
Cepstral Domain (2/3)



- The **spectral richness** can be quantified by means of the **amplitude and width** of the cepstral components corresponding to the **first and second harmonic**.
- The pathological voices have lower spectral richness.



Cepstral Domain (3/3)



- The amplitude and narrowness of the **cepstral peak corresponding to the second harmonic** are characteristics that can discriminate between high quality voice and low quality voice.
- High quality voices have a **cepstral peak corresponding to the first harmonic which it is narrower than the cepstral peak corresponding to the second harmonic**.



Nonlinear behavior angular modulations

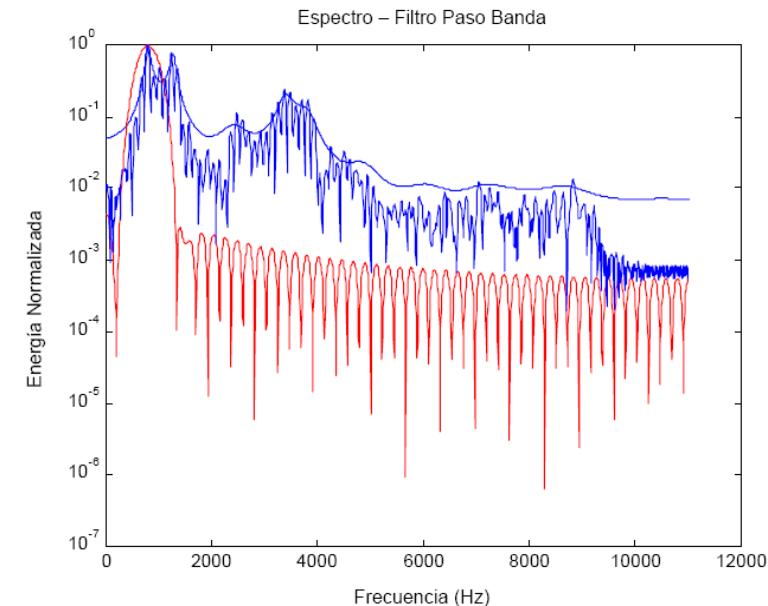
Voiced Signal

- Bandpass filter centered on the first formant
- Teager operator

$$R(t) = a(t) \cos \left\{ 2\pi \left[f_c t + \int_0^t q(\tau) d\tau \right] + \theta \right\}$$

$$\psi[s(t)] = [\dot{s}(t)^2 - (s(t)\ddot{s}(t))]$$

$$\psi[s(n)] = [s^2(n) - (s(n+1)s(n-1))]$$



Amplitude Modulation

$$a(n) \approx \frac{2\psi[x(n)]}{\sqrt{\psi[x(n+1) - x(n-1)]}}$$

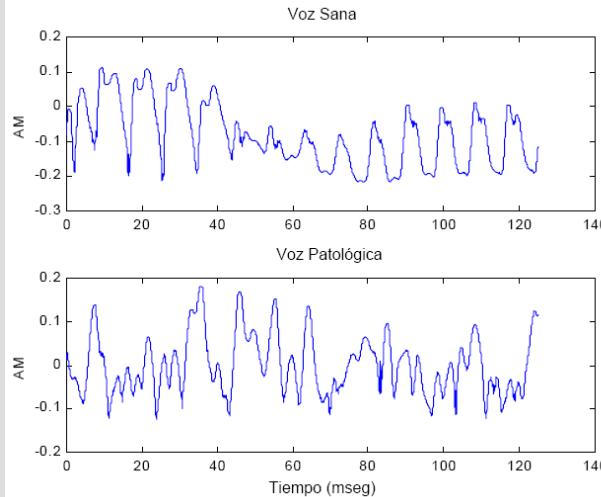
Frequency Modulation

$$f_i(n) \approx \frac{1}{4\pi T} \arccos \left\{ 1 - \frac{\psi[x(n+1) - x(n-1)]}{2\psi[x(n)]} \right\}$$

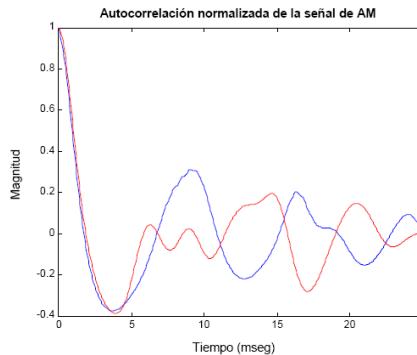
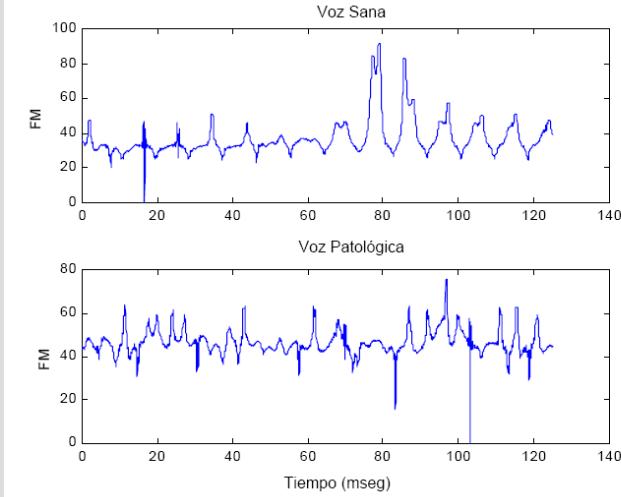


Nonlinear behavior angular modulations

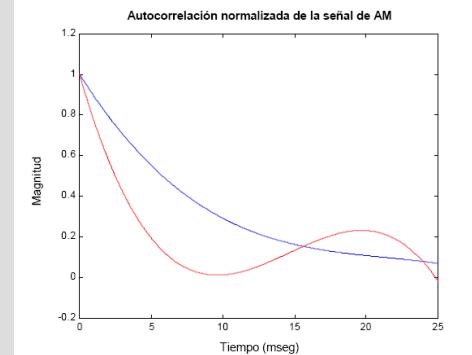
AM modulating signal



FM modulating signal



Normalized
autocorrelation of
the AM modulating
signal



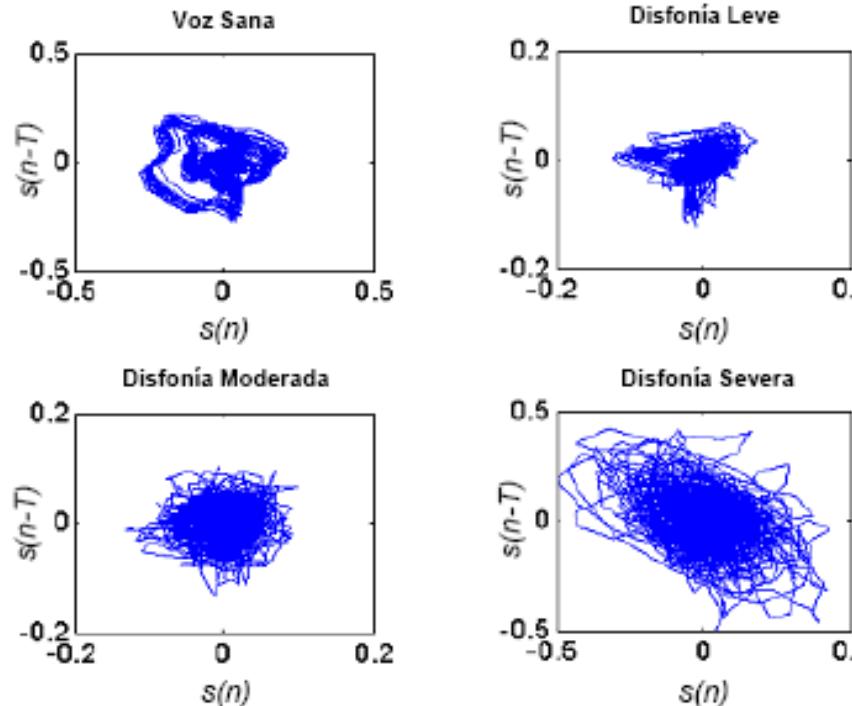
Nonlinear behavior

Deterministic nonlinear dynamical systems: Chaotic Systems

The reconstruction of this space with the method of delays is implemented by means of delayed values of the sequence of observations

$$s_n = [s(n), s(n-T), \dots, s(n-(d-1)T)]^T$$

defining a new space where T is the delay d is the dimension (**embedding dimension**). Their representation is called **attractor**.



The attractor of pathological voice signal is **more irregular and less predictable**



Nonlinear behavior Higher order statistics (HOS)

The value of kurtosis for each time window is not constant (e.g. 33 ms). This value varies along the time for the different time windows. This change is more significant for pathological voices.

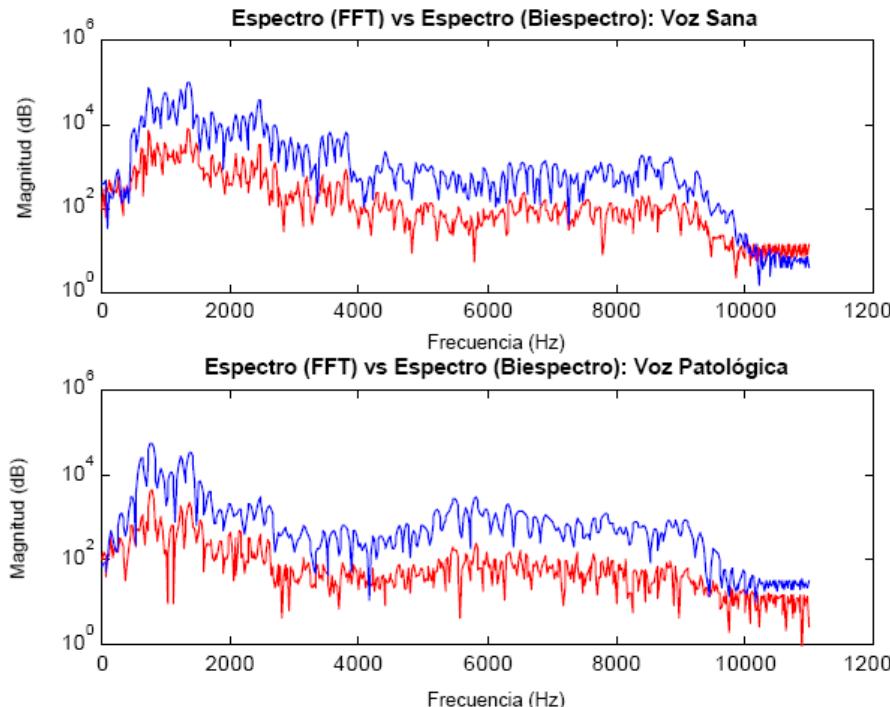
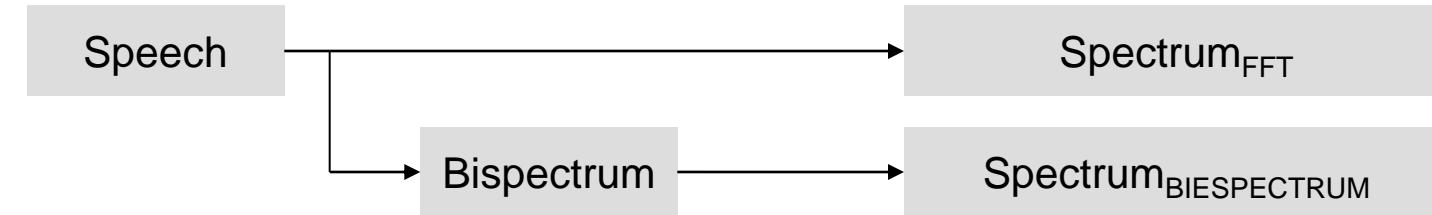
$$\gamma_4 = E\left\{ \left(\frac{x - \mu}{\sigma} \right)^4 \right\}$$

$$\hat{\gamma}_4 = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4$$



Nonlinear behavior Higher order statistics (HOS)

Quantification of Glottal Noise



Through the bispectrum is possible to separate the harmonic component and the noise component of a voice temporal window.

The noise component is not stationary, and it is more significant in the pathological voices.



Nonlinear behavior Higher order statistics (HOS)

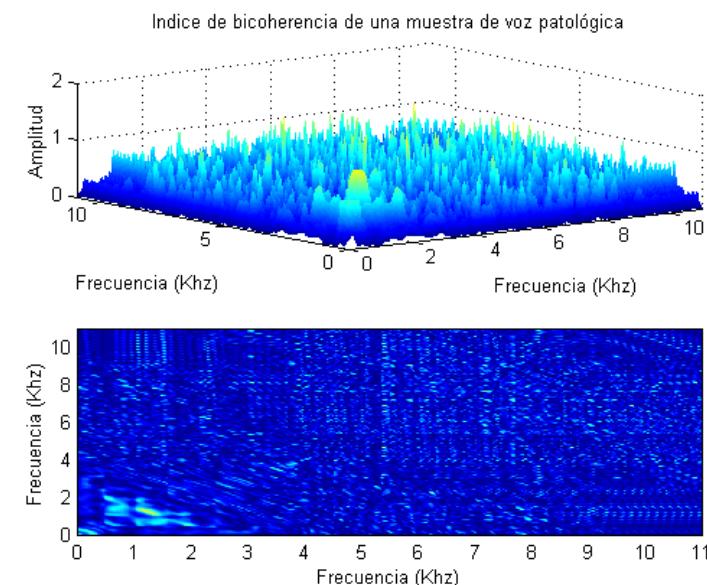
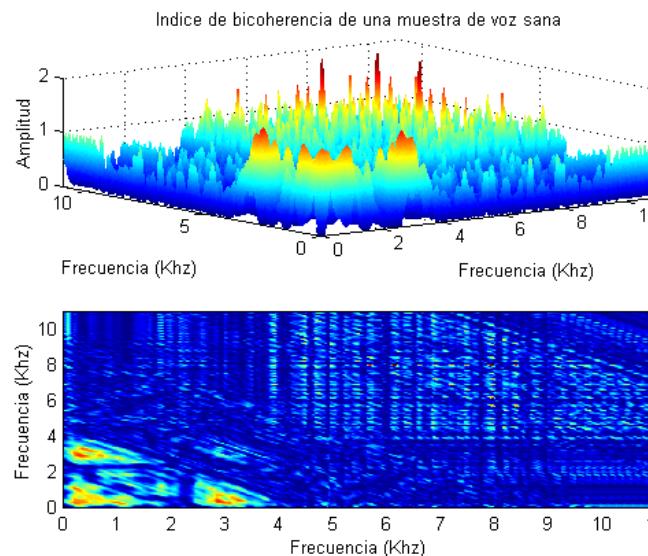
Bicoherence index

$$C_{3,x}(\tau_1, \tau_2) = E\{x(t)x(t+\tau_1)x(t+\tau_2)\}$$

$$b(\omega_1, \omega_2) \equiv \frac{B(\omega_1, \omega_2)}{P(\omega_1)P(\omega_2)P(\omega_1 + \omega_2)}$$

$$B(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3,x}(\tau_1, \tau_2) \times e^{-j \sum_{i=1}^2 w_i \tau_i}$$

There is greater roughness in the bicoherence Index in the healthy voice.





Our proposal



Proposed Solution

- There are different anomalies in the laryngeal functioning that can produce a phonation with abnormal quality.
- We have identified four quantifiable anomalies:
 - ✓ ***Problems of stability in the voice.***
 - ✓ ***Problems of glottal closure.***
 - ✓ ***Problems of rhythm in the beating of the vocal folds.***
 - ✓ ***Problems of irregularities in the masses.***
- A pathological voice implies the presence of at least one of these four anomalies.



Proposed Solution

Problems of stability in the voice:

This phenomenon characterizes the flow of air that is expelled through the lungs and generates the movement of the vocal folds (opening and closing movements).

To evaluate this phenomenon, It is used the ability of a speaker to produce an air flow for exciting the vocal folds with a constant intensity during the phonation of a sustained voiced sound.



Proposed Solution

Problems of glottal closure

This phenomenon characterizes the ability to make proper contact of both vocal folds during the phonation. The production of a voiced sound requires a correct movement of opening and closing of the vocal folds, but if there is an incorrect closing of the vocal folds, it will produce the presence of an excess noise in the speech signal.



Proposed Solution

Problems of rhythm in the beating of the vocal folds.

This phenomenon characterizes the quality of the beating (the impact between the vocal folds), in particular, the rate and force of the beating. The existence of an irregular rhythm reveals a physical difficulty which is characteristic in certain diseases. Similarly, if the beating does not occur with sufficient intensity, the signal of voice will produce a poor and abnormal sonority.



Proposed Solution

Problems of irregularities in the masses.

It is necessary that both vocal folds realize synchronously the different phases of movement (the opening and closing of the vocal fold) to get a correct beating. This phenomenon quantifies the mismatch of the different phases of movement between the vocal folds. It also quantifies the deterioration of the viscoelastic properties of the mucus that covers the vocal fold



Proposed Solution

“Any abnormal quality voice presents a deviation from the normal pattern of at least one of the four features.

- ✓ **Problems of stability in the voice.**
- ✓ **Problems of glottal closure.**
- ✓ **Problems of rhythm in the beating of the vocal folds.**
- ✓ **Problems of irregularities in the masses.**



Protocol proposed



Clinical assessment of the phonatory system

Recording process

- a **sustained vowel /aa/**; for at least 2 seconds. Must be a non whispered phonation.
- The recording can be made with a **common PC** with a **standard sound card**.
- Recording is done with a **sampling frequency of 22050 Hz** and a resolution of **16 bits per sample**.
- It may use a **conventional desktop microphone**, which present at least a bandwidth of 11000 Hz.
- It is recommended desktop microphones at a distance of **15 centimeters from the speaker**.



Clinical assessment of the phonatory system

Physical phenomena

The protocol for the quantification of voice quality, indicated the need **to measure four characteristics of voice:**

- ✓ ***Problems of stability in the voice.***
- ✓ ***Problems of glottal closure.***
- ✓ ***Problems of rhythm in the beating of the vocal folds.***
- ✓ ***Problems of irregularities in the masses.***



Clinical assessment of the phonatory system

Measures used to quantify the Physical phenomena involved

Problems of stability in the Voice	Amplitude Perturbation Quotient
Problems of rhythm in the beating of the vocal folds	Average value of the first peak Cepstral of the cepstrum derived spatially
Problems of glottal closure	Average value of Subharmonic - Harmonic Relation
Problems of irregularities in the masses	Average of the lower area of the integrated Bicoherence index



Problems of stability in the Voice

Amplitude Perturbation Quotient (APQ)

$$APQ = \left(\frac{1}{N-2} \right) \sum_{i=1}^{N-2} \frac{\left| [(P(i-1) + P(i) + P(i+1))/3] - P(i) \right|}{\frac{1}{N} \sum_{i=1}^N P(i)}$$

where $P(i)$ is the maximum amplitude of the frame i and N is the number of frames.

REF: Kay Elemetrics Corporation. 1994. Disorder Voice Database Model 4337.
Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA.



Problems of rhythm in the beating of the vocal folds

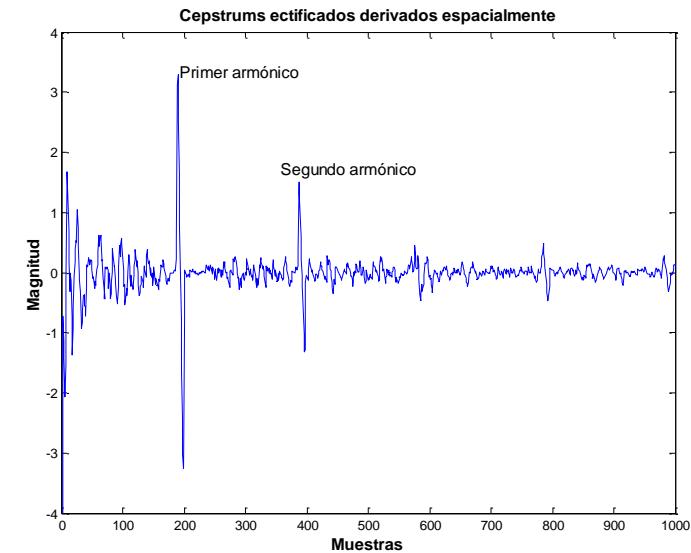
Average value of the first peak Cepstral of the cepstrum derived spatially

This measurement is obtained using a spatial derivative of a window cepstral $s(n)$

$$\frac{d}{dn} s(n) \approx \sum_{m=-N_d}^{N_d} m \cdot s(n+m) = \Delta s(n)$$

N_d equal to 5 samples for a sampling frequency 22050 Hz

REF: Alonso, J.B., de León ,J., Alonso, I., Ferrer, M.A.
(2001) "Automatic Detection of pathologies in the voice by HoS based parameters". Eurasip journal on Applied signal processing. 2001(4), 275-284



Problems of rhythm in the beating of the vocal folds

Average value of Subharmonic - Harmonic Relation

$$SHR = \frac{SS}{SH}$$

$$SH = \sum_{n=1}^N X(nf_0T)$$

X spectrum of the voice signal

f_0 fundamental harmonic

T sampling period

N number of harmonics

$$SS = \sum_{n=1}^N X((n - \frac{1}{2})f_0T)$$

REF: Sun, X., 2002. "Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio". IEEE International Conference on Acoustics, Speech, and Signal Processing. 1, 333-336.



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

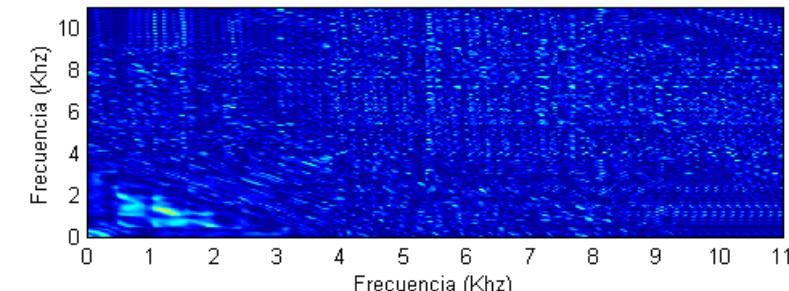
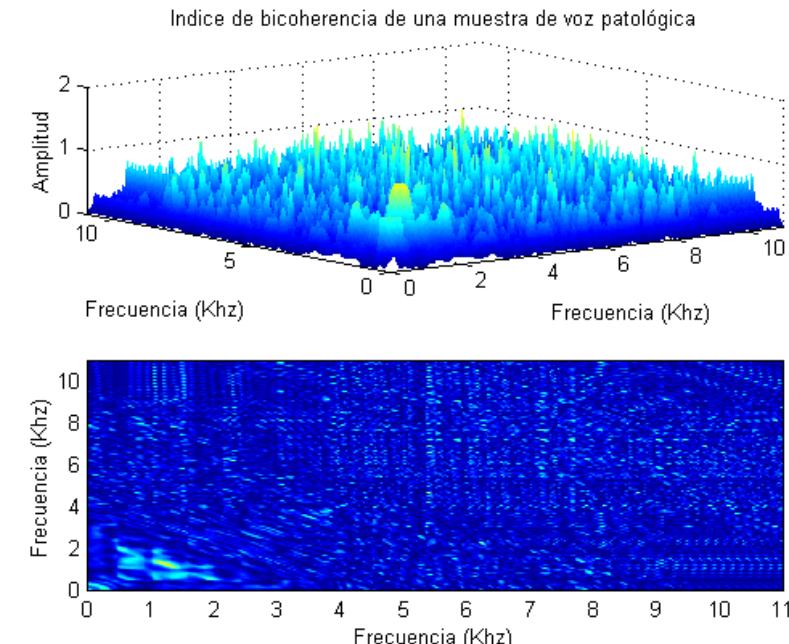
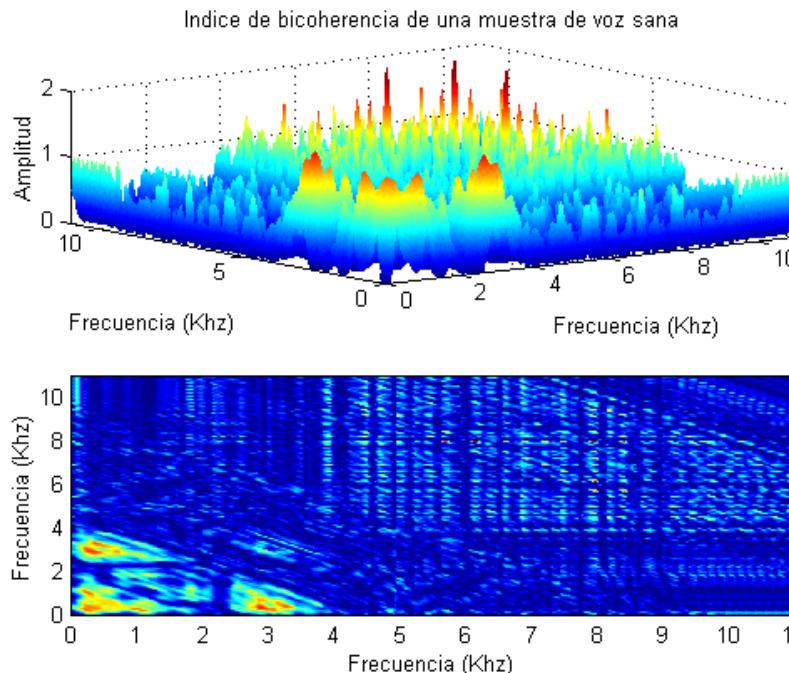
Jesús B. Alonso Hernández

Problems of irregularities in the masses

Index of Bicoherence index

$$C_{3,x}(\tau_1, \tau_2) = E\{x(t)x(t+\tau_1)x(t+\tau_2)\} \longrightarrow B(\omega_1, \omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} C_{3,x}(\tau_1, \tau_2) \times e^{-j\sum_{i=1}^2 w_i \tau_i}$$

$$b(\omega_1, \omega_2) \equiv \frac{B(\omega_1, \omega_2)}{P(\omega_1)P(\omega_2)P(\omega_1 + \omega_2)}$$



Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

Jesús B. Alonso Hernández



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

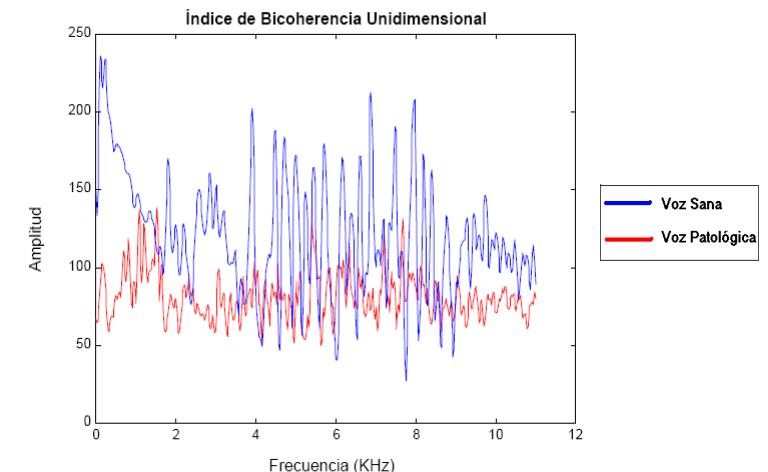
Problems of irregularities in the masses

Average of the lower area of the integrated Bicoherence index

$$b_{uni}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} b(\omega_i, \omega_j) d\omega_i$$

The threshold is set at 5300 Hz for a sampling frequency of 22050 Hz

$$Energy_{lower} = \frac{\sum_{\omega_i=0}^{\omega_{max}} b_{uni}(\omega_i)}{\text{Threshold} \sum_{\omega_i=0}^{\omega_{max}} b_{uni}(\omega_i)}$$



REF: Alonso, J.B., de León ,J., Alonso, I., Ferrer, M.A. (2001) "Automatic Detection of pathologies in the voice by Hos based parameters". Eurasip journal on Applied signal processing. 2001(4), 275-284



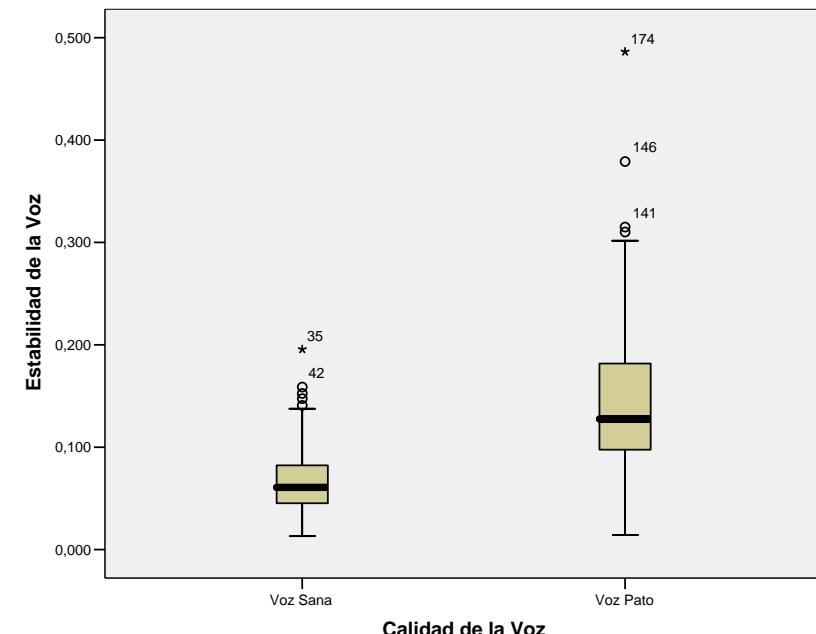
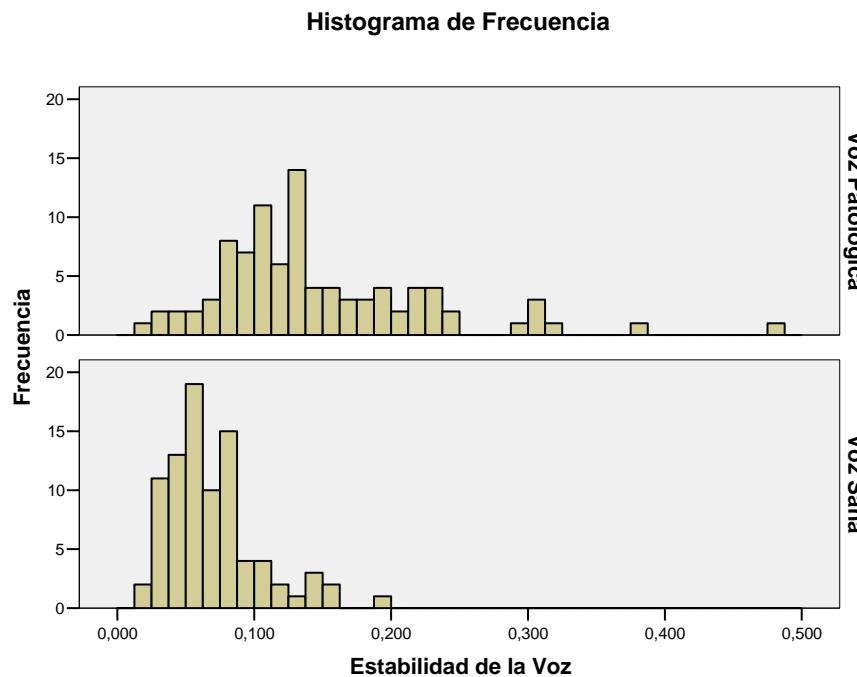
UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno ,October 24 to October 26 2012

Jesús B. Alonso Hernández

Clinical assessment of the phonatory system

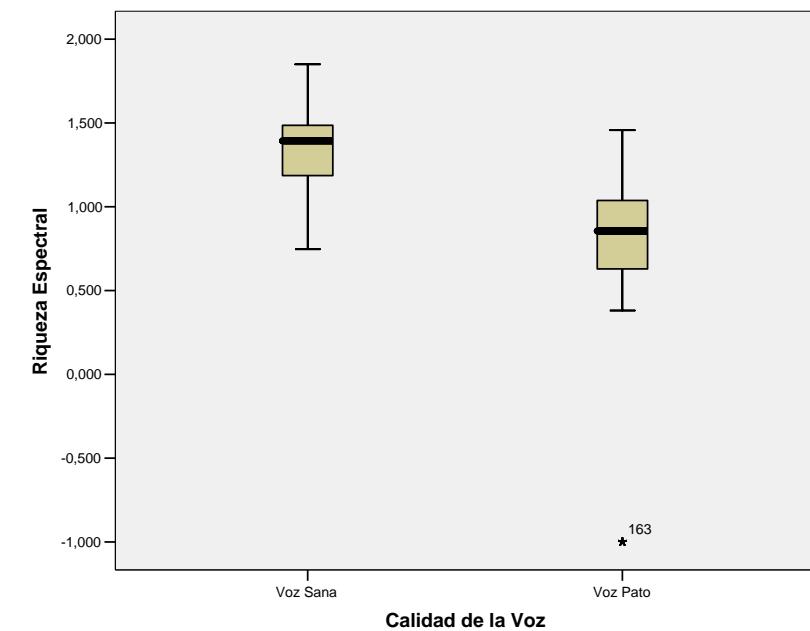
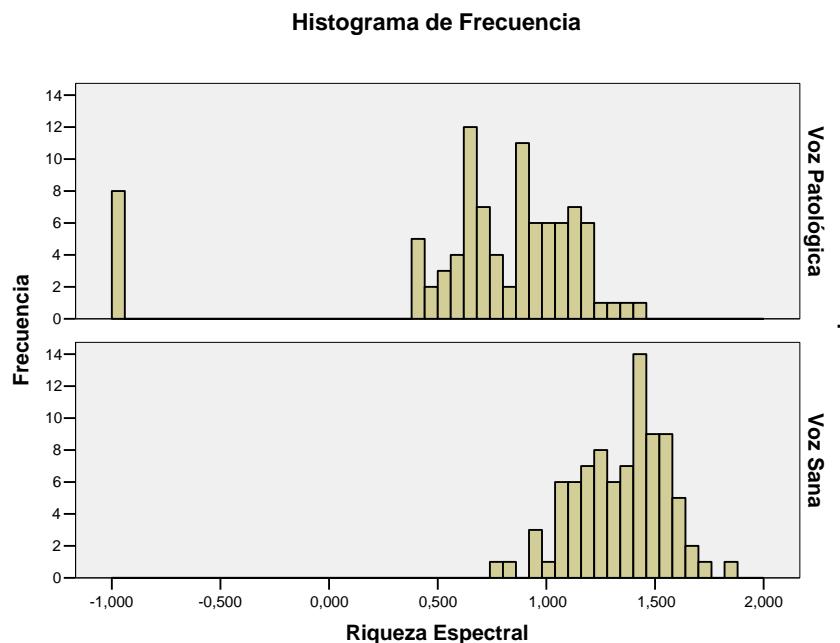
Problems of stability in the Voice **Amplitude Perturbation Quotient**



Clinical assessment of the phonatory system

Problems of rhythm in the beating of the vocal folds

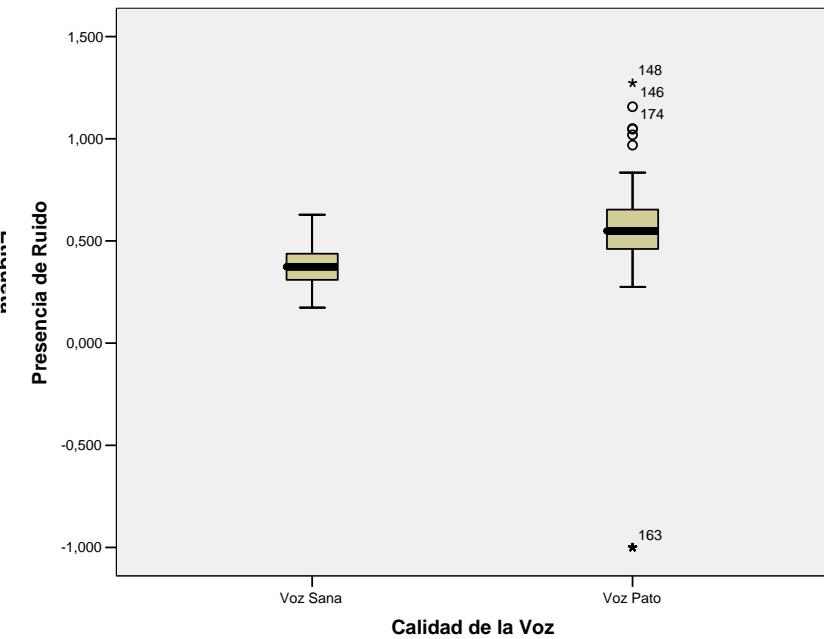
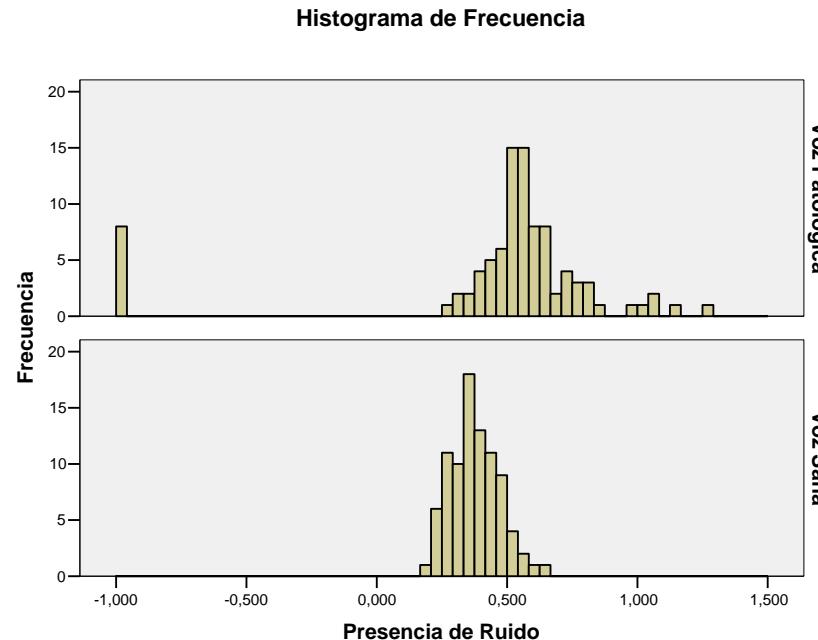
Average value of the first peak Cepstral of the cepstrum derived spatially



Clinical assessment of the phonatory system

Problems of glottal closure

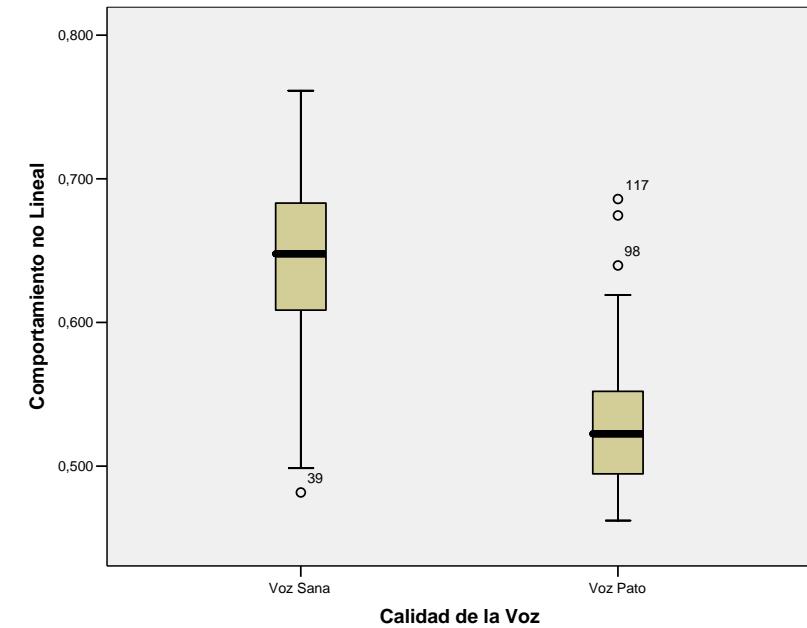
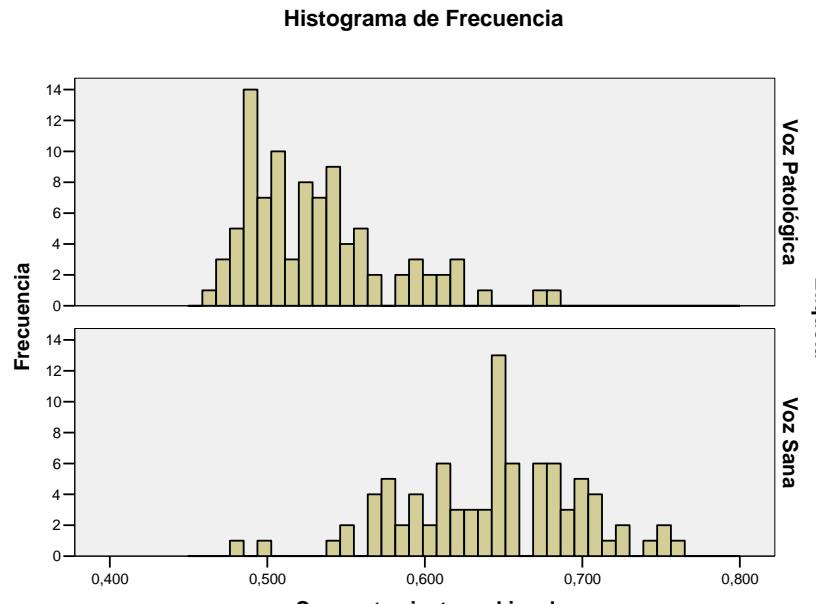
Average value of Subharmonic - Harmonic Relation



Clinical assessment of the phonatory system

Problems of irregularities in the masses

Average of the lower area of the integrated Bicoherence index



Aplicaciones

Protocolo de Evaluación

Assessment Protocol

Normal ranges for the different measures of the voice quality

Problems of stability in the Voice

0.000.....0.136

Problems of rhythm in the beating of the vocal folds

0.919.....1.759

Problems of glottal closure

0.189.....0.561

Problems of irregularities in the masses

0.530.....0.754

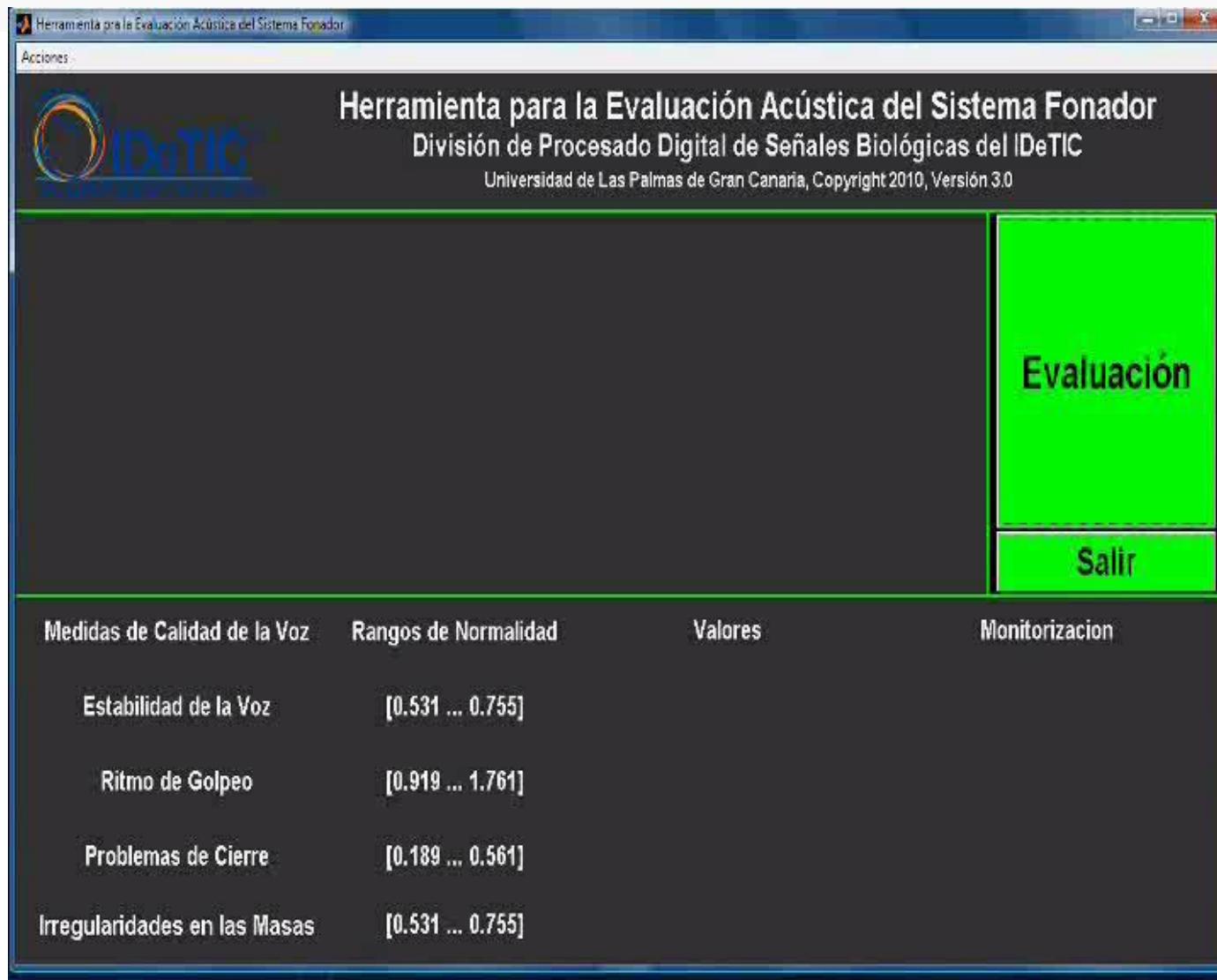


UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

Jesús B. Alonso Hernández

Video



Conclusion

- The protocol has been evaluated in a **clinical assay** in a Hospital. The conclusions of the evaluation are the following:
 - Acceptance of the tool by doctors
 - Low time-consuming in the assessment by patient.
 - Easy Interface
 - High success rate of correct diagnostic.
 - Interest to correlate measures with pathologies by doctors
- Future work:
 - To adjust the normal ranges with large databases.
 - To re-validate the proposed tool.
 - Web implementation





2nd SPLab Workshop 2012

Brno , October 24 to October 26 2012



Acoustic Analysis for the Clinical Evaluation of the phonatory system

Jesús B. Alonso Hernández (Ph.D.)
jalonso@dsc.ulpgc.es



Problems of stability in the voice

Temporal Domain

- ✓ Amplitude perturbation: *Shimmer*
- ✓ Relative average perturbation (RAP) => using the max value of the temporal interval
- ✓ Amplitude Perturbation Quotient (APQ)
- ✓ Fstr
- ✓ STCCF (Short Time Cross Correlation Function)
- ✓ Mean Waveform Matching Coefficient (MWC)
- ✓ Cycle-to-cycle mainform variation
- ✓ Kurtosis

Spectral Domain

- ✓ Normalized First Harmonic Energy (NFHE)

Inverse model domain

- ✓ Kurtosis residual signal
- ✓ Mean square value of the residual signal
- ✓ Glottal waveform
 - Closing Quotient (CQ)
 - Speech Quotient (SQ)
 - Peak value and dynamic range of a period of the residual signal



Problems of rhythm in the beating of the vocal folds

Spectral Domain

- ✓ Pitch perturbation: Jitter
- ✓ Relative average perturbation (RAP)
- ✓ Pitch period perturbation quotient (PPQ)
- ✓ HFPV (high frequency pitch variation)
- ✓ Tremor
- ✓ Fftr
- ✓ DUV (Degree of unvoicenesss)
- ✓ SHR (Subharmonic-to-harmonic ratio)
- ✓ UPR (Unvoiced periodicity Ratio)
- ✓ GST (Degree of loudness of a frame)

Inverse model domain

- ✓ Open quotient
- ✓ Speech quotient
- ✓ Closing quotient
- ✓ Normalized excitation moment
- ✓ Standard closing moment
- ✓ Standard opening moment

Cepstral Domain

- ✓ CE (Cepstral Energy)
- ✓ PCEM (Pitch energy cepstral measure)
- ✓ APR (Aperiodic periodic ratio)
- ✓ Second Harmonic Cepstral Energy(SHECM)
- ✓ Cepstrum derived spatially



Problems of glottal closure

Spectral Domain

- ✓ TNI (Turbulence noise index)

Spectral Domain

- ✓ Relations between the energies of the formants
- ✓ Relations between Regions Energy
- ✓ Relations between the energies of the formants
- ✓ Relations between Regions EnergyDistancia
- LPC
- ✓ HNR (Harmonic noise ratio)
- ✓ NNE (Normalized Noise Energy)
- ✓ Spectral distance measures

Cepstral Domain

- ✓ CHNR (Cepstral Harmonic Noise Ratio)
- ✓ HNRS (Harmonic to Noise Ratio from Speech)
- ✓ GNE (Glotal-to-Noise Excitation Ratio)
- ✓ NEP (Normalised Error Prediction)

Inverse model domain

- ✓ Roughness spectral residual signal(SFR and SFF)
- ✓ HNRR(Harmonic to Noise Ratio from Residual)



Problems of irregularities in the masses

Angular modulations

- ✓ The FM signal variation along the time
- ✓ On the normalized autocorrelation envelope of the modulating signal AM (vector area measurement, slope)

Higher Order Statistics (HOS)

- ✓ Perturbation of Integrated Bicoherence Index
- ✓ Quantification of Glottal Noise

Deterministic nonlinear dynamical systems: Chaotic Systems

- ✓ The correlation dimension
- ✓ The Lyapunov exponent
- ✓ Estimator minimum mutual information
- ✓ Estimation of the correlation dimension (Takens estimator)
- ✓ Estimator of the average entropy correlation (Grassberg-Procaccia)
- ✓ Mean estimator Renyi entropy of order 1
- ✓ Mean estimator Renyi entropy order 2
- ✓ Estimator Shannon entropy of the vowel



Estabilidad de la Voz

Análisis

- Las medidas de “estabilidad de la voz” presentan especial sensibilidad a dos aspectos:
 - al ruido ambiental, especialmente al ruido no estacionario,
 - al movimiento del locutor respecto al micrófono.
- Por este motivo, resulta recomendable medidas que cuantifican la variación de energía de la señal de voz a largo plazo.
- Determinadas medidas de calidad de la voz, necesitan información sobre el valor del pitch. Esta situación se convierte en un inconveniente frente a las muestras de voces de muy baja calidad de voz.



Riqueza Espectral

Análisis

- Las medidas presentan especial sensibilidad al detector de *pitch*.
 - Este sensibilidad se incrementa frente a la presencia de ruidos no estacionarios y relaciones señal ruido muy bajas.
 - Es posible minimizarlo por medio de detectores de pitch que minimicen la posibilidad de detectar frecuencias doble o mitad de *pitch*
 - resulta de interés la medida de la variabilidad del *pitch* a largo plazo.
 - Los casos de fonaciones “sonoras” de muy baja calidad, las medias que necesitan información del *pitch* no pueden ser estimadas, pudiéndose solventar con una codificación adecuada.
- La cantidad de riqueza espectral ...
 - ... simplemente una característica más de las peculiaridades de la voz de un locutor aunque sin embargo, niveles muy bajos de armónicos son indicios de una baja calidad de la voz.
 - ... generalmente, la información no está en la cantidad de armónicos existentes en una fonación sonora sostenida, está en la variación a lo largo del tiempo de la cantidad de armónicos.
- Respecto a los dominios de representación:
 - el dominio de los Cepstrum, o cualquier otro dominio de representación derivado, permite cuantificar adecuadamente los niveles de riqueza espectral.
 - Las medidas basadas en filtrados inversos del tracto vocal, son muy sensibles a los algoritmos de filtrado utilizados, principalmente frente a:
 - la presencia de ruidos estacionarios
 - niveles de relación señal a ruido bajas
 - señal de voz presenta bajos niveles de periodicidad.



Comportamiento no Lineal

Análisis

- Los dominios no lineales presentan gran interés debido a la capacidad potencial de manifestar información que en otros dominios no está presente o es difícilmente cuantificable.
- **Los estadísticos de alto orden**, permiten cuantificar la presencia de alinealidades o la posibilidad de separar la componente armónica de la componente de ruido. Sin embargo, presentan un coste computacional significativo.
- Las técnicas basadas en características obtenidas a partir de **modelos dinámicos no lineales** presentan un gran potencial. Sin embargo, también presentan el inconveniente de que si los modelos están perfectamente sintonizados (dimensión del espacio de fases y el *delay*) probablemente no sea posible obtener las máximas prestaciones.



Presencia de Ruido

Análisis

La presencia de ruido en la señal de voz, sea estacionario o no, implica:

- Dificultad para estimar medidas de calidad en el dominio temporal
- Gran dificultad para estimar medidas basadas de la identificación de determinadas características de la señal de voz: primer y tercer formante, el pitch, etc.
- Resulta interesante determinadas medidas de calidad que se estiman en función de la variación a lo largo del tiempo de una determinada magnitud o relación:
 - Niveles de energía o relaciones de energía
 - Relaciones de señal a ruido



Methods for assessment the phonatory system

Hearing-based methods: GRABS Scale

Each item is rated on a 4-point scale:
(0 = normal, 1 = slight, 2 = moderate, 3 = severe)

GRADO (G) GRADE	Indicates the overall degree of affectation of the voice, with 0 corresponds to a normal voice, 3 corresponds to severe pathological voice.	
VOZ RASPOSA (R) ROUGH	The psychoacoustic impression of irregularity of the vocal folds vibration or defect of vibration (friction), or even the absence of vibration.	
VOZ AÉREA (B) BREATHY	The psychoacoustic impression of loss of air in the glottis.	
VOZ ASTÉNICA (A) ASTENIC	It would be the degree of asthenia or fatigue	
TENSIÓN EN LA VOZ (S) STRAIN	The psychoacoustic impression of effort and hyperfunction.	

Methods for assessment the phonatory system

Hearing-based methods: GRABS Scale

inconvenient of the GRABS scale:

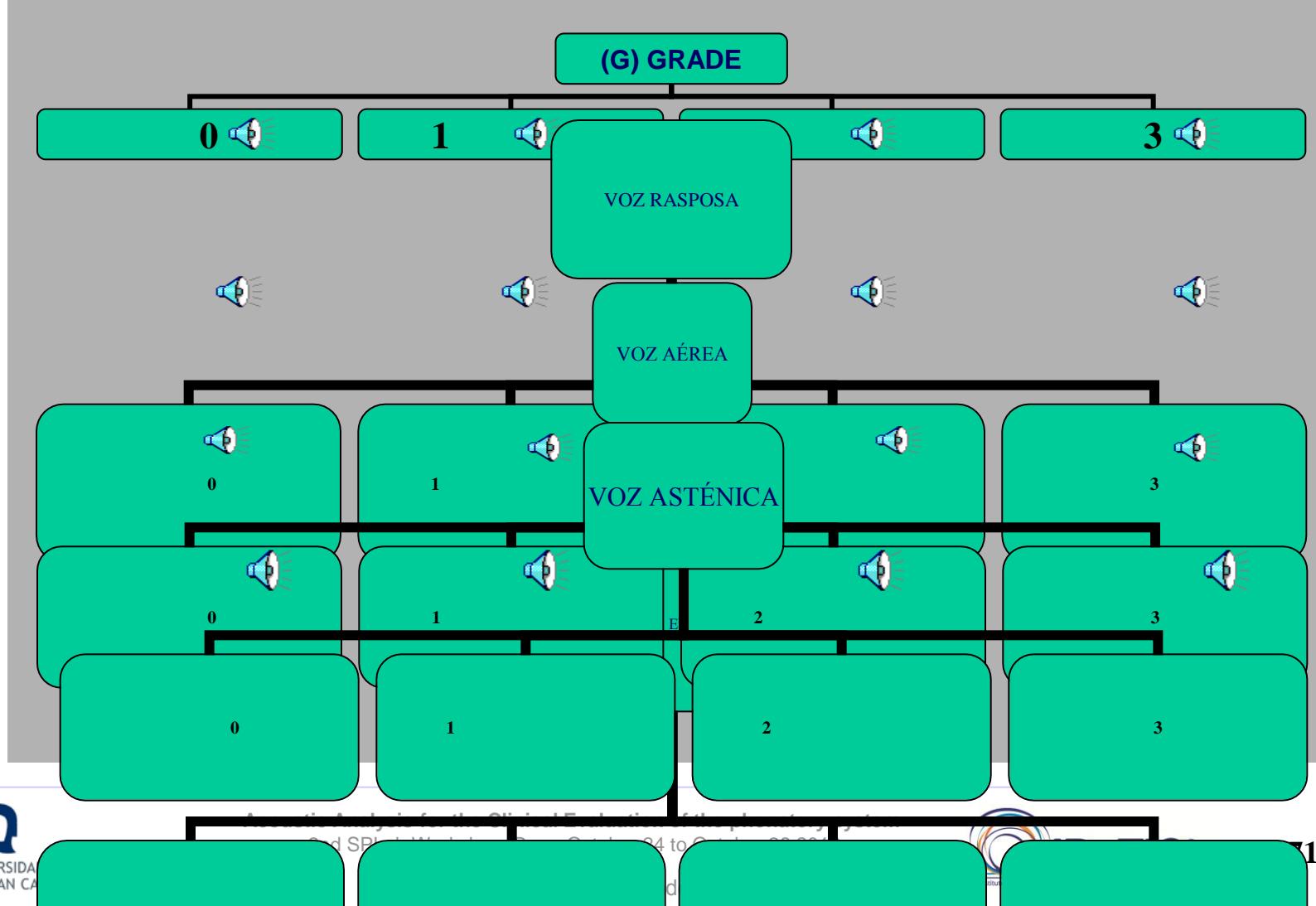
- Subjective technique
- Based on the experience
- The results in the evaluation depend of the evaluator
- Effect of the Fatigue in the evaluator



Métodos basados en la audición – Técnicas Subjetivas

Escala GRABS : Ejemplos

Fuente: Cristina A. Jackson-Mehaldi, Ph.D.
Director Lakeshore Professional Voice Center, (LENT) and Adjunct Associate Professor Department. Otolaryngology,
Wayne State University, Detroit, Michigan



Origin and Motivation

Balance:

Publications: international journals and conferences (technical field and medical field). Some citations:

- Jesús B. Alonso and Patricia Henríquez, "Speech-based Clinical Diagnostic Systems", in *Encyclopedia of Artificial Intelligence. IGI Global publication*, ISBN: 978-1-59904-849-9, pp. 1439-1446, July 2008.
- Jesús B. Alonso, José de León, Itziar Alonso, and Miguel A. Ferrer, "Automatic Detection of Pathologies in the voice by HOS based Parameters", in *EURASIP Journal on Applied Signal processing*, vol. 2001, no 4, pp. 275-284, December 2001.
- Patricia Henríquez, Jesús B. Alonso, Miguel A. Ferrer, Carlos M. Travieso, Juan I. Godino-Llorente, Fernando Diaz-de-Maria, "Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics", in *IEEE Transactions on Audio, Speech and Language Processing*, ISSN: 1558-7916, vol. 17, no. 6, pp.1186-1195, USA, August 2009.
- Jesús B. Alonso, Fernando Díaz-de-Maria, Carlos M. Travieso, Miguel A. Ferrer, "Optimal Size of Time Window in Nonlinear Features for Voice Quality Measurement", in *Nonlinear Analyses and Algorithms for Speech Processing in Springer-Verlag Lectures Notes in Computer Science*, ISSN: 03029743, vol. 3817/2006, pp.206-217, February 2006.
- Patricia Henríquez, Jesús B. Alonso, Miguel A. Ferrer, Carlos M. Travieso, Rafael Orozco-Arroyave, Application of Nonlinear Dynamics Characterization to Emotional Speech, *Non-Linear Speech Processing (NoLISP 2011)*, Lecture Notes in Artificial Intelligence. Ed. Springer-Verlag, ISSN: 0302-9743, ISBN: 978-3-642-25019-4, Vol I, pp.127-136, Las Palmas de Gran Canaria, España, 07-09 de Noviembre, 2011
- José de león, J.F. Rivero, F. Ayudarte, M.A. Nuñez, J.B. Alonso, M.A. Ferrer, "Morphological Análisis of the Glottal Excitation Curve", *XVIIth World Congreso of The internacional Federation of Oto-Rhino-Laryngological Societies*, Egipto, 2002
- José de león, J.F. Rivero, F. Ayudarte, M.A. Nuñez, J.B. Alonso, M.A. Ferrer, "Temporal Evolution Study of Pathological Voice Features", *XVIIth World Congreso of The internacional Federation of Oto-Rhino-Laryngological Societies*, Egipto, 2002



Origin and Motivation

Balance:

Research National Projects:

- (01-03) **Proyecto CICYT** (TIC2000-0586) : Diagnóstico Automático en la Patología de la Voz
- (04-06) **Proyecto CICYT** (TIC2003-08956-C02-02) : DAMAVOZ: Diagnóstico Automático con medidas objetivas de calidad mediante evaluación de la voz y la fonación
- (07-09) **Proyecto CICYT** (CICYT TEC2006-13141-C03-01/TCM) CICYT BIO-PASS: Uso de modalidades transparentes en esquemas de reconocimiento multibiométrico para entornos de aplicación realistas de baja vulnerabilidad y alta conveniencia de usuario
- (10-12) **Proyecto CICYT** (TEC2009-14123-C04-01) BIOMET: identificador Biométrico con Metabiometrías

PhD thesis:

– “Metodología del Análisis Acústico para la Evaluación Clínica de la Voz” (June 2006).

Books:

“La Evaluación Acústica del Sistema Fonador” Ed. ULPGC



UNIVERSIDAD DE LAS PALMAS
DE GRAN CANARIA

Acoustic Analysis for the Clinical Evaluation of the phonatory system
2nd SPLab Workshop -Brno , October 24 to October 26 2012

Jesús B. Alonso Hernández