

La identificación forense de voces sustentada en bases de datos locales, regionales e internacionales.

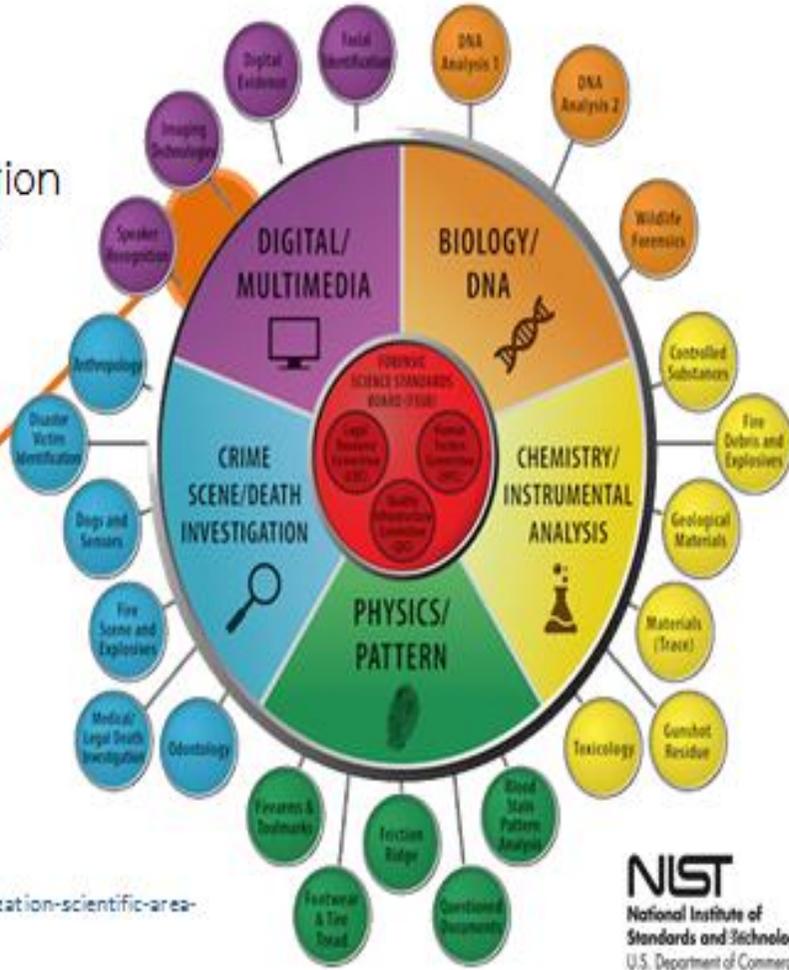
Jorge Gurlekian

Laboratorio de Investigaciones Sensoriales (LIS)

jgurlekian@hospitaldeclinicas.uba.ar

- Diciembre 7, 2016 -

OSAC: Organization of Scientific Area Committees for Forensic Science



OSAC: > 500 miembros y >25 aéreas científicas.

SWGDOC y SWGMAT. FBI

ENFSI: Representa a 33 países del consejo europeo de redes de instituciones de la Ciencia Forense.

SAG: Australia y NZ

UKAS: Gran Bretaña

IRCGN: Francia

IAFP: Internacional, Phonetics and Acoustics

IAFL: Internacional, Lingüística

<https://www.nist.gov/forensics/organization-scientific-area-committees-forensic-science>

Identificación forense de hablantes

Es decidir si la voz de una persona no conocida (evidencia, dubitada, “offender” en inglés) corresponde o no a la voz de una persona que es conocida (indubitada, sospechoso).



“Conseguime la guita”



Identificación Forense

Enfoques empleados para la identificación forense de hablantes:

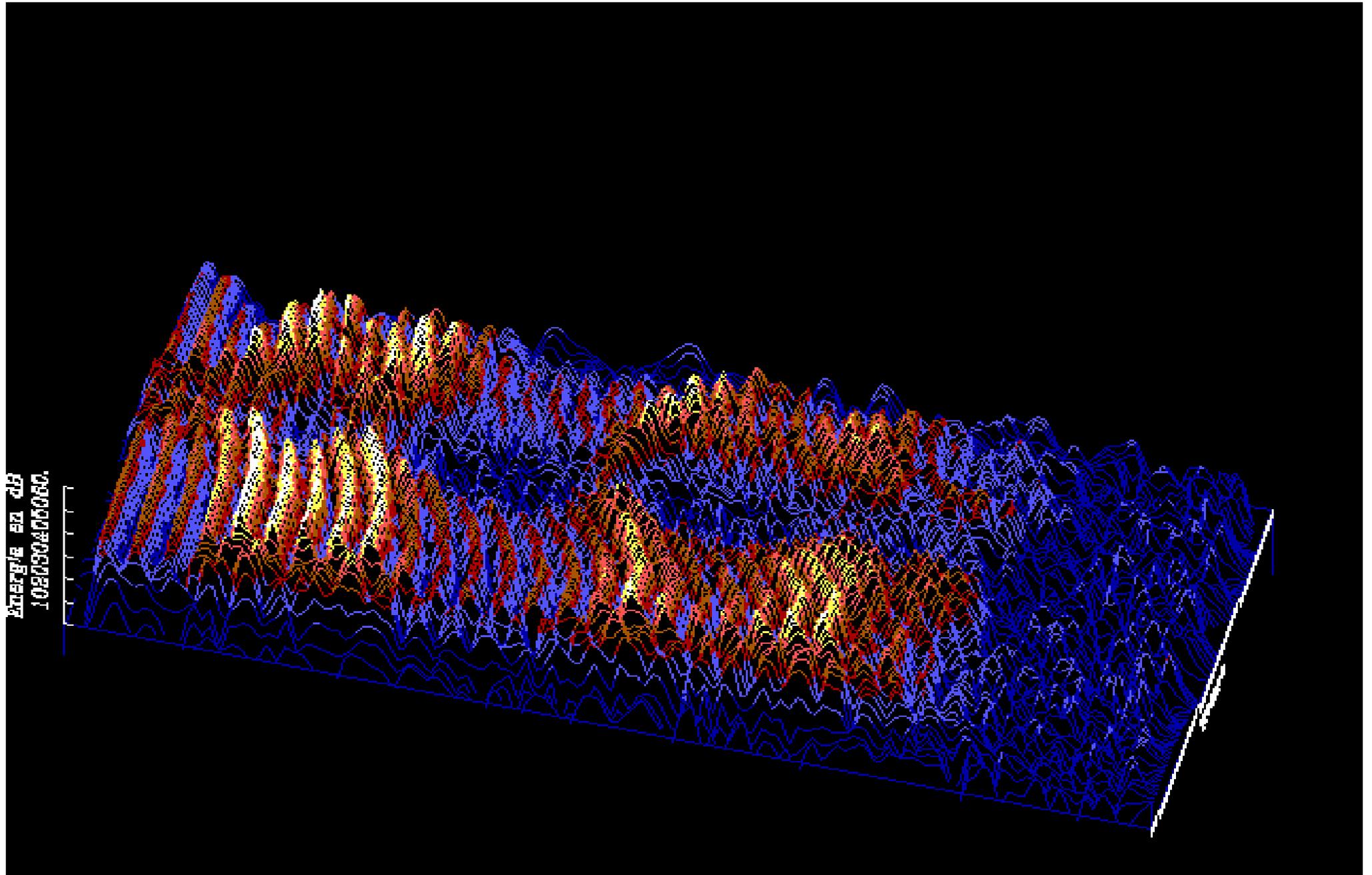
1. Análisis espectrográfico-auditivo
2. Análisis auditivo-perceptual
3. Análisis fonético-acústico
4. Métodos automáticos
5. Métodos semi-automáticos
6. Métodos combinados

¿Donde interviene el experto humano?

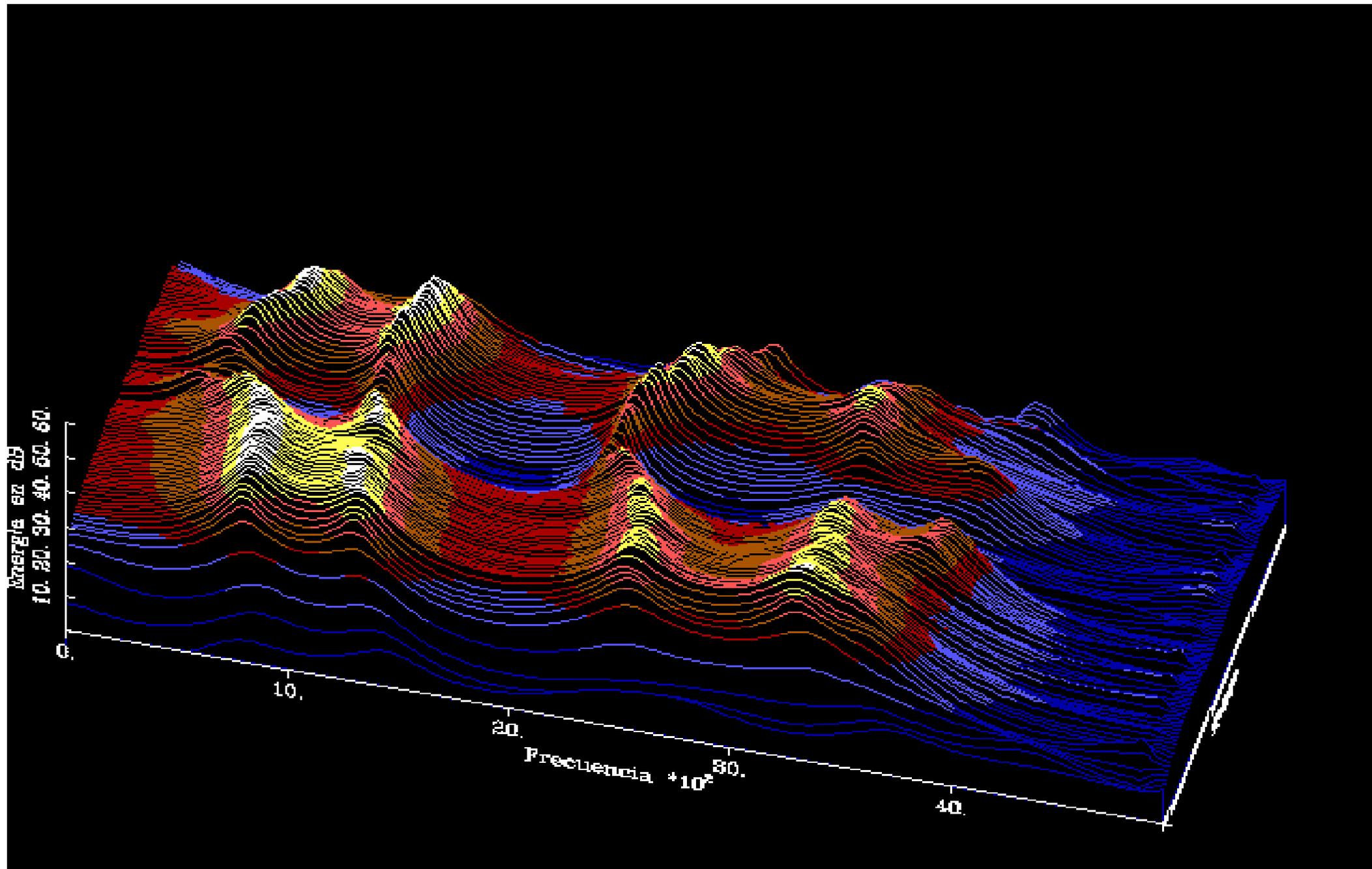
Diseña el vector de características, Anota la variabilidad con la transcripción asignando etiquetas a eventos fonéticos y acústicos

Equipo:

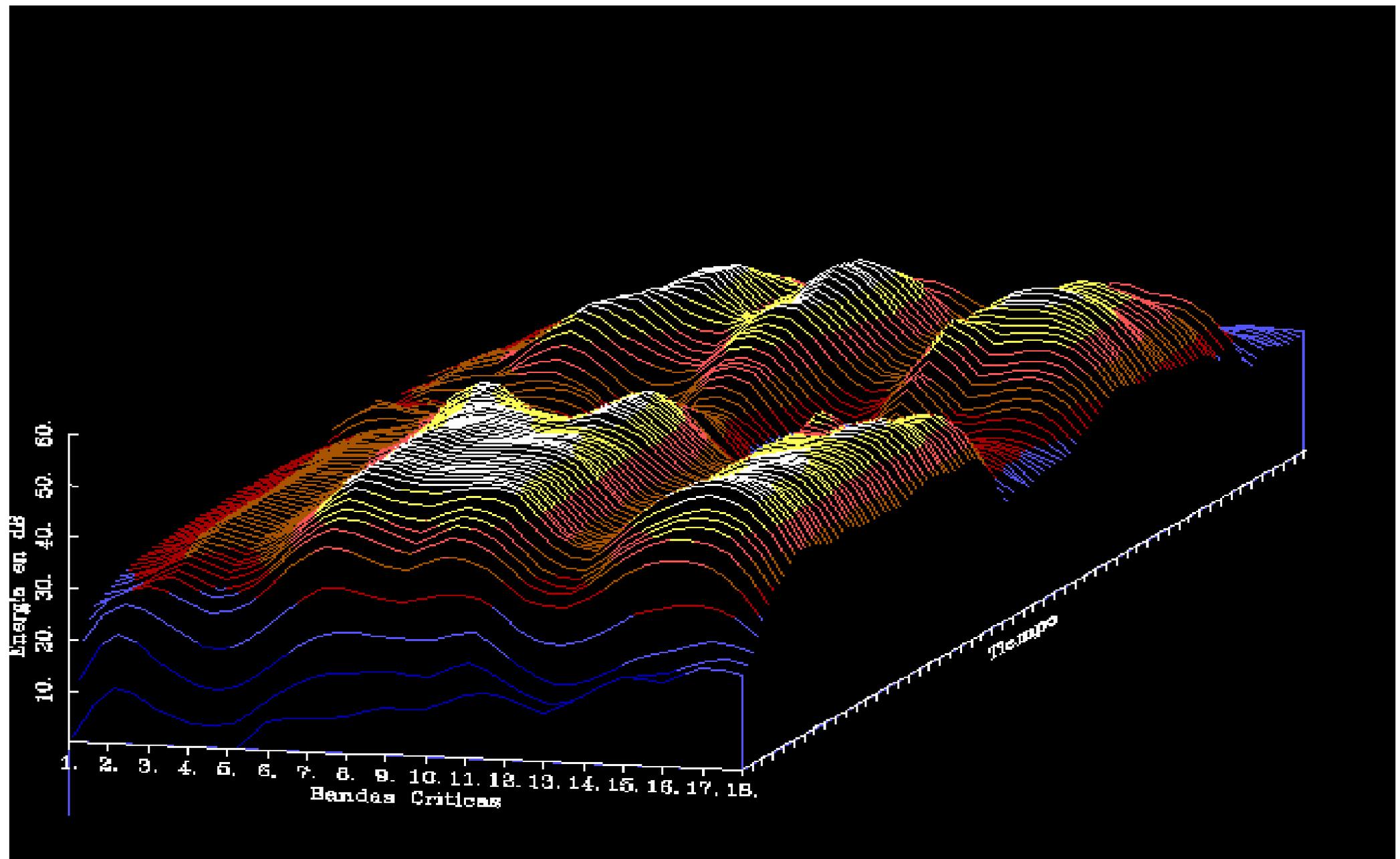
- Ingeniero
- Lingüista
- Estadístico
- Fonaudiólogo



Espectro 3D de Armónicos



Espectro 3D de LPC



Espectro 3D de Bandas Críticas

FORENSIC COMMUNICATION ASSOCIATES

Case Name:

FCA REF:

Aural-perceptual Approach to Speaker Identification Score Sheet

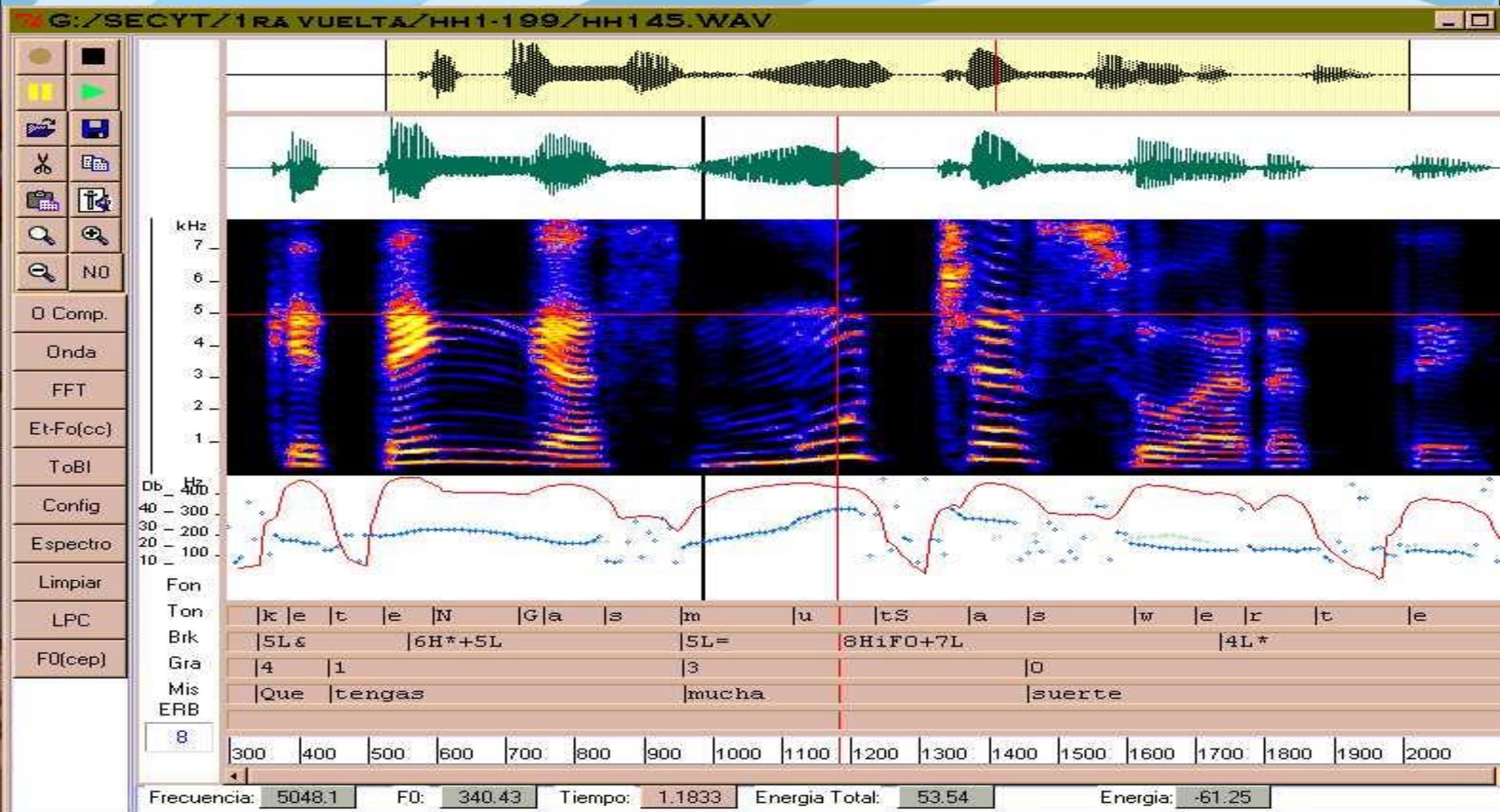
0 = U-K least alike; 10 = U-K most alike

		SCORE	RANGE
1.	PITCH		
	a. Level	0 5 10	
	b. Variability	0 5 10	
	c. Patterns	0 5 10	
2.	VOICE QUALITY		
	a. General	0 5 10	
	b. Vocal Fry	0 5 10	
	c. Other	0 5 10	
3.	INTENSITY		
	a. Variability	0 5 10	
4.	DIALECT		
	a. Regional	0 5 10	
	b. Foreign	0 5 10	
	c. Idiolect	0 5 10	
5.	ARTICULATION		
	a. Vowels	0 5 10	
	b. Consonants	0 5 10	
	c. Misarticulations	0 5 10	
	d. Nasality	0 5 10	
6.	PROSODY		
	a. Rate	0 5 10	
	b. Speech Bursts	0 5 10	
	c. Other	0 5 10	
7.	OTHER		
	a. Nonfluencies	0 5 10	
	b. Speech Disorders	0 5 10	
	c. Other	0 5 10	

Etapas de un método de identificación forense de hablantes

1. Recepción del material - cadena de custodia
2. Pre-procesamiento - formato y segmentación
3. Extracción de parámetros – Segmentales y Supra
4. Modelado de los hablantes - i-vectors y PLDA
5. Cálculo del cociente de verosimilitud – LR , UBM
6. Calibración y Fusión. BD Regresión Logística
7. Validación del sistema de identificación. BD y C_{llr}
8. Análisis de los resultados
9. Informe

Primer Simposio Nacional de Ciencia y Justicia



Vector de características

LPC

MFCC

MLLR Max Likelihood Linear Regression (Stolcke et al 2005)

Energía, delta (diferencia), delta delta (aceleración) (Reynolds, 2008),

shifted delta

Duración de las ventanas de análisis

Prosodia:

Snerf (onset núcleo y coda) N-grams

Parametros de Fujisaki

Otros....

Etapas de un método de identificación forense de hablantes

1. Recepción del material - cadena de custodia
2. Pre-procesamiento - formato y segmentación
3. Extracción de parámetros – Segmentales y Supra
4. Modelado GMM-UBM-SVM, i-vectors y PLDA
5. Cálculo del cociente de verosimilitud – LR , UBM
6. Calibración y Fusión. BD Regresión Logística
7. Validación del sistema de identificación. BD y C_{llr}
8. Análisis de los resultados
9. Informe

El problema de la variabilidad intrahablante

Ambientales

- Ruido Ambiente
- Variabilidad de Canales
- Micrófonos

Extralinguísticos

- Condición Física
- Emociones
- Edad

Lingüísticas/ Paralingüísticas

- Idioma
- Acentos
- Nivel Sociolingüístico

Variabilidad intrahablante debida al canal y al ambiente

- I-Vectors (Dehak et al. 2009)

Se modela con una sólo matriz de variabilidad total:

Este modelo lleva los supervectores de rasgos distintivos a un espacio intermedio más pequeño para aplicar la compensación más adelante.

¿Por qué lo llamamos i-vector?

- Cada mezcla tiene 40 componentes en la media
- En un modelo de 512 mezclas hay 20480 componentes
- El modelo de variabilidad total reduce el espacio a una representación intermedia de 400-600 dimensiones

Normalización de los i-vectors: PLDA

- Una vez que pasamos a la representación intermedia podemos separar la variabilidad entre locutores de la variabilidad restante usando **PLDA** (Prince & Elder 2007). Se requiere una base de datos relevante
- Además, no hace falta etiquetar datos para entrenar

Etapas de un método de identificación forense de hablantes

1. Recepción del material - cadena de custodia
2. Pre-procesamiento - formato y segmentación
3. Extracción de parámetros – Segmentales y Supra
4. Modelado: GMM-UBM-SVM, i-vectors y PLDA
5. Cálculo del cociente de verosimilitud – LR , LLR
6. Calibración y Fusión. BD Regresión Logística
7. Validación del sistema de identificación. BD y C_{llr}
8. Análisis de los resultados
9. Informe

Aproximación Bayesiana

- Hipótesis del método directo:

H1: E **es** del hablante sospechoso

H0: E **no es** del hablante sospechoso

$$p(H_1|E) = \frac{p(E|H_1) p(H_1)}{p(E)}$$

$$p(H_0|E) = \frac{p(E|H_0) p(H_0)}{p(E)}$$

Relación de Verosimilitud

(Likelihood Ratio)

$$LR = \frac{p(E|H_1)}{p(E|H_0)}$$

$LR > \theta$ aceptar H_1

$LR < \theta$ aceptar H_0

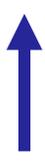
$$\frac{p(H_1|E)}{p(H_0|E)} = \frac{p(E|H_1)}{p(E|H_0)} \times \frac{p(H_1)}{p(H_0)}$$



Corte



Perito



Corte

LR es una declaración de la fuerza de la evidencia

Log LR

LR	LLR	Interpretación
0,0001	-4	Falta de certeza muy fuerte
0,001	-3	Falta de certeza fuerte
0,01	-2	Falta de certeza moderada
0,1	-1	Falta de certeza limitada
1	0	Indefinición
10	1	Certeza limitada
100	2	Certeza moderada
1000	3	Certeza fuerte
10000	4	Certeza muy fuerte

Resultados presentados al Juez

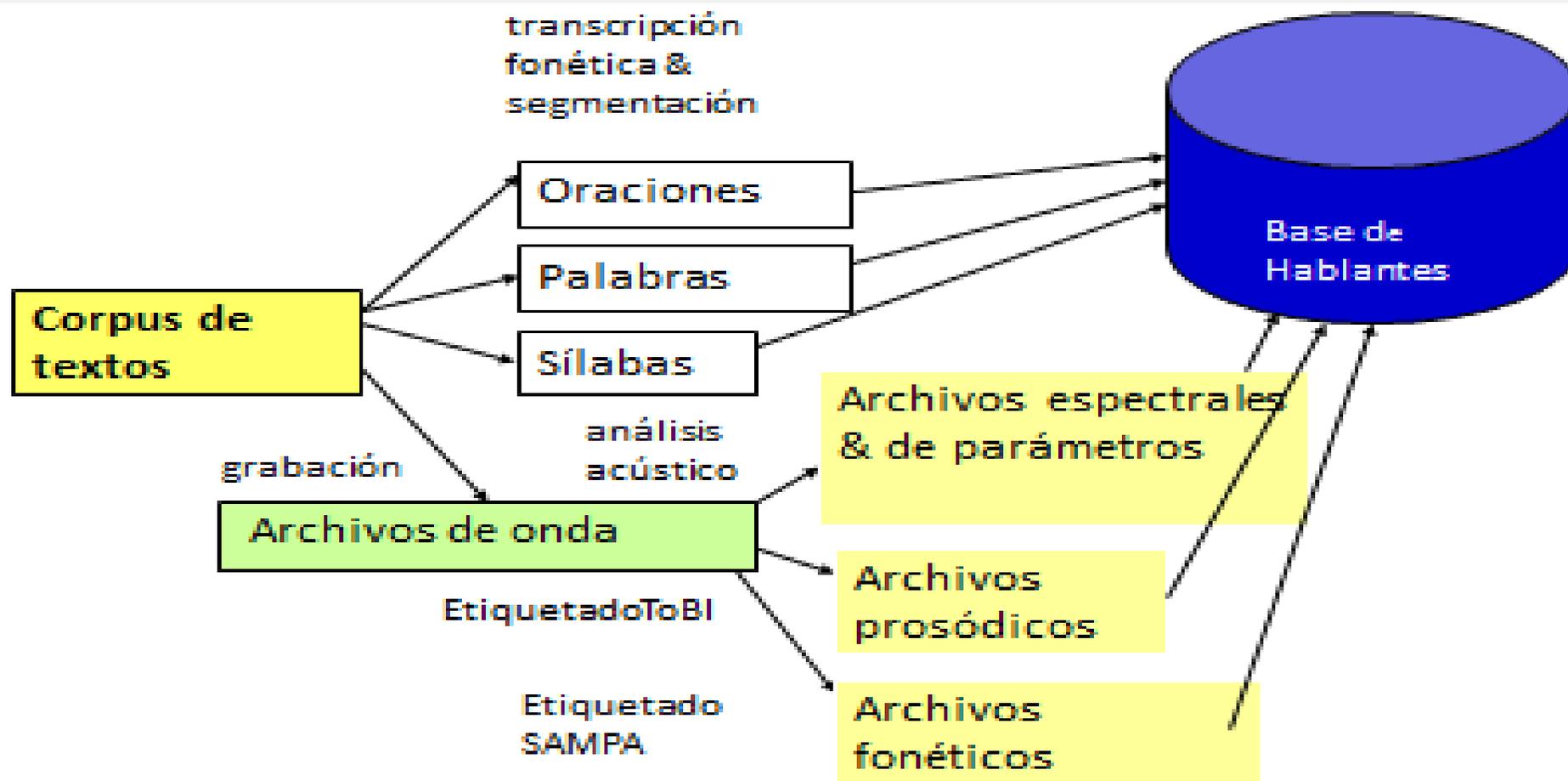
Metodo utilizado: I-vectors/PLDA

Cociente de verosimilitudes (LR) = XX

Es XX veces mas probable que la evidencia haya sido producida por el sospechoso que por cualquier otro hablante.

El juez o el jurado deberán usar esa declaración de la fuerza de la evidencia, agregar toda información a priori que no dependa de la evidencia para su deliberación y decisión final.

Diseño de la Base de Datos de Voces



Bases de datos

1. Base de datos de la población (*UBM*).
2. Base de datos del sospechoso, BD control.
3. Población de referencia o relevante.
4. Base de datos de calibración. Casos periciales
5. Base de datos de validación del sistema
6. Base de datos de evaluación del desempeño del sistema para un caso específico.

Base de datos UBM

- UMB para la H0, base con muestras de habla de múltiples **hablantes de la región e idioma de interés**. Este modelo de referencia universal (UBM) sirve para estimar cuán típico es observar el habla a cotejar, en la población en cuestión.
- Además puede entrenarse usando datos que representen al hablante impostor con la BD relevante.
- Para tener el mismo tipo de habla que el empleado en el enrolamiento (modalidad, idioma, canal)

Empleo de UBM obtenido de otras Lenguas

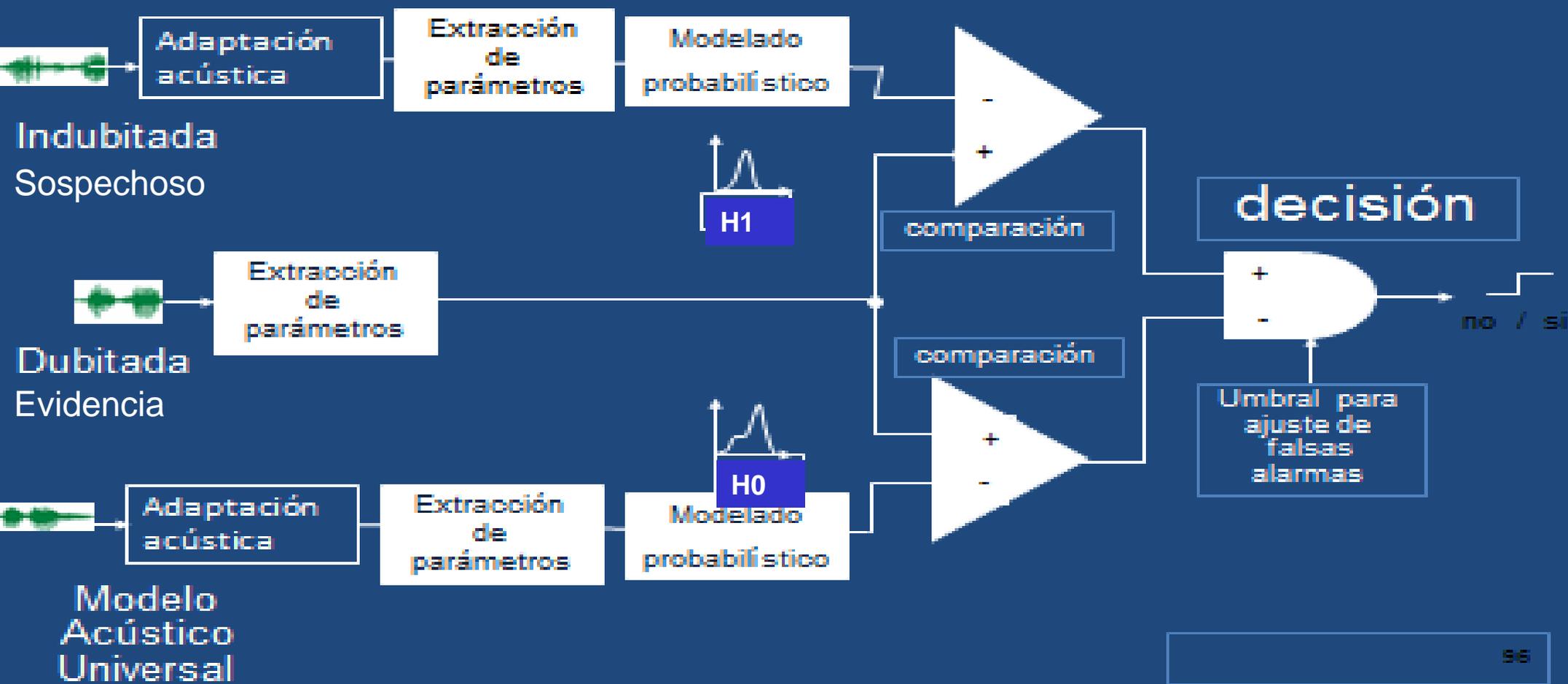
- La diferencia de lenguajes genera degradación en el desempeño.
- Estudios previos:
 - *Entrenamiento de un modelo a partir de un corpus multilingüe* (Ma 2004)
 - *Normalización de lenguaje* (Akbaşcak 2007)
 - *Compensación del Factor Lenguaje* (Lu 2009)
 - *Combinación de atributos* (Nagaraja 2013)

Base de datos local: Posee información del sospechoso y de hablantes similares. Contiene ejemplos de canales y ambientes variados. BD relevante.

Base de datos regional: da información de las variaciones dialectales por región (alófonos mas usados)/país (lengua) o por zona geográfica (geolectos). da información sociolectal: grupo étnico (etnolecto), o de clases: económico, educativo, profesional (tecnolecto). Forma de hablar de hombres y mujeres (sexolecto).

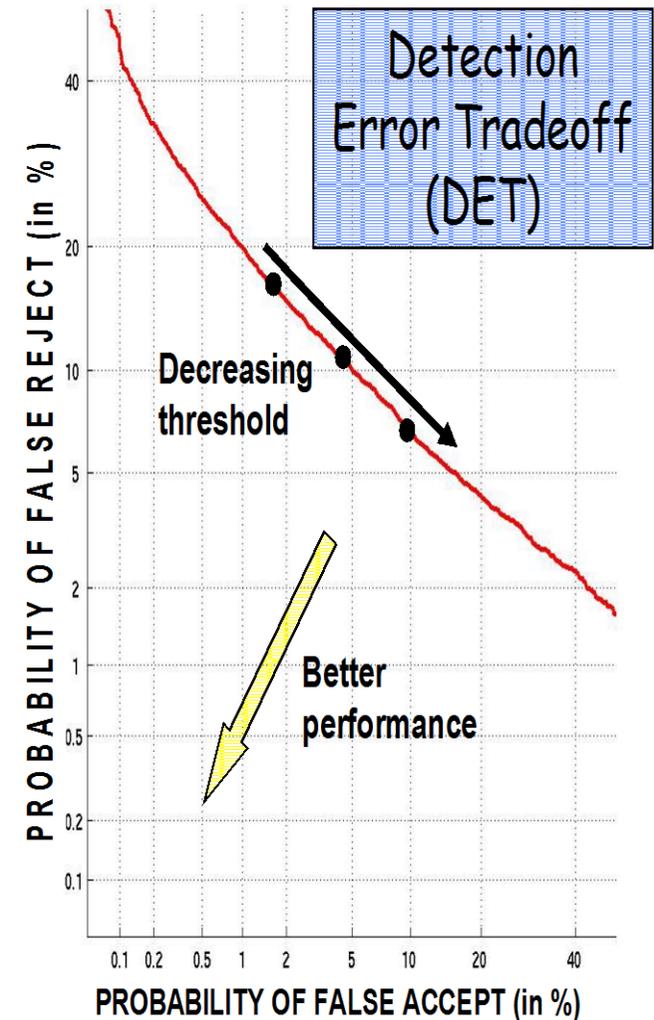
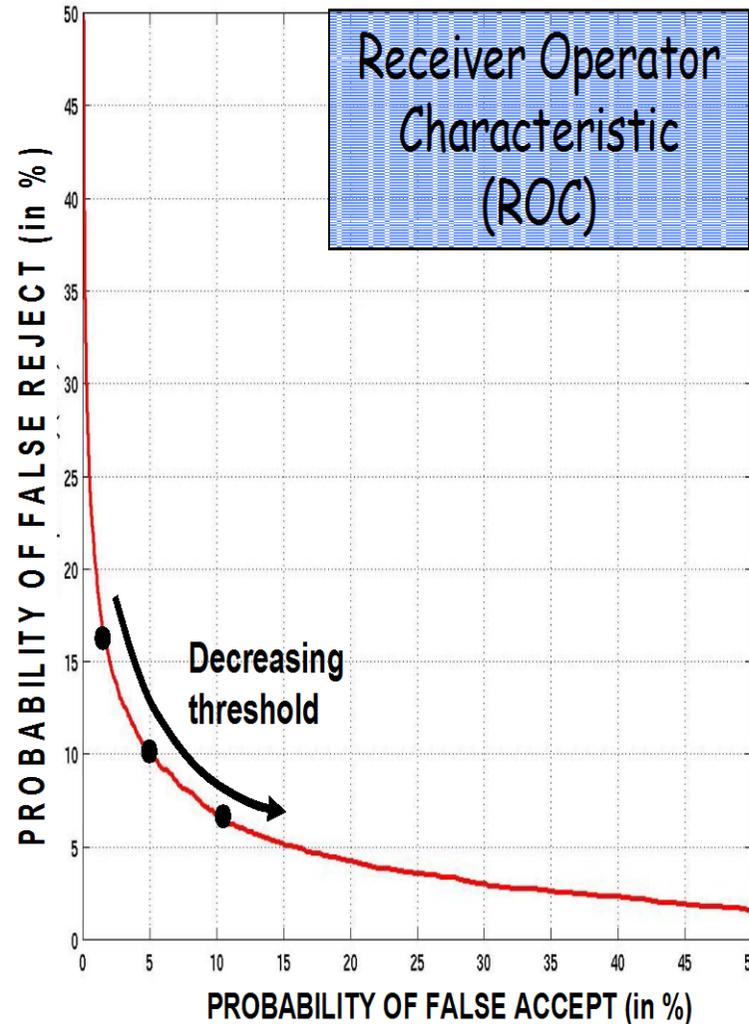
Base de datos internacional: da información sobre una gran cantidad de sonidos (multilingue) de habla producidos por cualquier ser humano.

Sistema de reconocimiento del hablante



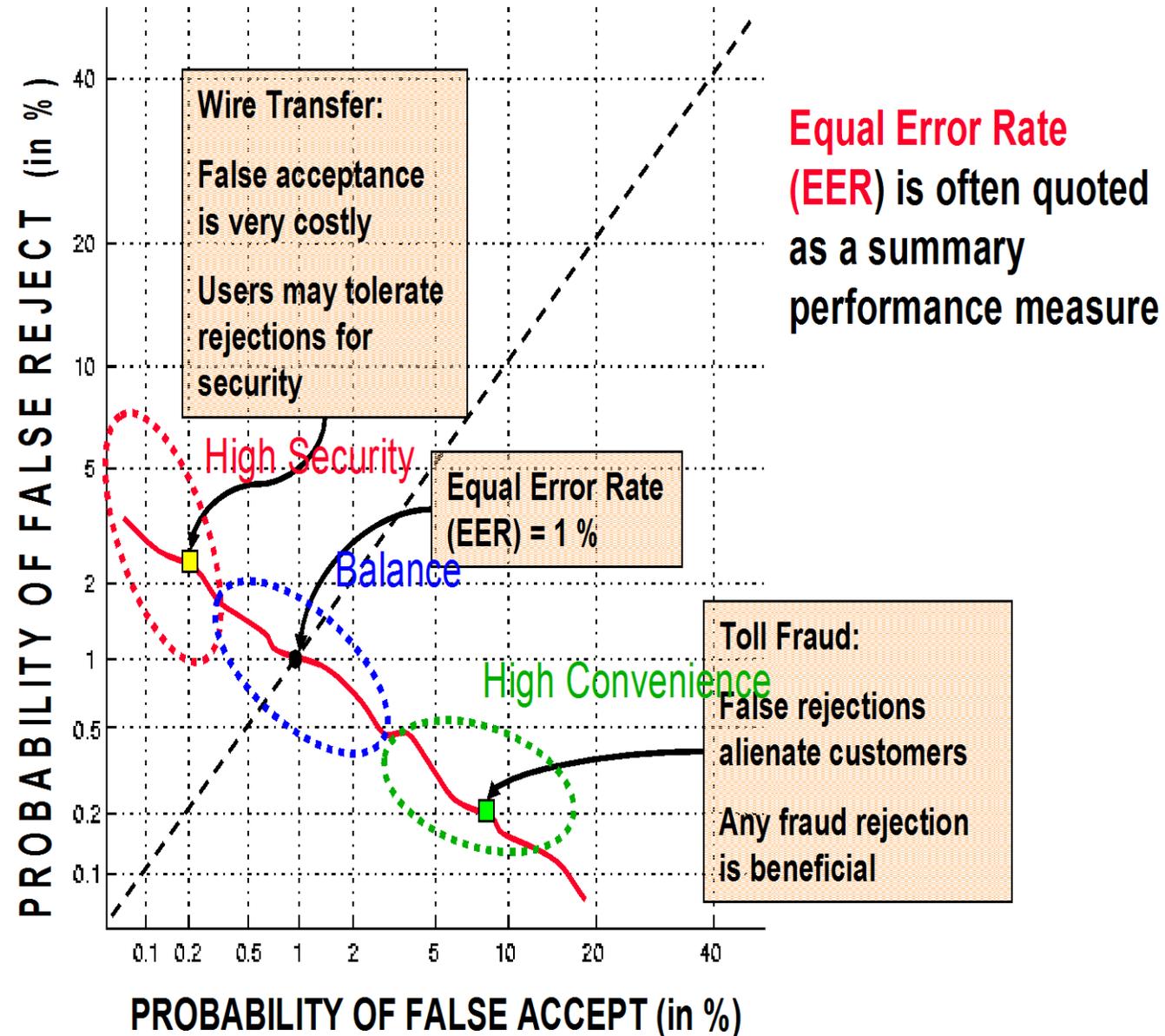
Curvas ROC y DET

Caracterizan el desempeño del Sistema en función de:
P(falso rechazo)
vs.
P(falsa aceptación)



Curva DET

La aplicación del punto de operación depende de los costos relativos de dos tipos de errores: Falsas Aceptaciones y Falsos Rechazos



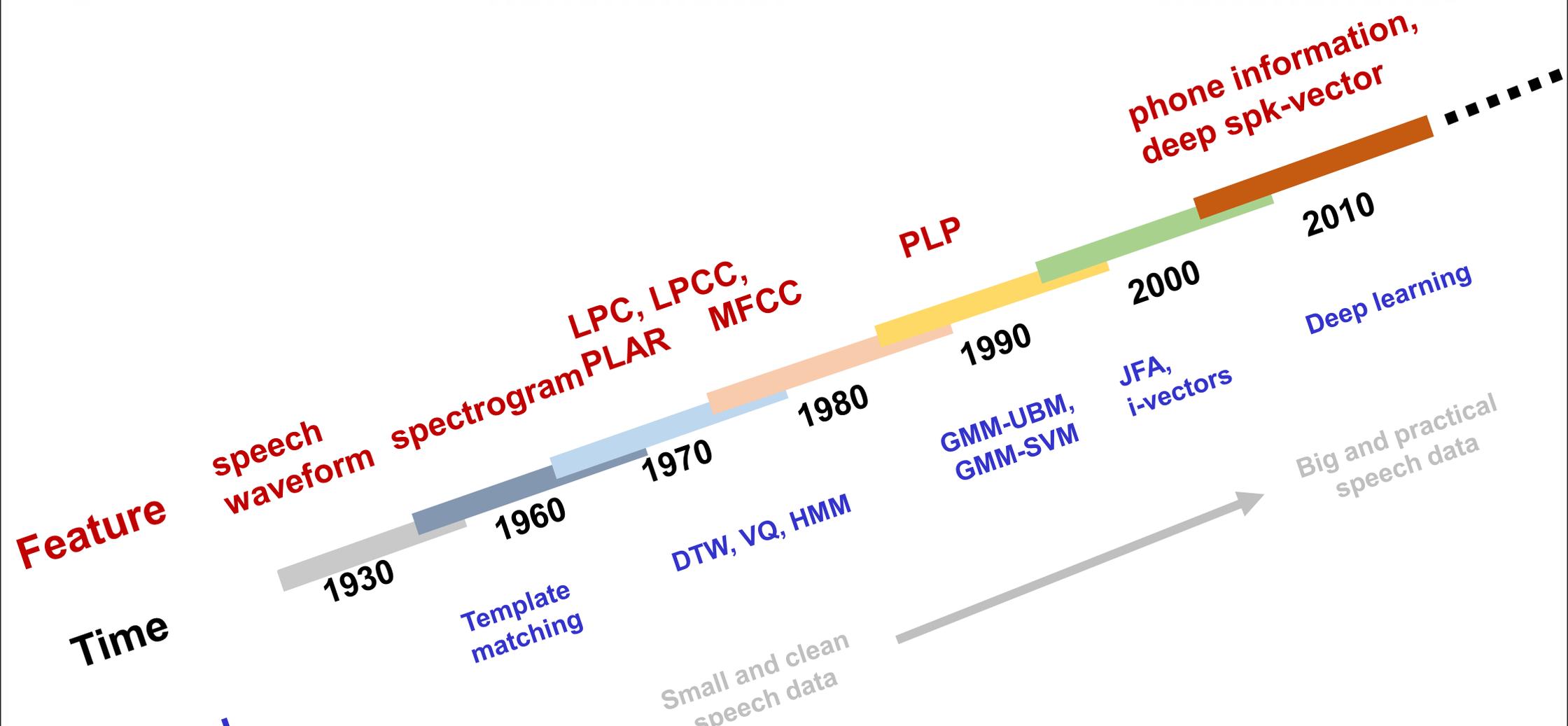
Evaluación de reconocedores de hablantes con una base de datos de Validación.

- Las diferentes aplicaciones de reconocedores de hablantes tienen costos diferentes para las pérdidas y las falsas alarmas
- ¿Cómo evaluar el desempeño independientemente de la aplicación?

Brümmer & du Preez 2006

$$C_{lr} = \frac{1}{2} \left(\frac{1}{N_{ss}} \sum_{i=1}^{N_{ss}} \log_2 \left(1 + \frac{1}{LR_{ss_i}} \right) + \frac{1}{N_{ds}} \sum_{j=1}^{N_{ds}} \log_2 \left(1 + LR_{ds_j} \right) \right)$$

Evolución de la Tecnología



Feature
Time

1930 speech waveform
1960 spectrogram
1970 LPC, LPCC, PLAR, MFCC
1980 DTW, VQ, HMM
1990 GMM-UBM, GMM-SVM
2000 JFA, i-vectors
2010 phone information, deep spk-vector
Deep learning
Big and practical speech data

Sistema de Identificación Forense de Hablantes Sistema Argentino i-vectors y PLDA

Este sistema formó parte de la evaluación *Speakers in the Wild* (SITW) desarrollada por SRI International durante el 2016 y fue verificado con datos de la evaluación NIST HASR 2012 .

Evidencia

N°1



Es posible agregar archivos de audio propios para probar el sistema de identificación:

Browse... N°1.wav

Enviar

100%

Uploaded File: N°1.wav

Plana de Voz

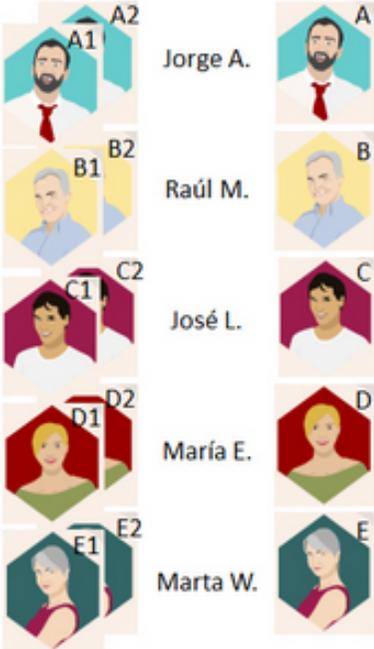
A +0.542

B +0.195

C +0.317

D +0.363

E +0.414



Copyright © Univaso, Martínez Soler, Gurlekian.

<http://181.28.244.40/>

Contacto

Laboratorio de Investigaciones Sensoriales

www.lis.secyt.gov.ar

(011) 5950-9024

jgurlekian@hospitaldeclinicas.uba.ar

Programa Ciencia y Justicia

cienciayjusticia@conicet.gov.ar

(011) 4899-5400 int. 2819/2817

Modelo de la Producción de Habla Gunnar Fant (1960)

Fuente
de Excitación

Onda Glotal
Ruido

Función de
transferencia
del Tracto Vocal

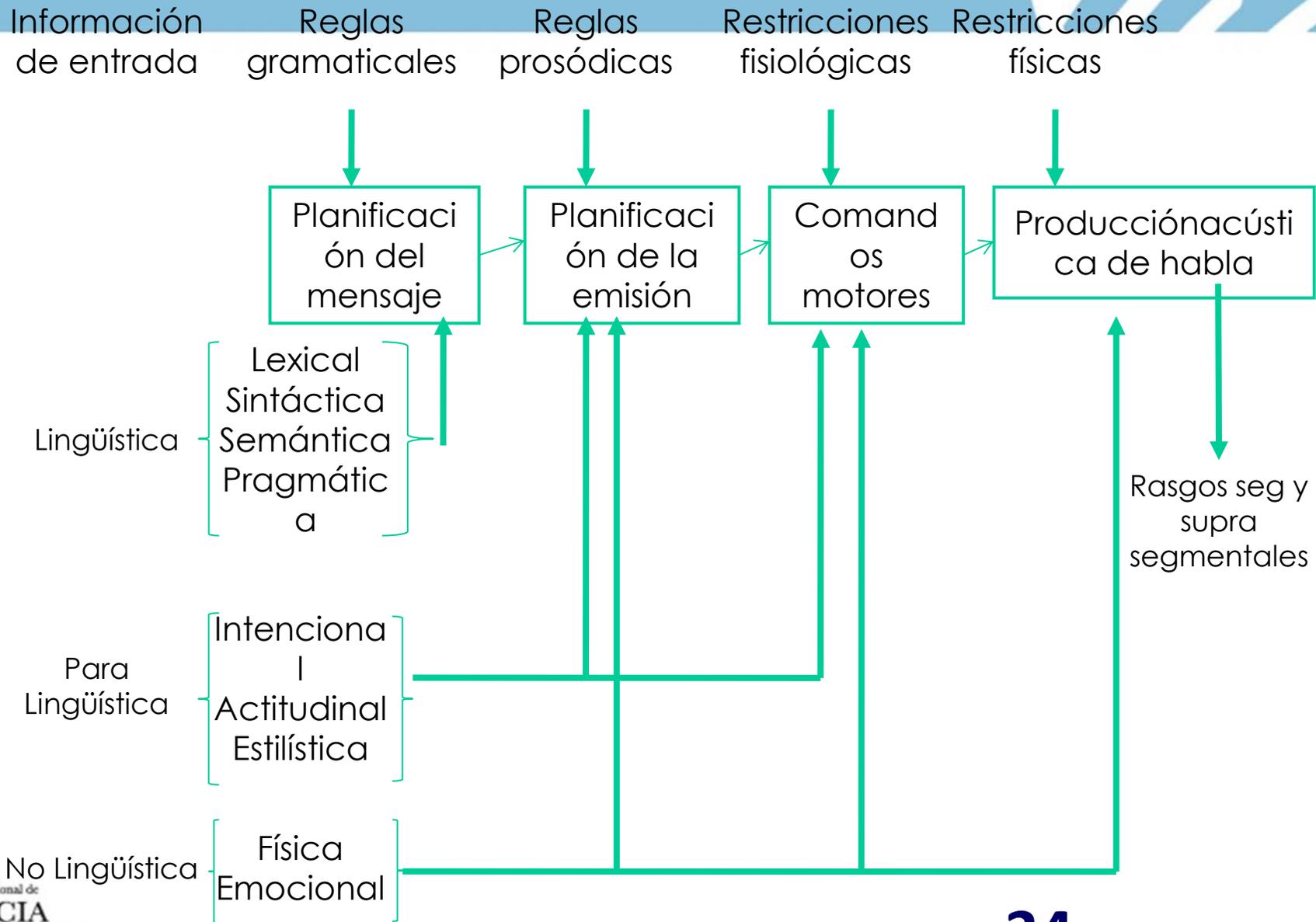
Lengua,
Paladar,
Labios, etc.

Radiación

Desde Labios
al Oído
Externo

Producción de Habla

Procesos de Información (Fujisaki, 2010)



Análisis Perceptual

