



**HAL**  
open science

## Voice and Speech therapy using VOCALAB - From research to practice

Étienne Sicard, Anne Menin-Sicard, Stéphanie Perrière, Julie Mauclair

► **To cite this version:**

Étienne Sicard, Anne Menin-Sicard, Stéphanie Perrière, Julie Mauclair. Voice and Speech therapy using VOCALAB - From research to practice. 1st International Interdisciplinary Voice Congress, Nov 2016, Madrid, Spain. pp.0. hal-03658839

**HAL Id: hal-03658839**

**<https://hal.science/hal-03658839v1>**

Submitted on 4 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in: <https://oatao.univ-toulouse.fr/22034>

**To cite this version:**

Sicard, Etienne and Menin-Sicard, Anne and Perrière, Stéphanie and Mauclair, Julie *Voice and Speech therapy using VOCALAB - From research to practice*. (2016) In: 1st International Interdisciplinary Voice Congress, 24 November 2016 - 25 November 2016 (Madrid, Spain). (Unpublished)

Any correspondence concerning this service should be sent to the repository administrator: [tech-oatao@listes-diff.inp-toulouse.fr](mailto:tech-oatao@listes-diff.inp-toulouse.fr)



# Voice and Speech therapy using VOCALAB From research to practice

**E. Sicard**, A. Menin-Sicard, S. Perriere, J. Mauclair

- (1) Professor, INSA Toulouse, France – [etienne.sicard@insa-toulouse.fr](mailto:etienne.sicard@insa-toulouse.fr)
- (2) Speech Therapist, LURCO, Toulouse France – [anne.sicard2@orange.fr](mailto:anne.sicard2@orange.fr)
- (3) Speech Therapist, LURCO, Nice, France - [stephanie.perriere@gmail.com](mailto:stephanie.perriere@gmail.com)
- (4) Senior Lecturer, IRIT, France - [julie.mauclair@irit.fr](mailto:julie.mauclair@irit.fr)

- Objectives
- State of the art
- Methodology
- Key results
- Conclusion

# OBJECTIVES

- Help **speech therapists** in the evaluation and therapy of voice and speech
- Address patients of **all ages**, most **common** voice pathologies
- Improve speech therapy efficiency
- Provide the optimum tools for efficient voice/speech evaluation and rehabilitation
- Help to **interpret measurements**
- Gather voice therapy knowledge and provide free access to **data base**



*© Edward, D. Sicard & S. Moisan*

- **Furgal innovation** : low cost for the largest number of therapists
- Best possible tools for efficient evaluation and therapy
- **Proven methods** of therapy, keep seniors in their homes
- Analyze **groups of patients** (Cancer, Parkinson, Nodules, Kysts...)
- **Compare** approaches and tools
- Provide training, educational contents
- Cooperate with international experts



© Edward, D. Sicard & S. Moisan

**INSA**

INSTITUT NATIONAL  
DES SCIENCES  
APPLIQUÉES  
TOULOUSE

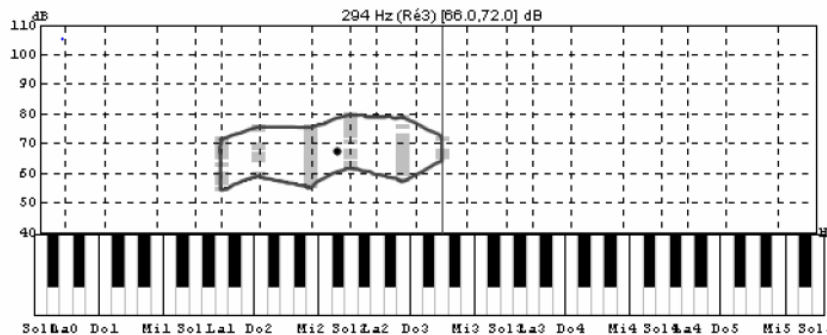
# STATE OF THE ART



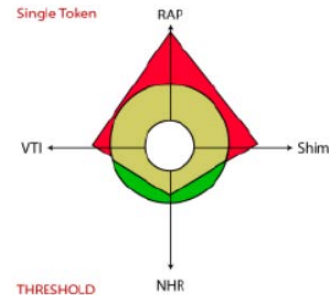
# VOICE EVALUATION SPEECH THERAPISTS

Type of data	Instant	Very short term	Short term	Medium term	Long term	Very long term
Vowel [a]	Attack quality	Variations of frequency and amplitude		Power, breathiness, dynamic voice range		Maximum phonation time
[a/s/z]						Ratio [s/z, a/z]
Siren				Vocal range		
Speech						Average fundamental Frequency, prosody

Dynamic voice Range EVA2 (Teston 2004)



MDVP  
Kay Elemetrics

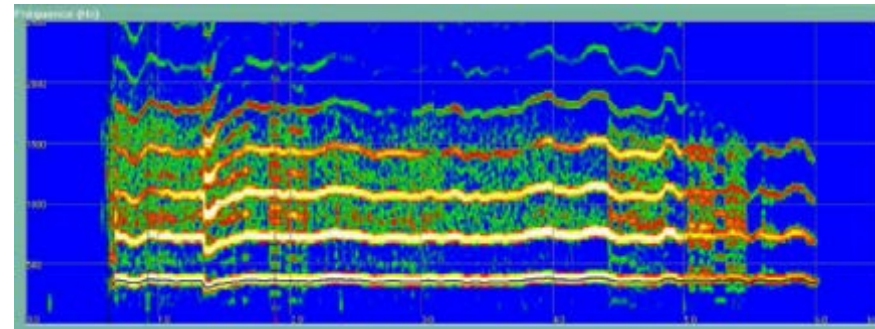


# VOICE EVALUATION INDICATORS

<i>PRAAT</i>	<i>EVA</i>	<i>MDVP</i>	<i>Our work</i>
Average F0 Maximum Phonation Time Dynamic phonation range	Average F0 Maximum Phonation Time Dynamic phonation range	Average F0 Maximum Phonation Time Dynamic phonation range	Average F0 Maximum Phonation Time Dynamic phonation range <b>Ratio s/z, a/z</b>
<b>Pitch:</b> Jitter , Jita, DDP, ratio, RAP, PPQ5, jDDP	<b>Pitch :</b> JittA, Jitt Factor, Jitt Ratio, vF0, RAP	<b>Pitch:</b> PPQ, Jitt, RAP, Jita, sPPQ , STD, vf0, ..	<b>Pitch : Instability</b>
<b>Amplitude:</b> Shim, ShdB, APQ3, APQ5, APQ11, ADDP	<b>Amplitude :</b> Shim Factor, APQ	<b>Amplitude:</b> APQ, Shim, ShdB, sAPQ, vAm	<b>Amplitude: Instability</b>
<b>Noise :</b> HNR	<b>Noise :</b> Sr, Sr>1KHz	<b>Noise :</b> VTI, NHR, SPI , FTRI, ATRI	<b>Noise: harmonic ratio</b>
<b>Breaks:</b> FLUF, DVB	<b>Pressure :</b> efficiency, glottal resistance...	<b>Harmonics:</b> DSH, NSH <b>Breaks:</b> DVB, DUV, NVB, NUV	<b>Harmonic poverty</b> <b>Attack alteration</b>

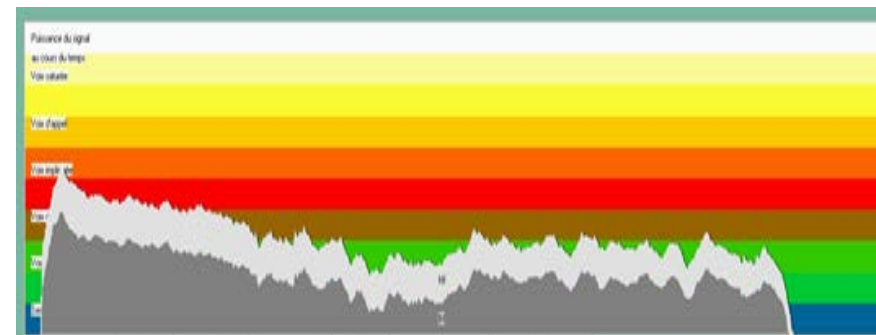
- Pitch instability. Important **variability** in Jitter results, families of indicators strongly **correlated**, difficulty of **detecting F0** in pathological voices (Maryn 2009, Werth 2010).
- Amplitude Instability. Different tools and Shimmer algorithms give **different results** (Maryn, 2009). Shimmer families **strongly correlated**, **sensitive to noise** (Werth, 2010)

Frequency



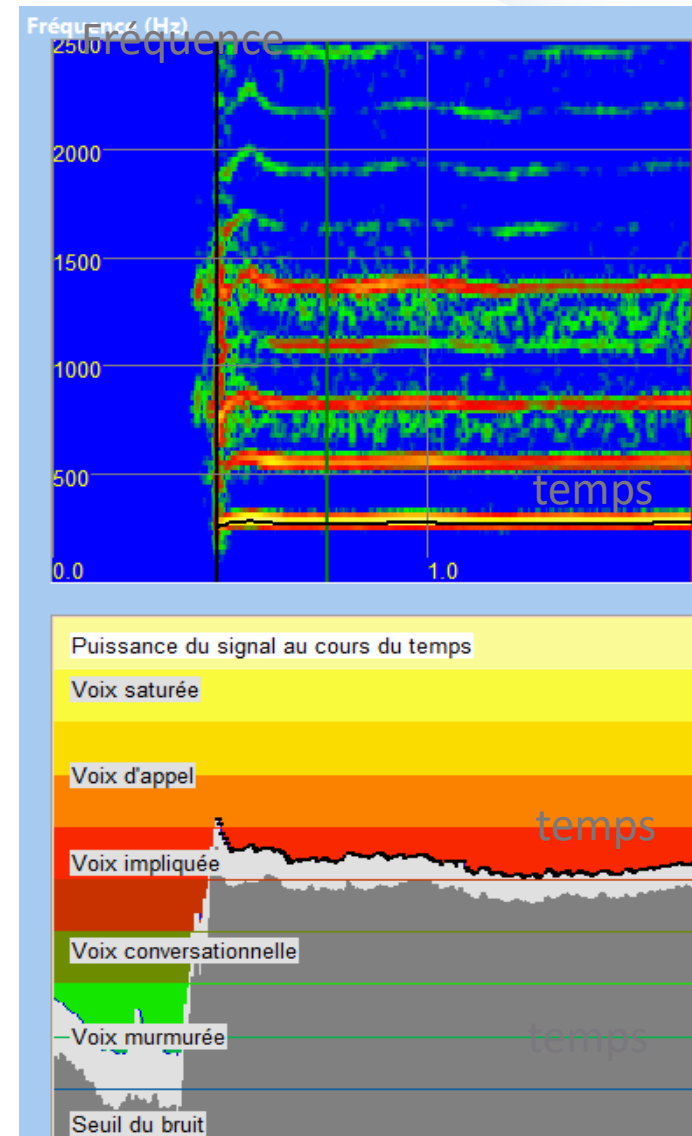
Time

Frequency



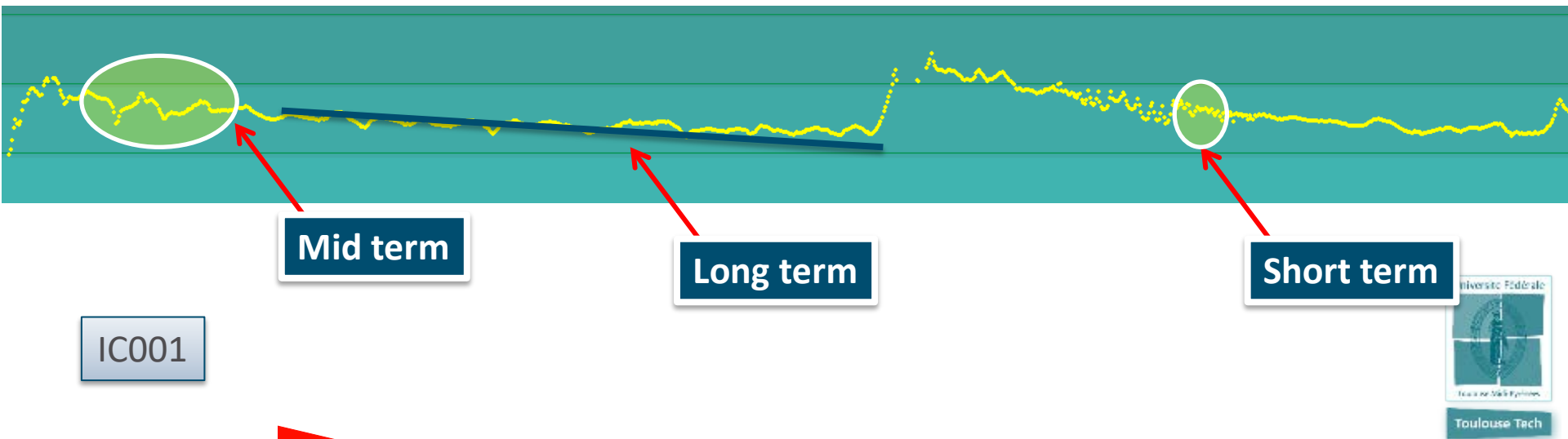
Time

- Noise. HNR, NNE, SNR... Doubts on the **relevance of HNR** (Werth 2010), correlation with Shimmer
- Quality of attack. **Important information** inside (Orlikoff 2009, Revis 1999), no indicator available.
- Most studies: one second after attack. **Do not include long term**, attack and end of sound



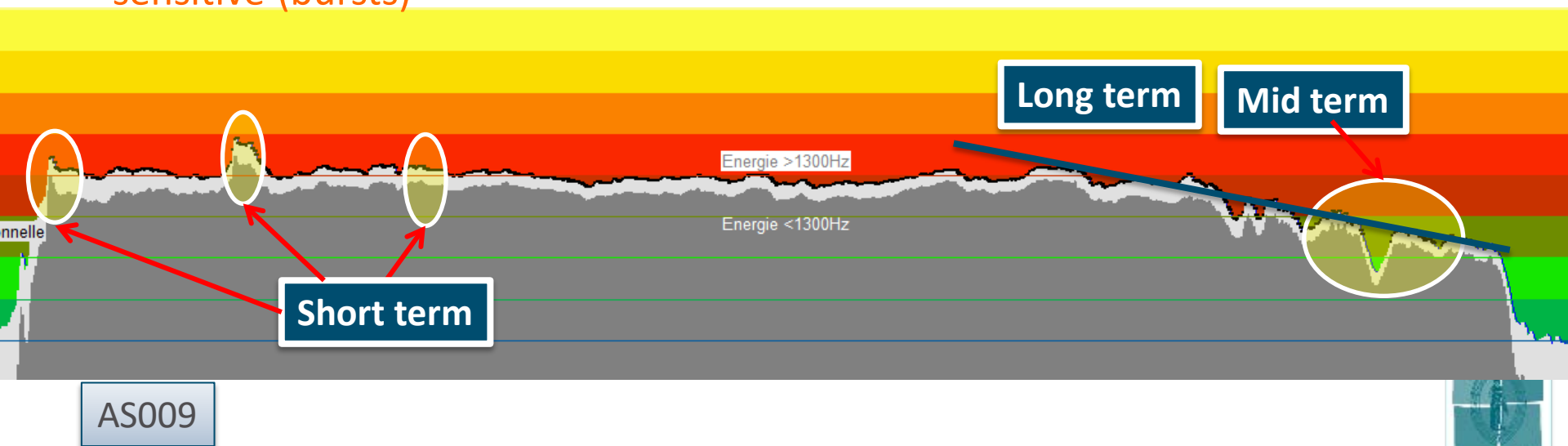
# SHORT, MID, LONG TERM JITTER

- **Short term** jitter is a very rapid variation of pitch. The human ear is **not so sensitive** to such jitter (10ms)
- **Mid term** jitter is a medium variation of pitch, close to fast vibrato. The human ear is **somehow sensitive** to such variation (100ms)
- **Long term** jitter is a slow variation of pitch, which characterizes the trend to decrease of increase pitch over some seconds. The ear is **very sensitive** to such fluctuations



# SHORT, MID, LONG TERM SHIMMER

- **Short term** shimmer is a very rapid variation of amplitude (10ms).
- **Somehow sensitive (bursts)**
- **Mid term** shimmer is a medium variation of amplitude (100ms).  
**Somehow sensitive.**
- **Long term** shimmer characterizes the trend to decrease or increase the amplitude over seconds.  
**Sensitive**



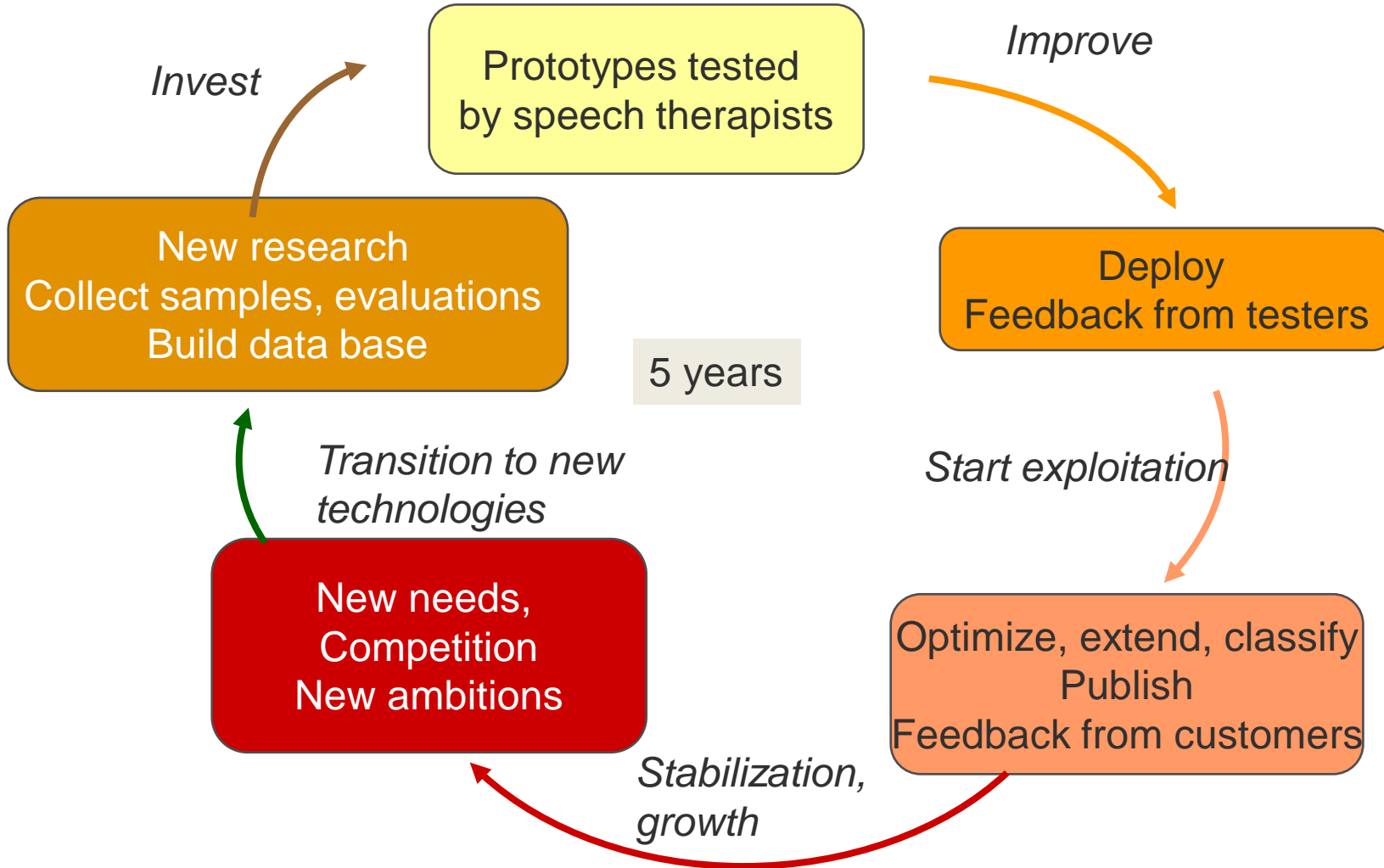
Scanisme / FUM ↓ Passage Méc

Octave 1 Octave 2



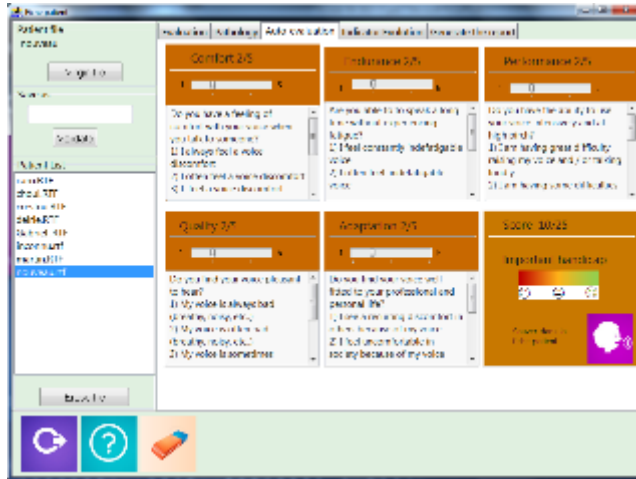
# METHODOLOGY

# METHODOLOGY 5 YEARS CYCLE





### 4 Main Screens



Patient



Evaluate



Therapy



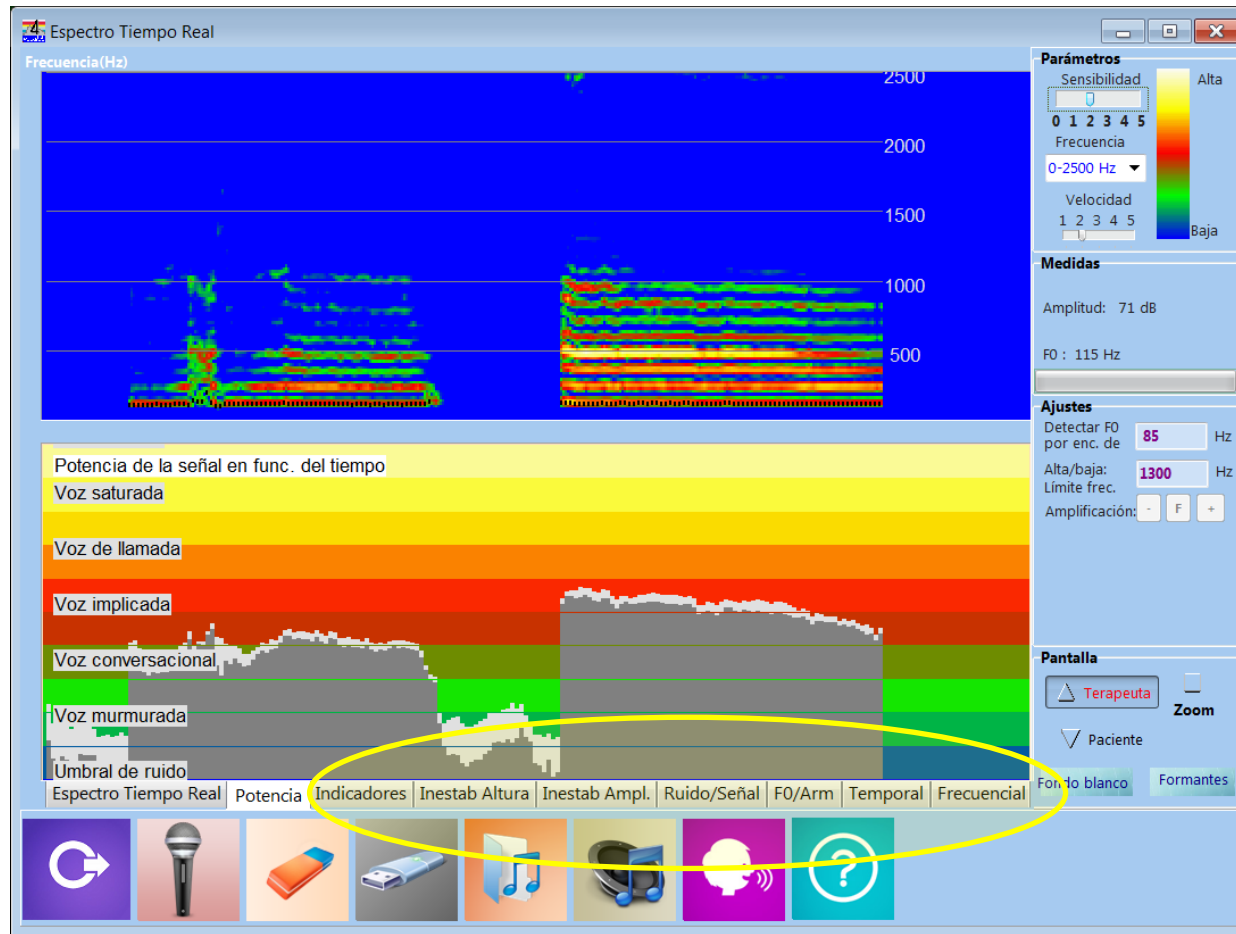
Media Library

Spectrogram

Energy

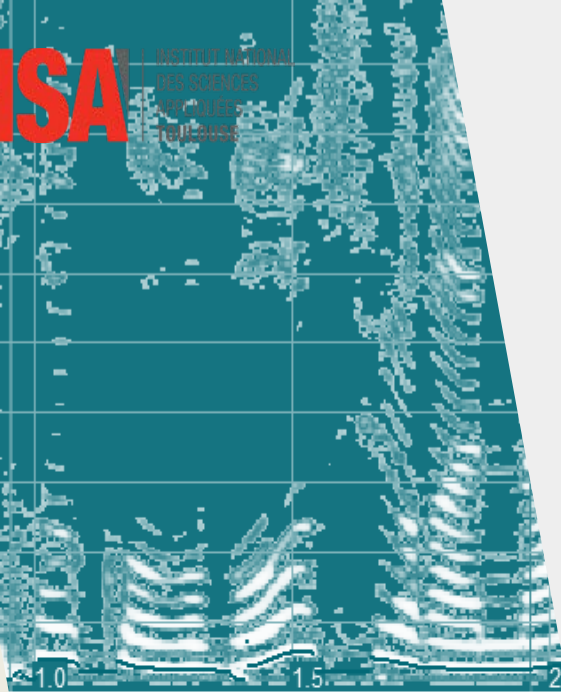
Configuration

Signal processing



Add tools and indicators to provide objective evaluation of voice

2. Software Structure



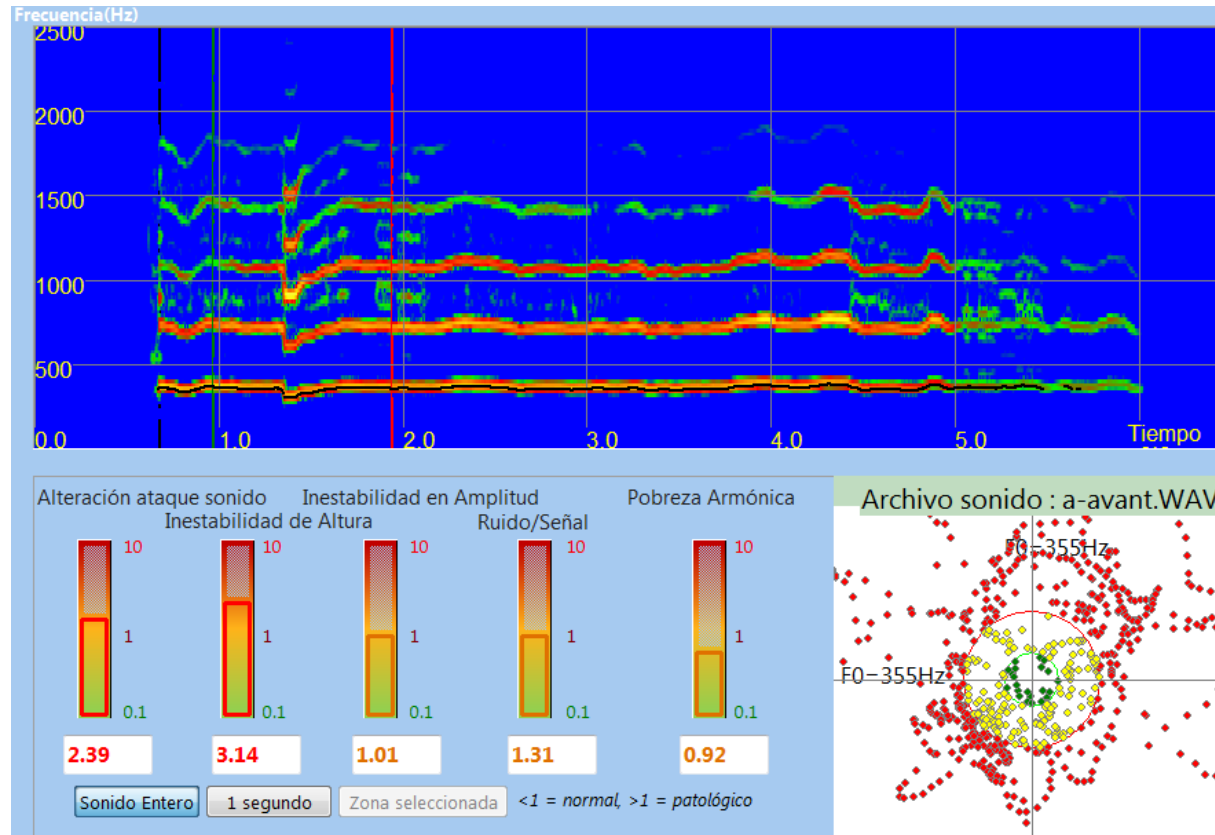
# KEY RESULTS

## Five indicators:

- alteration of the attack
- pitch instability
- amplitude instability
- noise / signal ratio
- harmonic poverty

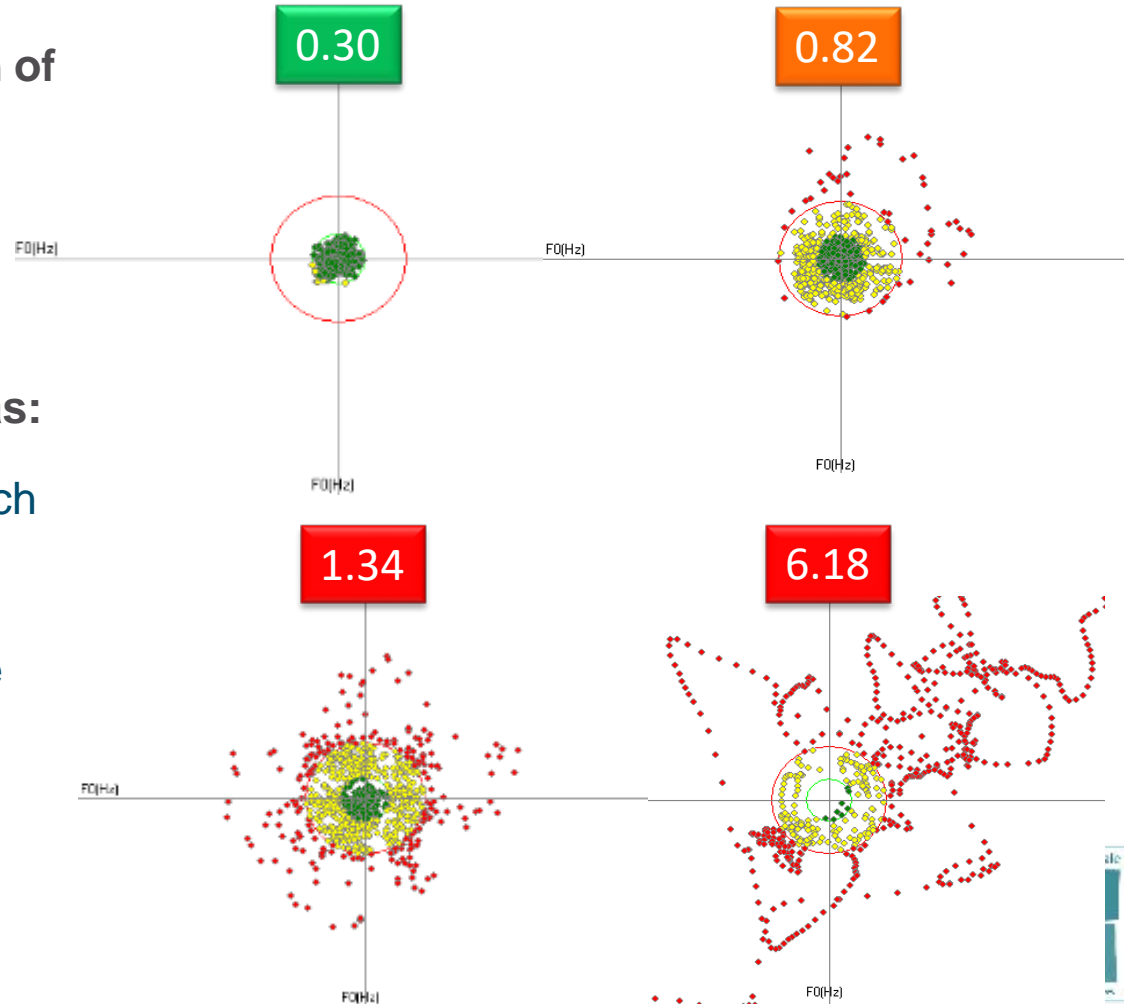
## Normalized

- Threshold **1.0** for all
- Value **<0.8: Green**
- **0.8 <value <1.5: Orange**
- Value **> 1.5: Red**



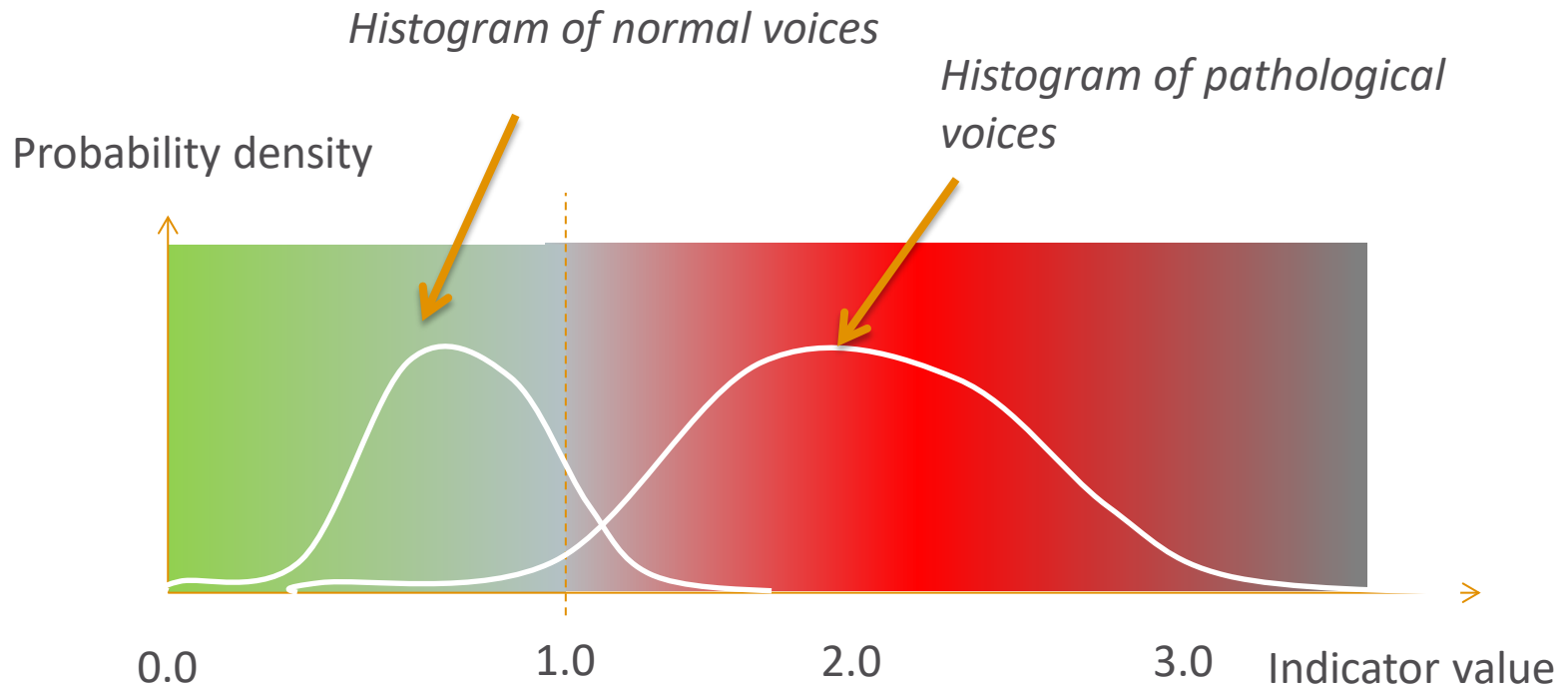
## Phase portrait

- Other graphic representation of pitch instability.
- Inspired from **chaos theory** (Chua, Mira)
- VOCALAB defines three areas:
  - In the center, the stable pitch (**green**)
  - A little distant, a less stable pitch (**yellow**)
  - More distant, chaotic pitch (**red**)

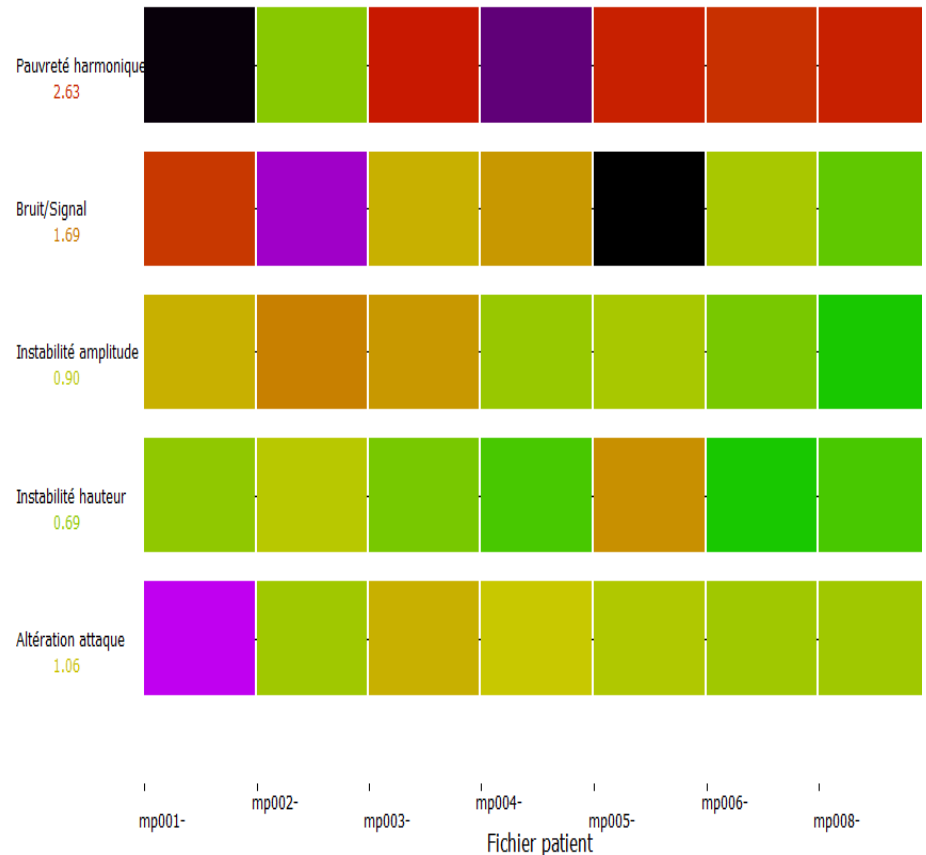


# VOICE SAMPLE DATA BASE NORMAL AND PATHOLOGICAL

- **1000 samples** collected from speech therapists (normal, altered, pathological)
- Subjective **evaluation** of voice quality : attack, jitter, shimmer, noise, harmonics
- **Correlation** optimization with speech therapist evaluation
- **Cross-analysis** with PRAAT



# VOICE SAMPLE DATA BASE VALIDATION



- **Normal voices** have most of indicators in green, sometimes orange. Average is below 1.0

- **Pathologiques** voices have some indicators in red. Some averages are above 1.0 (2.6 harmonic poverty)



# INDICATORS ATTACK ALTERATION

- Evaluates the characteristics within **initial 300 ms.**

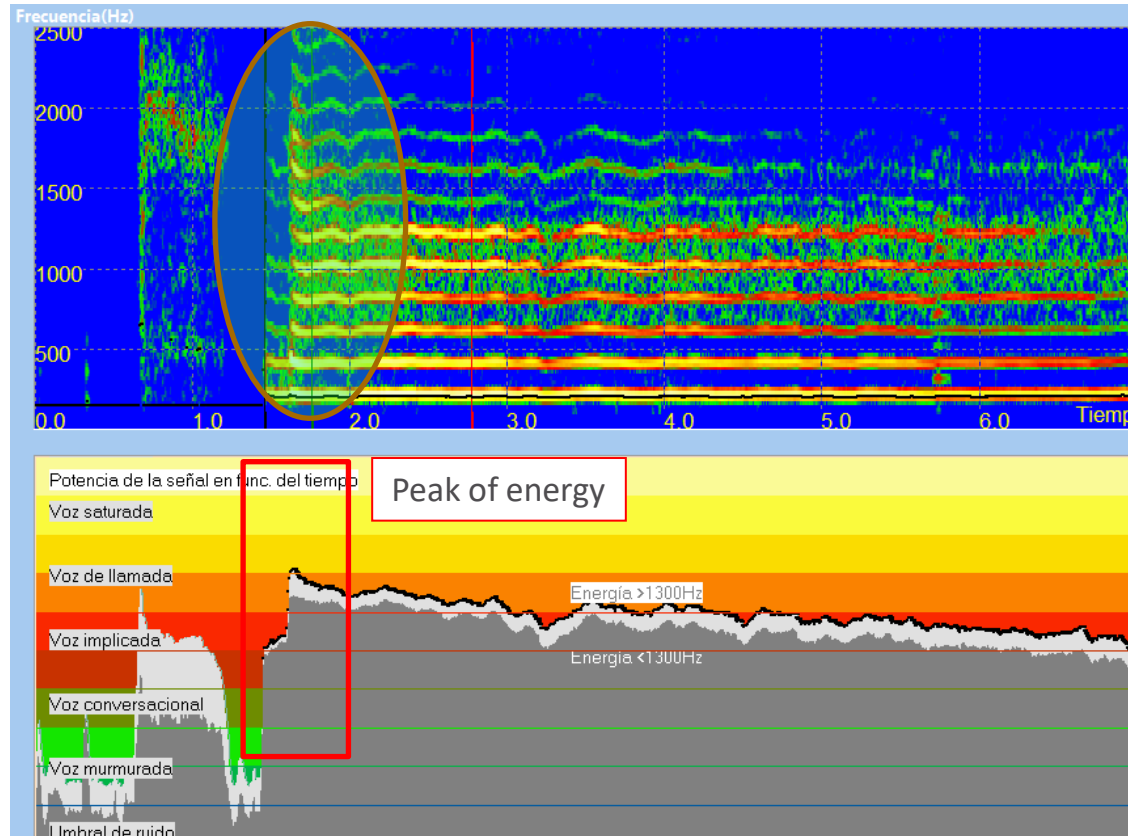
Start: **black** vertical line

End: **green** vertical line.

- The indicator evaluates:

- Pitch** Instability
- Instability in **amplitude** relative to an ideal ramp
- The **noise** from the harmonics of the voice

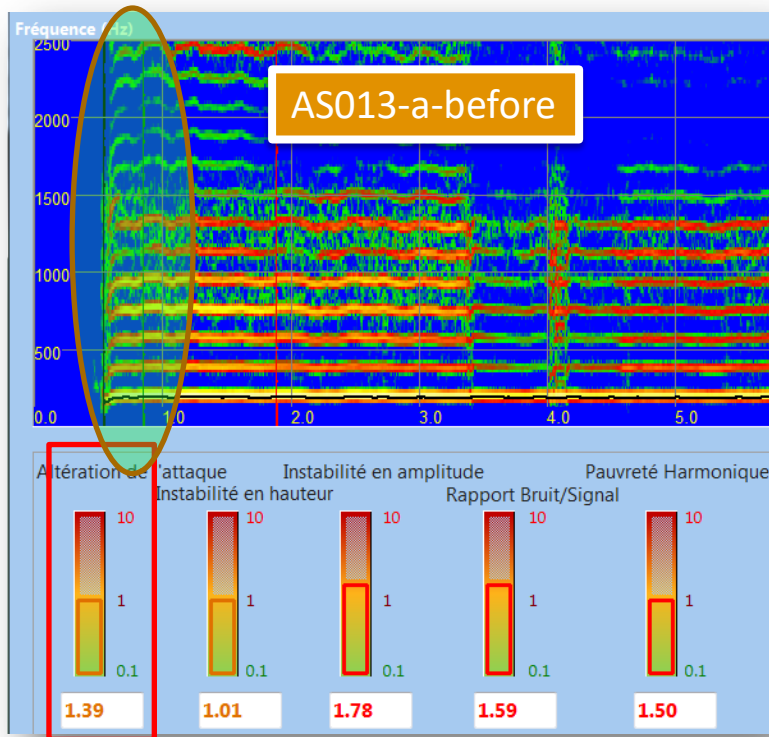
300 ms zone



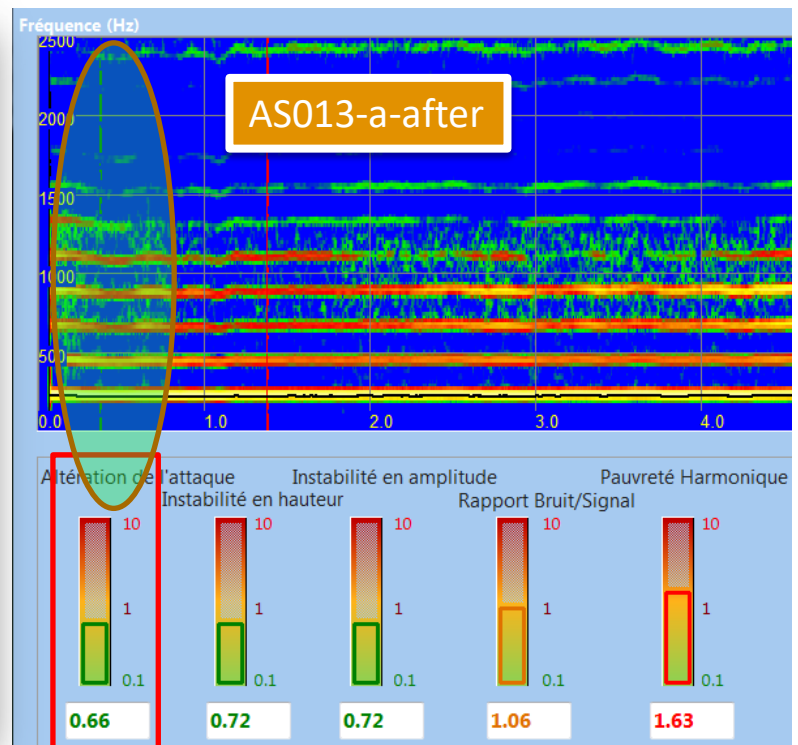
SP031-a-before



## Before/After therapy



**Before therapy**  
 Unstable, hard, noise

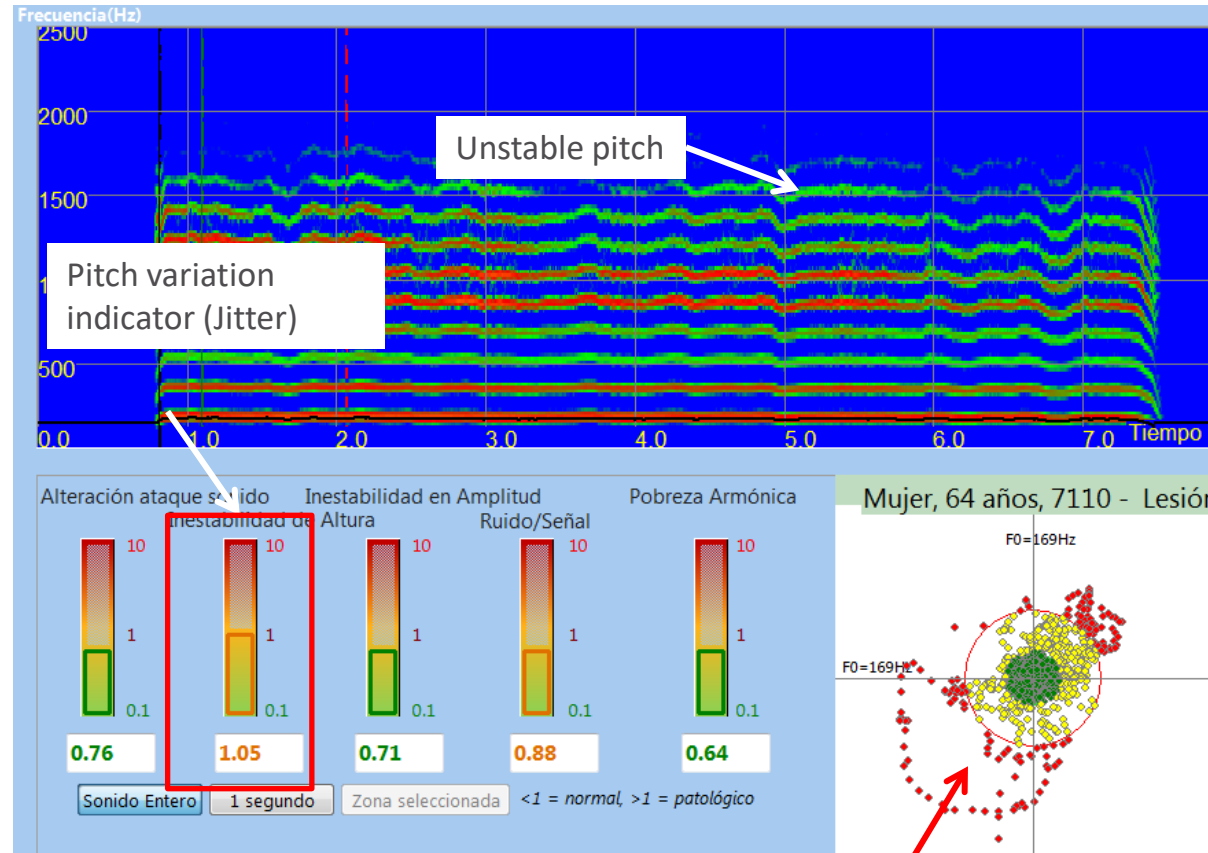


**After therapy**  
 Stable, soft, less noisy

# INDICATORS PITCH INSTABILITY

Short, mid and long term  
Jitter

- Based on the evaluation of F0, which appears as a black horizontal line superposition to the lowest harmonic.
- Visible variations of the fundamental frequency over time.
- Also illustrated by the phase portrait (Center is average F0)



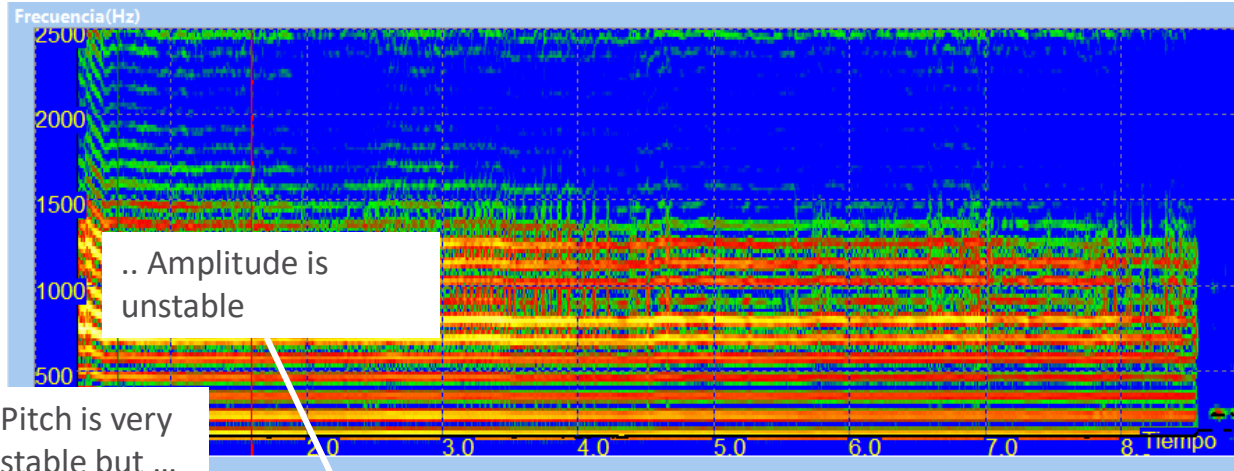
AS016

Phase portrait out of green center

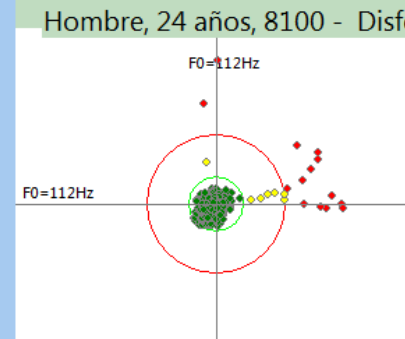
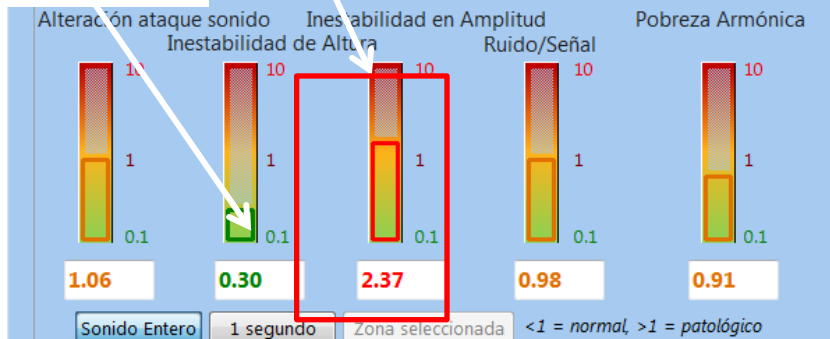
# INDICATORS AMPLITUDE INSTABILITY

Short, mid and long term Shimmer

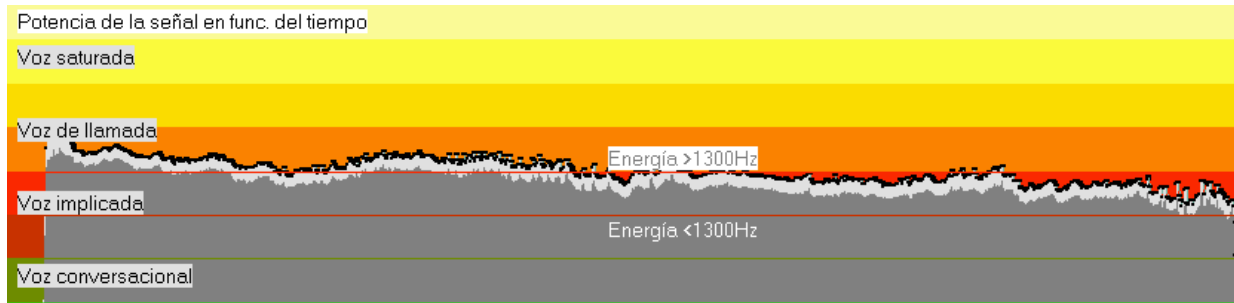
- The color gives the amplitude information in the spectrogram
- **Color variation** means shimmer
- More clear in the **power profile**



Pitch is very stable but ...



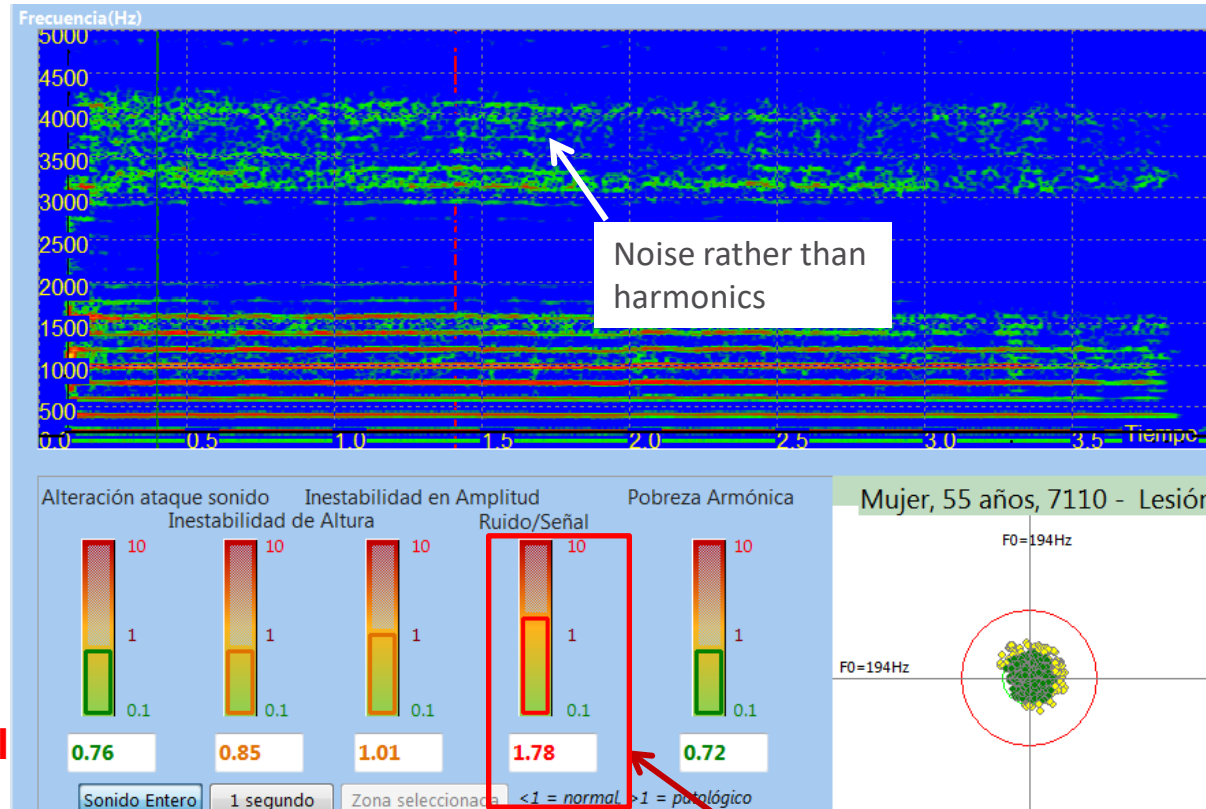
AS014



# INDICATORS NOISE SIGNAL RATIO

## Noise / Signal indicator

- Close to GRBAS “Breathy”
- Noise instead of harmonics, especially visible in the **3-4 KHz** bandwidth
- Noise **between harmonics** around 1KHz
- Noise indicator **above** the threshold (>1.0)
- Others indicators are **normal**

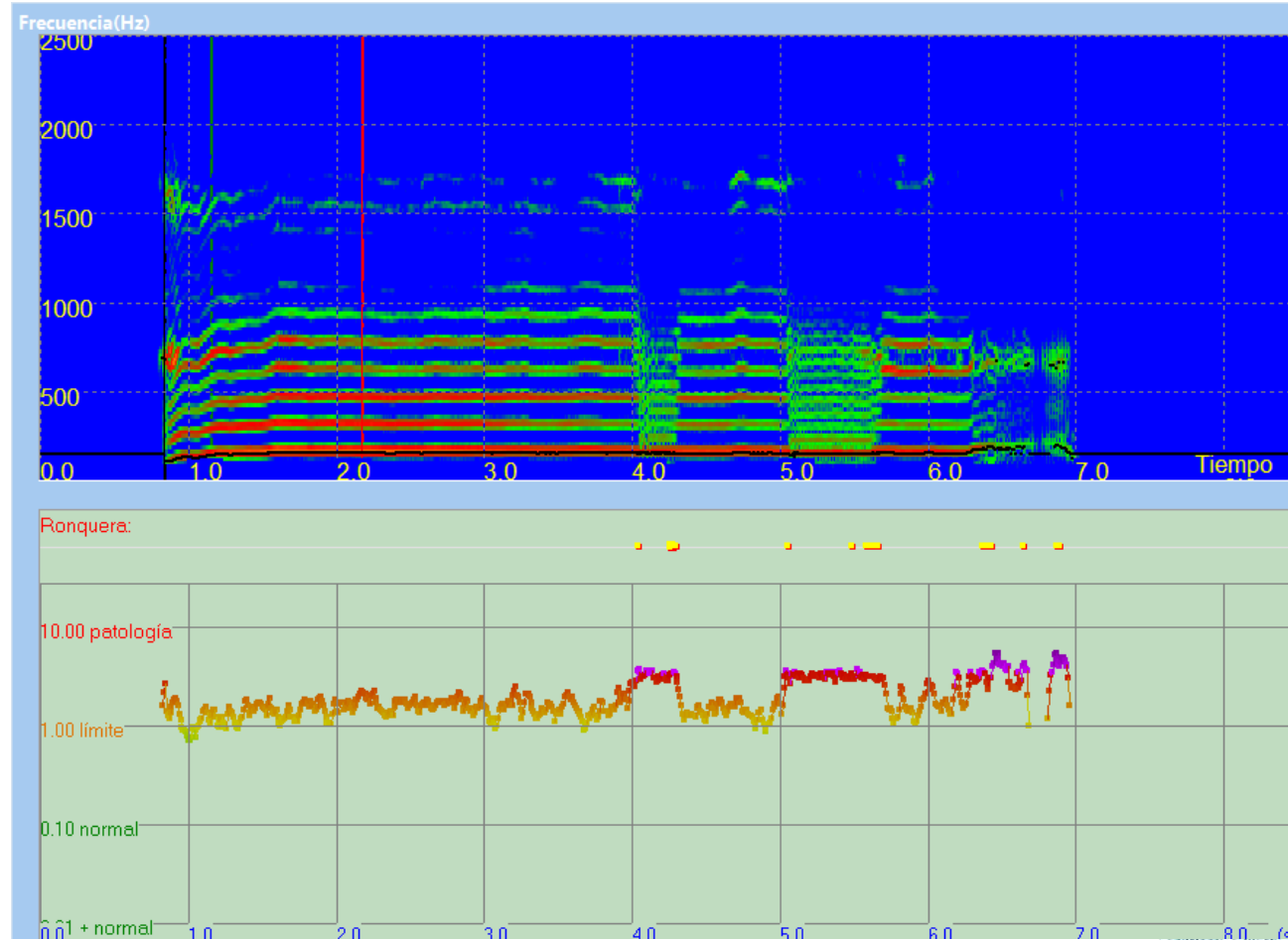


SP136

Noise/Signal Indicator in red

# INDICATORS NOISE SIGNAL RATIO

- Noise/Signal ratio also includes a **detection of Hoarsness**
- Increases NHR significantly when detected



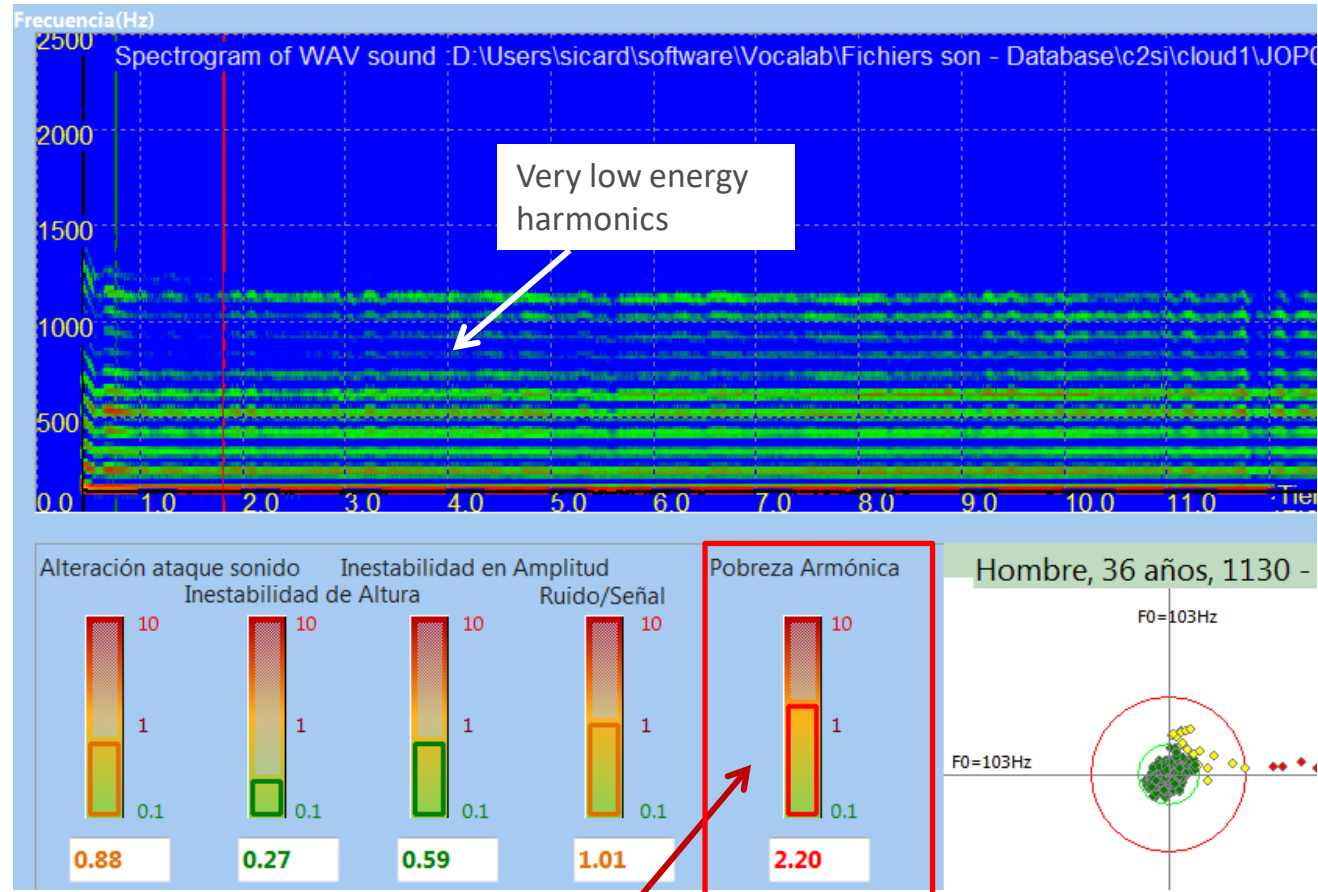
ANC150



# INDICATORS HARMONIC POVERTY

## Harmonic poverty indicator

- Counting of harmonics
- Lack of high-frequency** harmonics rises the indicator
- Noise** instead of harmonics also rises the indicator



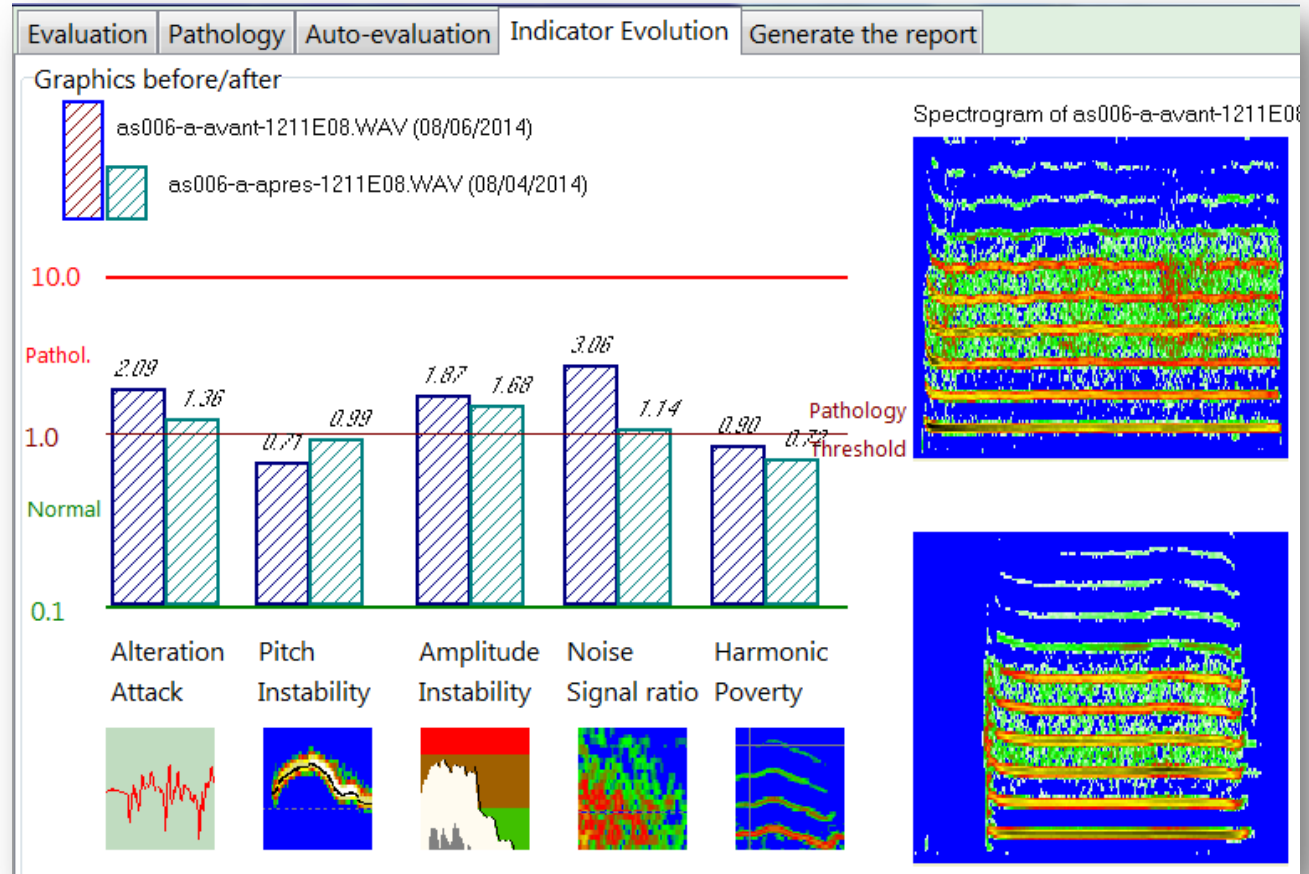
JOP013

Harmonic poverty in **red**

# INDICATORS BEFORE/AFTER THERAPY

AS006 - Child, age 8,  
vocal cord nodule

- After therapy
- **Less noise**
- Global **decrease** of alteration indicators



Sounds on-line at [www.vocalab.org](http://www.vocalab.org) >

## Voice Data Base

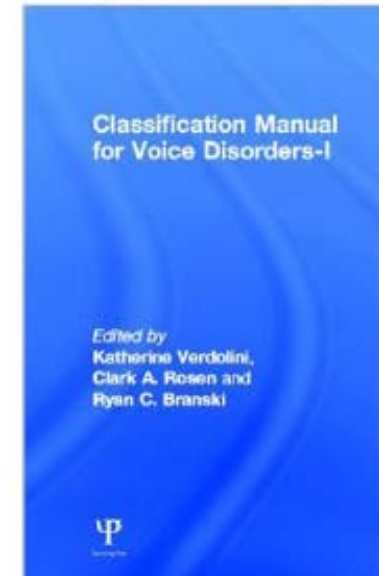
The screenshot shows a Mozilla Firefox browser window displaying a table titled 'Base de données de voix'. The table has the following columns: Pathologie, Code, Nom, âge, H/F, Enfant, /a/, Sirène /a/, /a/ après rééducation, Sirène /a/ après rééducation, Commentaires, and Axes thérapeutiques. The table contains 11 rows of data. Annotations include: 'Comments' pointing to the 'Commentaires' column; 'More details' pointing to the 'Sirène /a/ après rééducation' column; 'Before/After' pointing to the 'Sirène /a/' and 'Sirène /a/ après rééducation' columns; 'Pathology' pointing to the 'Pathologie' column; 'Classification' pointing to the 'Code' column; and 'Speech Therapist' pointing to the 'Nom' column. There are also small icons in the 'Axes thérapeutiques' column for some rows.

Pathologie	Code	Nom	âge	H/F, Enfant	/a/	Sirène /a/	/a/ après rééducation	Sirène /a/ après rééducation	Commentaires	Axes thérapeutiques
Bégaiement	9010	AS003	57	F	a				h hauteur et	
Crico-Hyoido-Epiglotto-Pexie (CHEP)	3330	JCF064	68	H	a				ilité en puissance après.	
cordectomie	1110	SP001	43	H	a	sirène	a	sirène	Bruit, instabilités importants avant rééducation	
dysarthrie	7300	SP002	73	H	a	sirène			Bruit important dans la bande 500-800 Hz	
dysodie	8150	SP074	26	F	a	sirène	a	sirène	Voix quasi normale.	
	8150	SP004	73						instabilités et bruit pathologiques	
	8150	SP005	53						Sirène sur /ou/	
	8150	SP006	53	H	a	sirène			Passage des mécanismes sur la sirène	
aphonie queluche	5310	SP008	64	F	a	sirène			Voix très instable, faible et bruitée. Sirène très peu étendue	
aphonie dysfonctionnelle	8100	SP009	40	F	a		a		Bruit dans la bande 700-1700 Hz	
	8100	AS008	66	F	a				Instabilité en hauteur et zones avec errâtures	
	7110	AS009	64	F	a	sirène	a	sirène	Très forte instabilité avant, sirène incontrôlée.	
	8100	AS011	52	F	a	sirène			Voix instable. Passage brusque du mécanisme L à	



## Classification according to ASHA (USA), from *Classification Manual For Voice Disorders, 2006*, by K. Verdolini

- 1000 - STRUCTURAL DISEASES OF LARYNX
- 2000 - INFLAMMATORY CONDITIONS OF THE LARYNX
- 3000 - TRAUMA, INJURY OR OPERATION OF LARYNX
- 4000 - SYSTEMIC CONDITIONS AFFECTING VOICE
- 5000 - AERODIGESTIVE DISORDERS AFFECTING THE VOICE
- 6000 - PSYCHIATRIC AND PSYCHOLOGICAL DISORDERS
- 7000 - NEUROLOGICAL DISORDERS AFFECTING THE VOICE
- 8000 - OTHER DISORDERS AFFECTING THE VOICE
- 9000 - OTHER VOICE DISORDERS



## Coding to ease patient identification

AS001-a-SHI36-AE12-7120F44.WAV

- AS001 : Speech Therapist Anne Sicard, patient 001,
- -a- : sound /a/
- SHI36 = Speech Handicap Index 36
- AE12 = Self evaluation 12
- 7120 : pathology
- F = women
- 44 = 44 years old

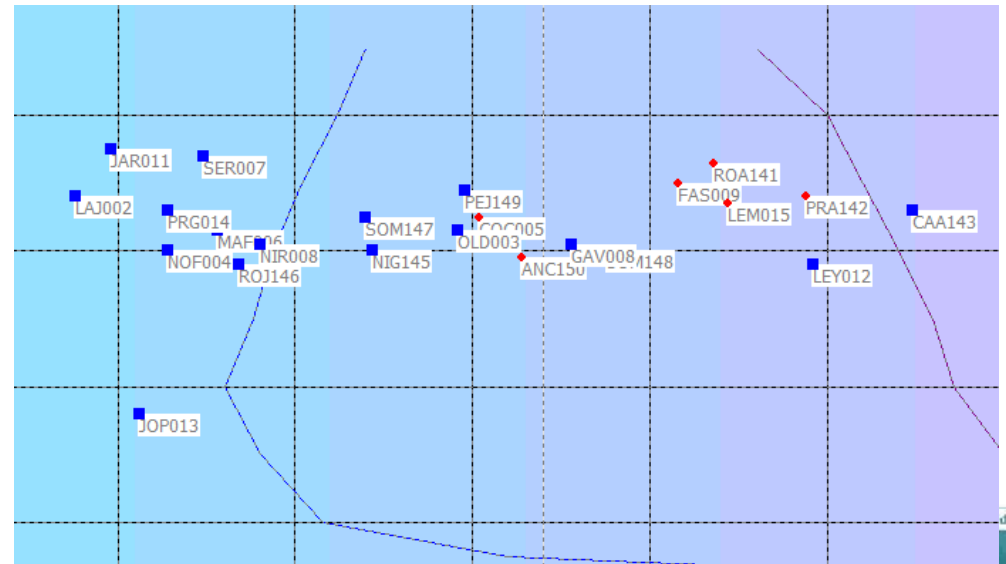
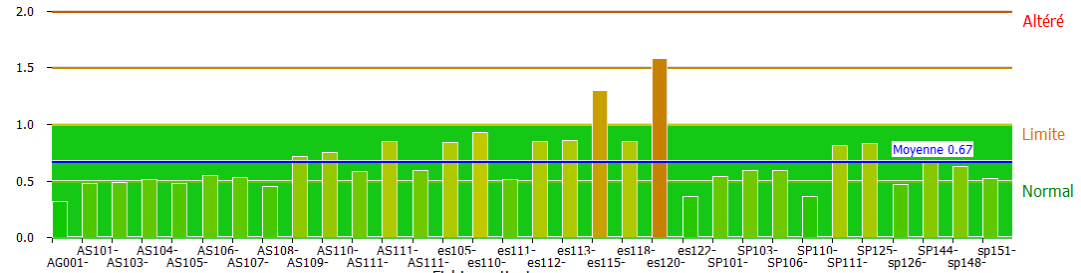
- Since 2014, around 10 training sessions have been organized in France, Belgium, Switzerland, Spain...
- Indicators help speech therapists to exploit spectrogram information
- On-line data-base eases illustration of each concept
- Daily speech therapy benefits from these tools



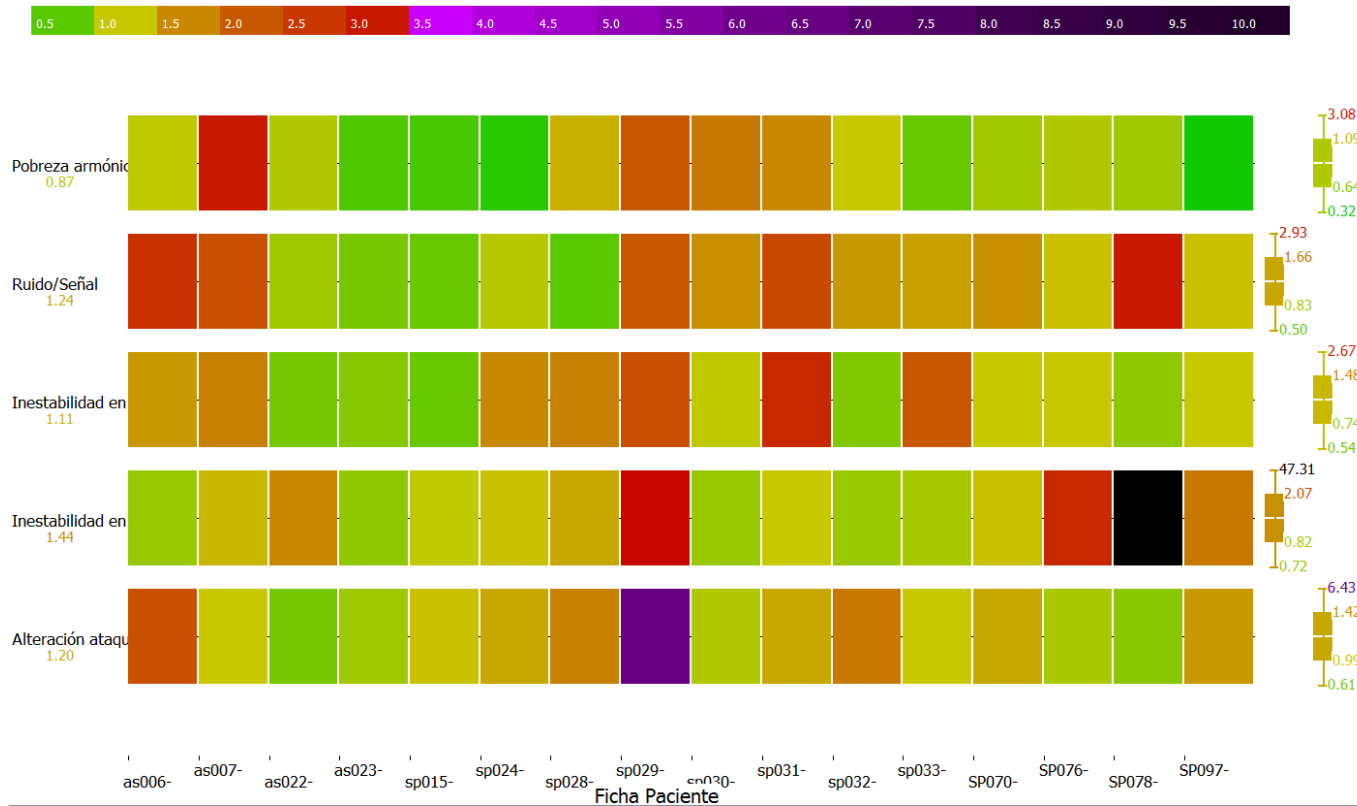
# STATISTICS

# STATISTICS ILLUSTRATING TRENDS

- Speech therapy in France is now a **5-years Master** program, instead of 4
- Statistic analysis very common in **Master thesis**
- Typically: normal vs patients
- Voice statistics help to extract trends, correlations, etc..



- Most indicator averages above threshold
- Highest alteration :  
pitch instability  
Noise/signal ratio



Master Report of Marie Daumet, Speech Therapist, Nice, 2015

# STATISTICS ON GROUPS OF PATIENTS

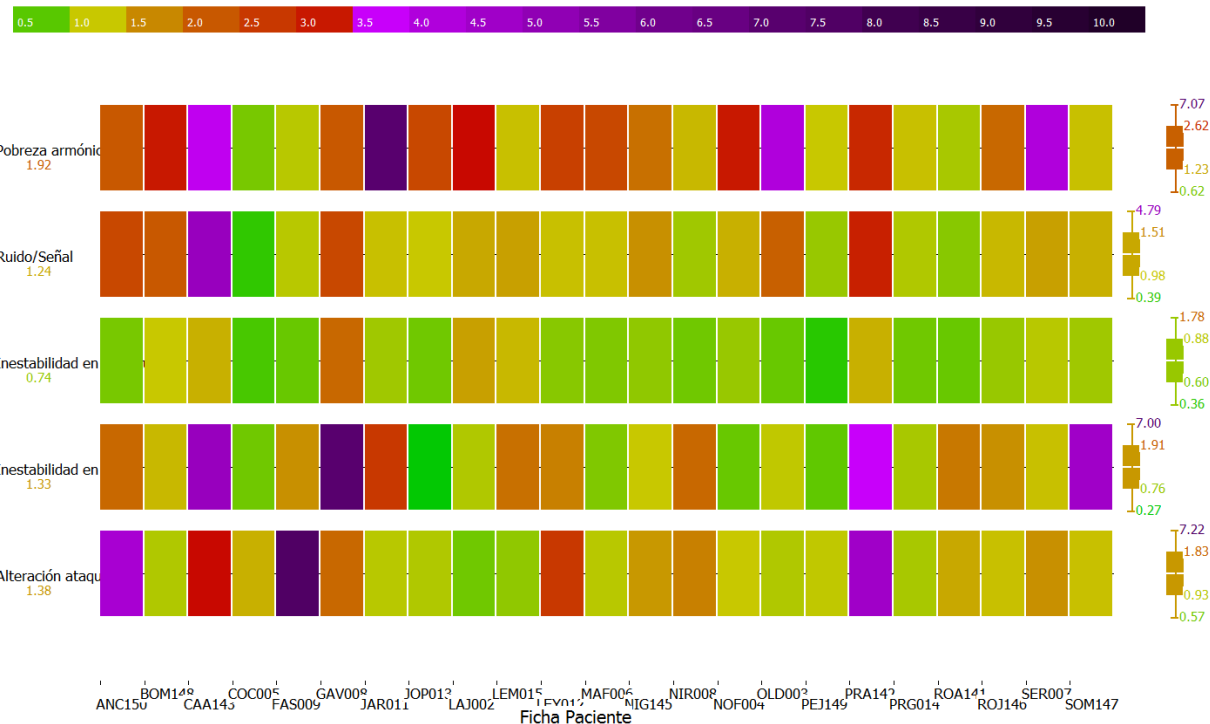
## CANCER VOICE PROJECT C2SI

- **Project C2SI :**  
intelligibility of patients after cancer treatment (/a/, siren, speech...)
- Statistical tools on indicators, spectrum, F0
- Most significant alteration

Harmonic poverty

Noise/Harmonic ratio

Pitch Instability



# STATISTICS ON GROUPS OF PATIENTS

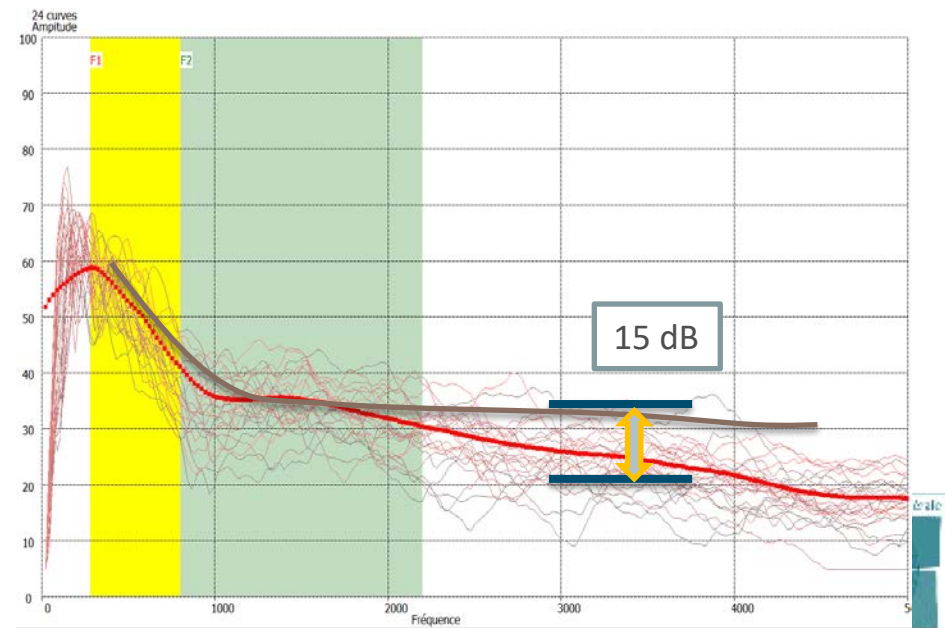
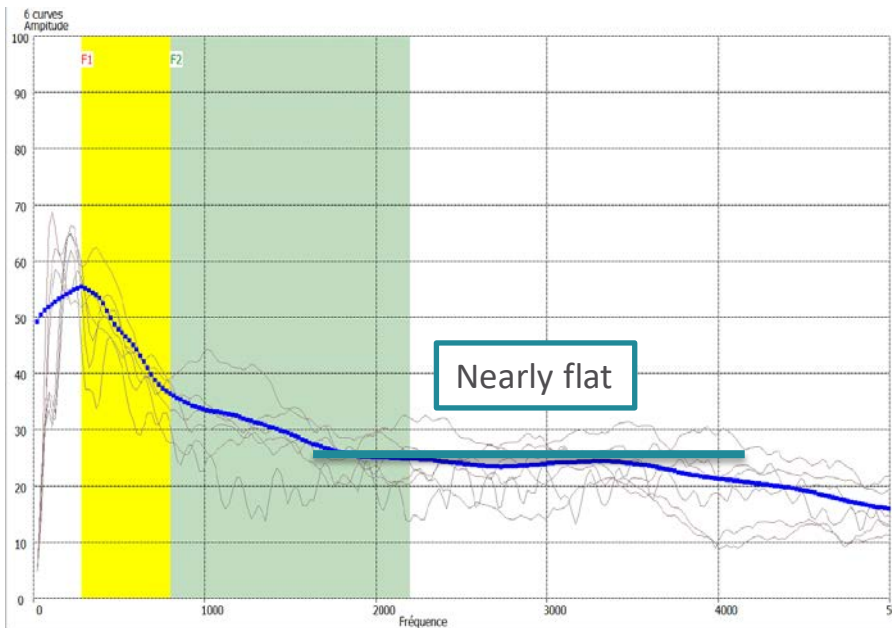
## CANCER VOICE PROJECT C2SI

- **Long-term Average Speech Spectrum (LTASS)**

Spectrograms of text reading sound files

6 Reference voices: nearly flat 1800-4000 Hz

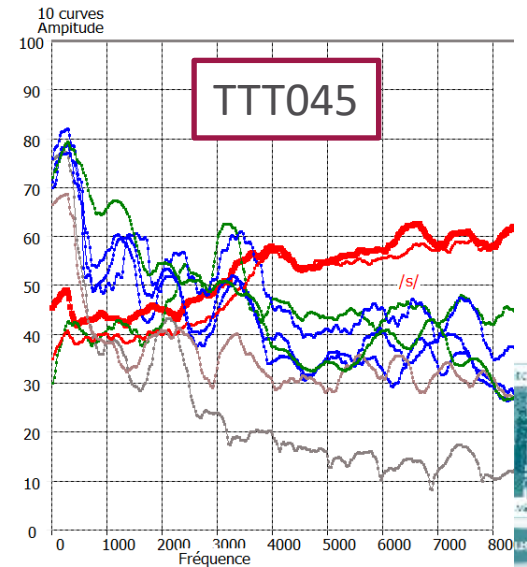
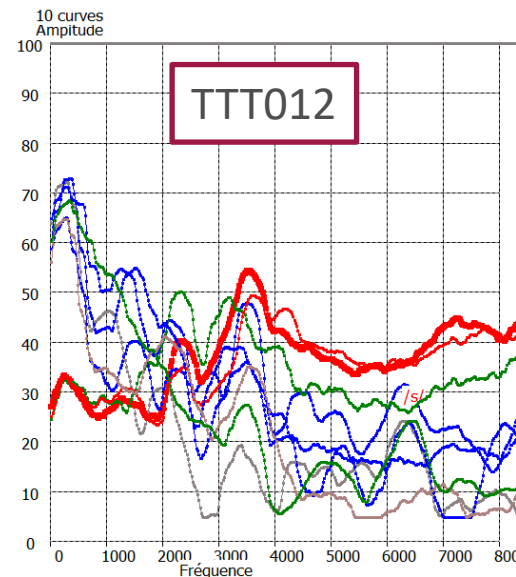
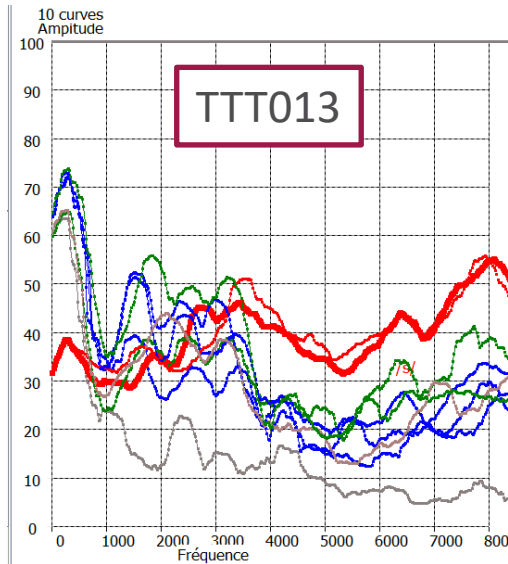
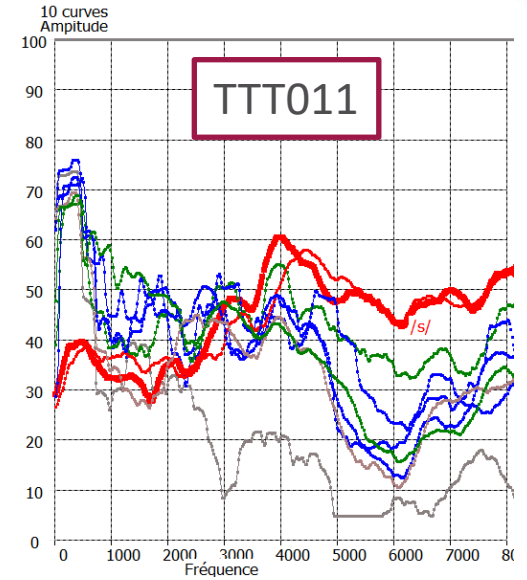
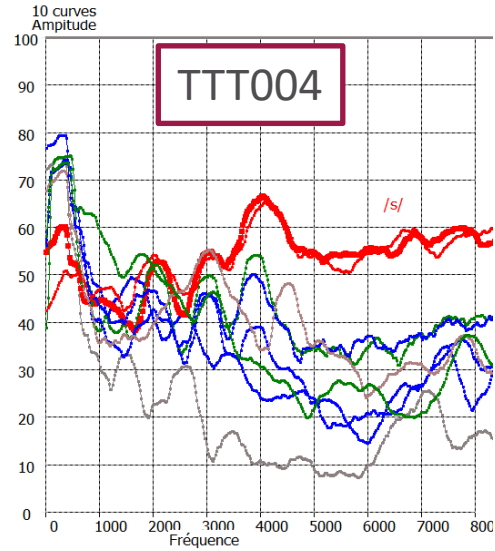
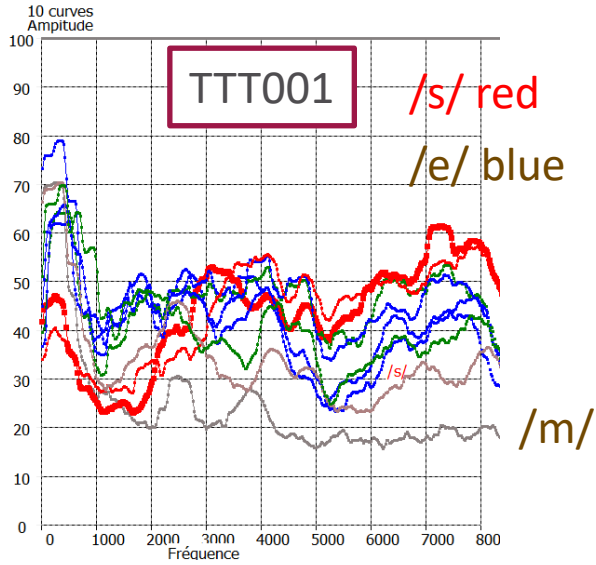
24 Patients's voice: -15dB in F3-F4 formants





# STATISTICS ON GROUPS OF PATIENTS

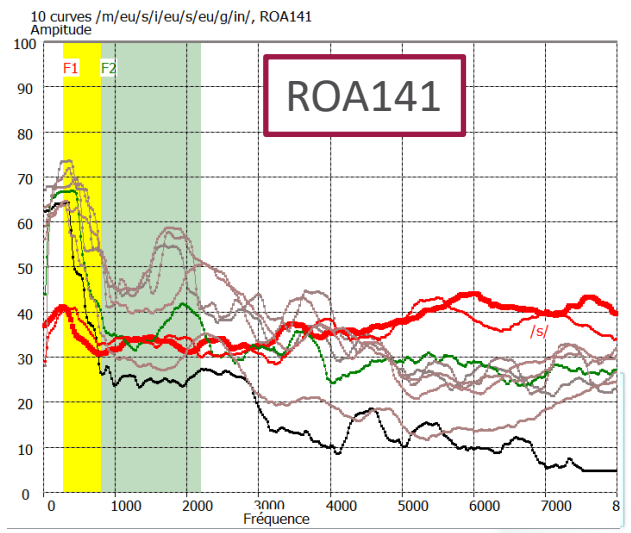
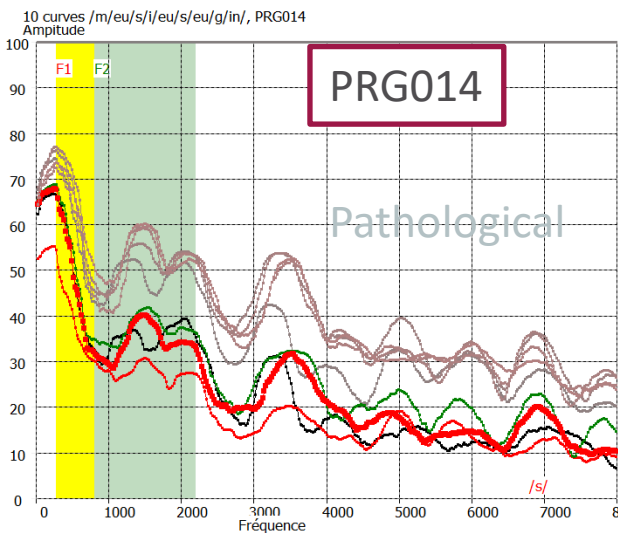
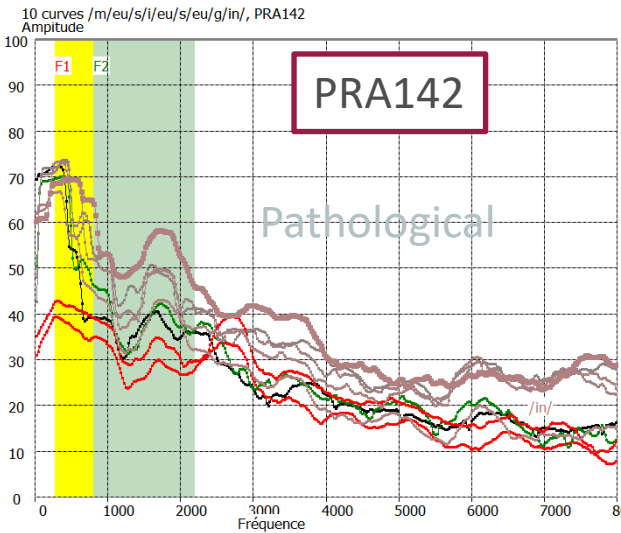
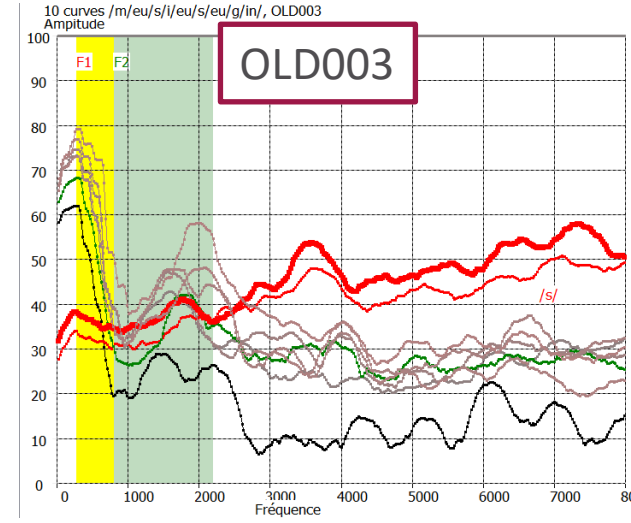
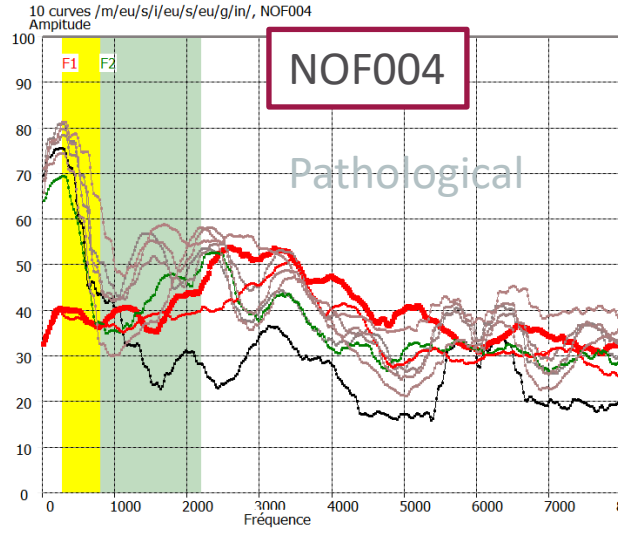
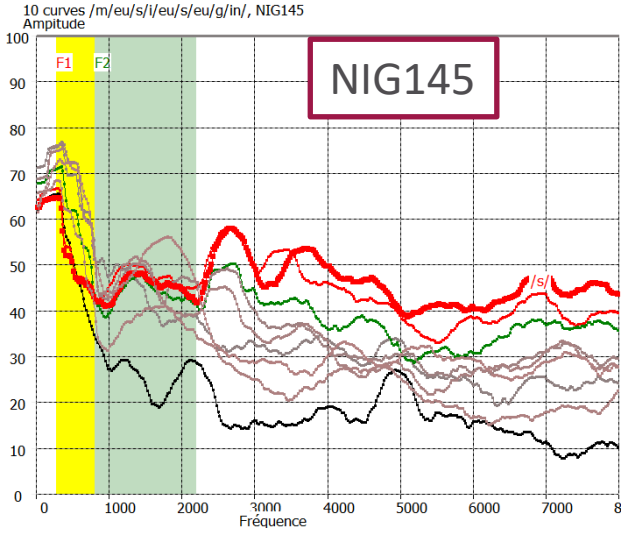
## VOWELS NORMAL SPEAKERS

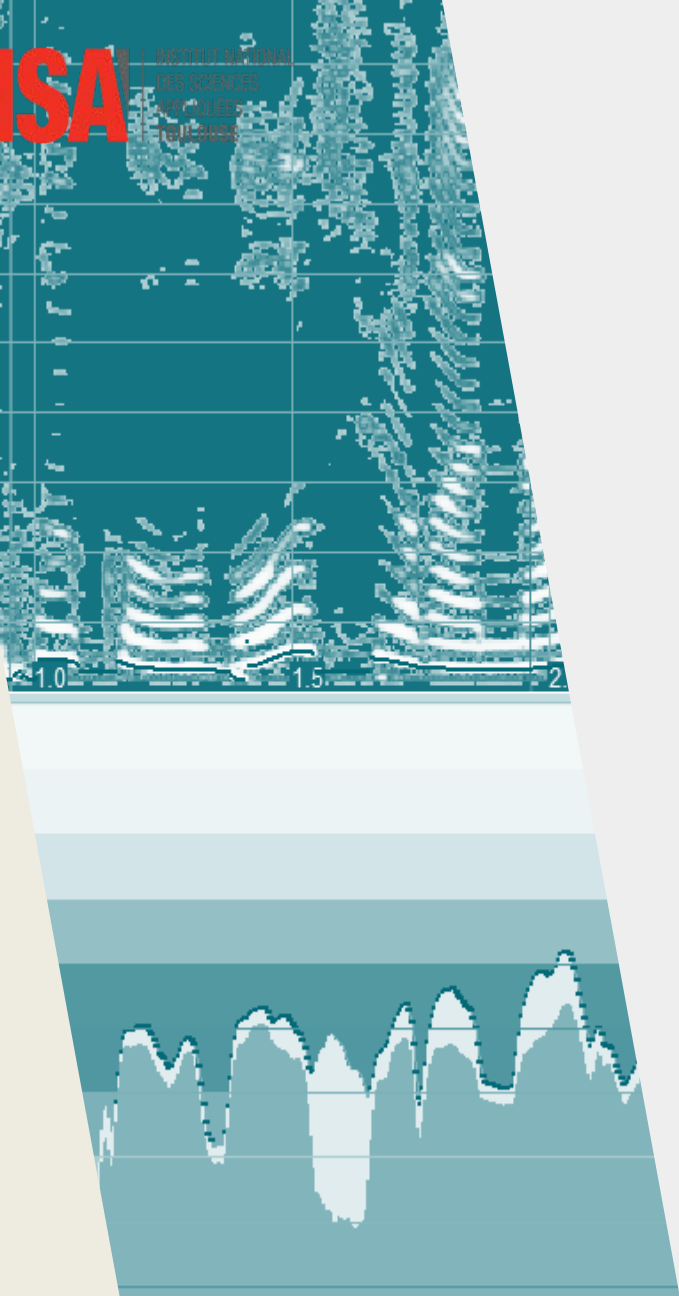




# STATISTICS ON GROUPS OF PATIENTS

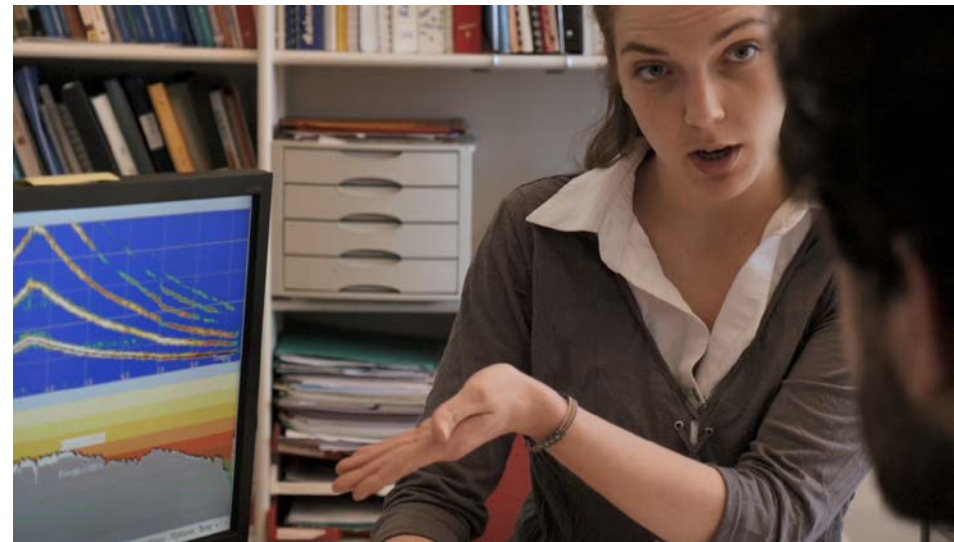
## VOWELS C2SI PATIENTS





# CONCLUSION

- A methodology for voice evaluation has been proposed
- The VOCALAB tool has been developed targeting speech therapy
- Voice alteration indicators include some novel approaches
- High degree of acceptance among French-speaking speech therapist
- Statistics on groups of patients successfully used by Master students



Scanisme / FUM ↓ Passage Méc

Octave 1

Octave 2

THANK YOU FOR YOUR  
ATTENTION