Speech Analytics

RECONOCIMIENTO DEL GENERO A PARTIR DE LA VOZ

JSOSABRI@EVERIS.COM

Estado del arte

AVANCE DEL ANÁLISIS DE LA VOZ EN NUESTROS DÍAS

¿Qué ofrece la tecnología?

- Síntesis de Voz – Text to Speech
- Reconocimiento de Voz – Speech to text
- Análisis del Lenguaje y la Minería de Texto
- Traducción en tiempo real
- Procesamiento de Lenguaje Natural (NLP)
- Motores de búsqueda inteligentes

¿Qué es el Speech Analytics?

CONCEPTOS GENERALES

Definición de Speech Analytics

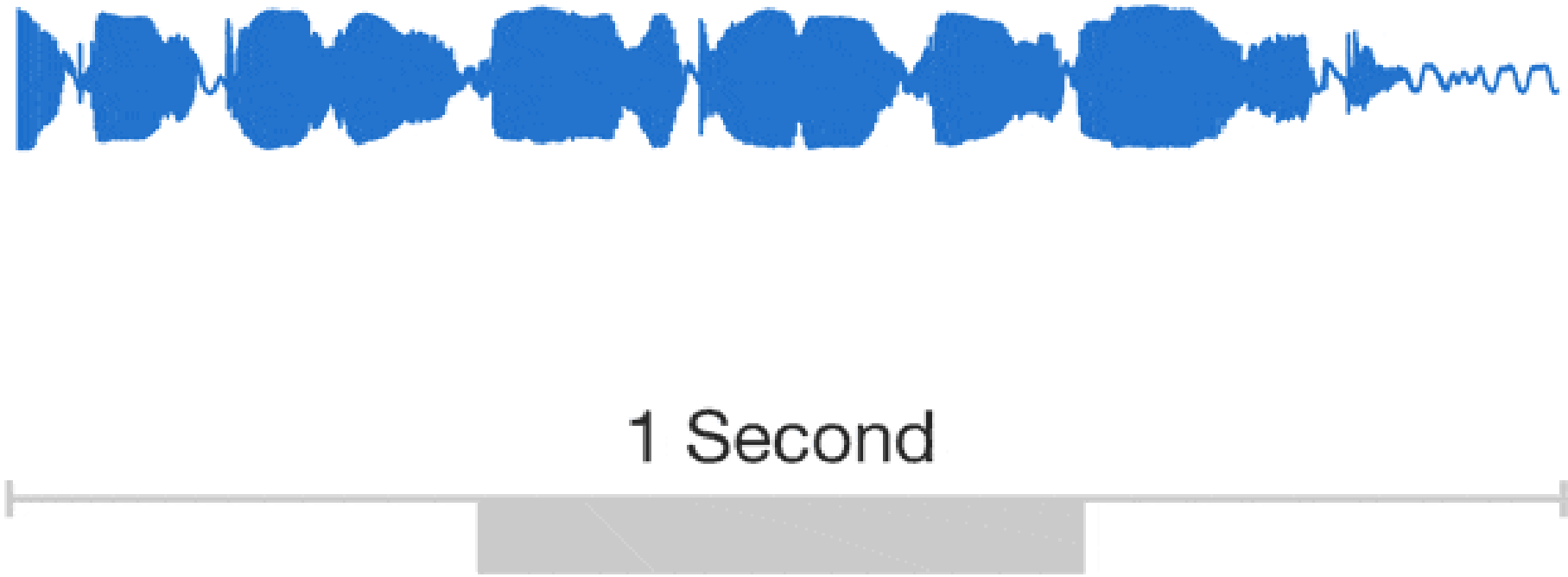
El análisis del habla o speech analytics trata de sistemas que permiten extraer información a partir del análisis de las conversaciones grabadas. Esto permite, por ejemplo recopilar información del cliente para mejorar la comunicación y la interacción futura. El proceso es utilizado principalmente por los centros de contacto de los clientes para extraer información oculta en las interacciones del cliente con una empresa.

Para ello se basa en diferentes herramientas:

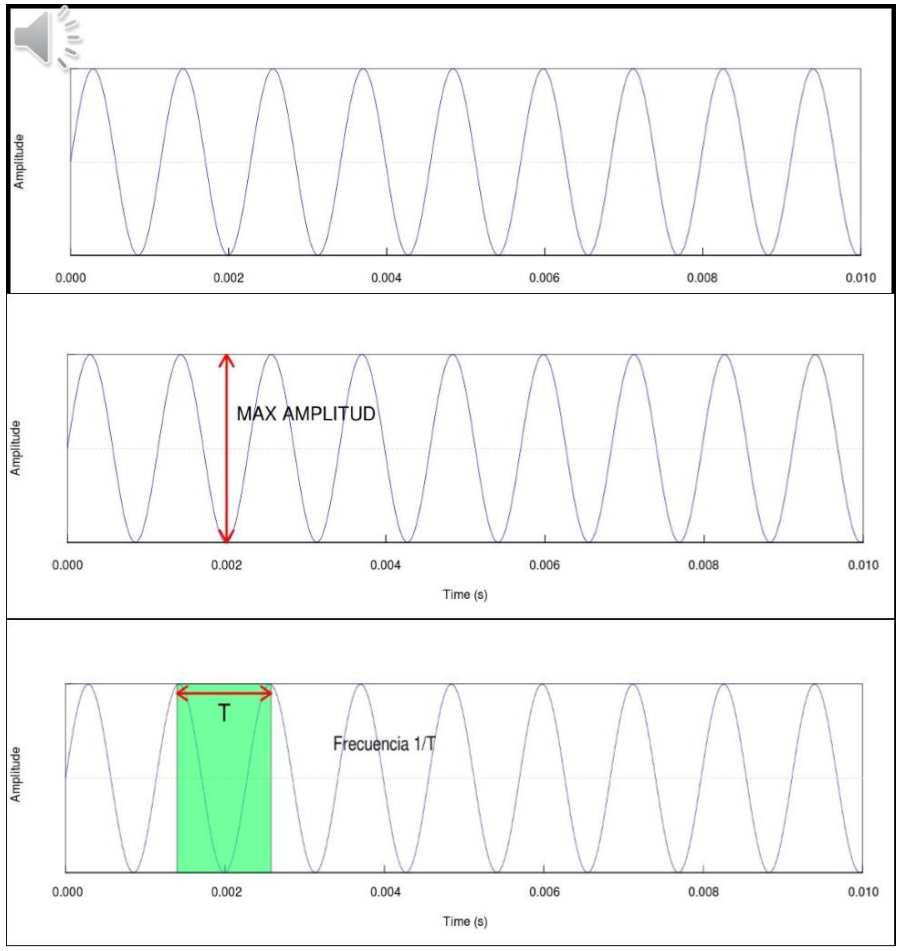
- Síntesis de voz.
- Reconocimiento del habla.
- Reconocimiento de patrones sonoros.

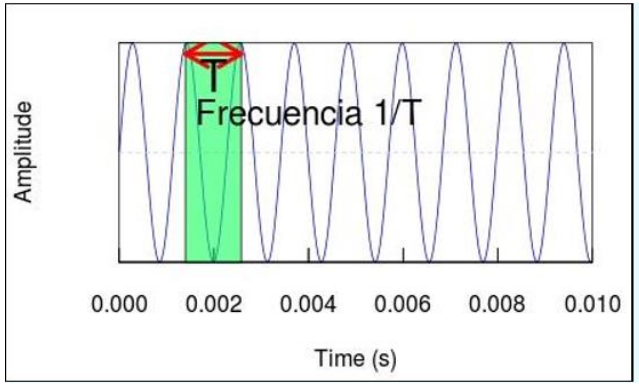
Fuente: Wikipedia

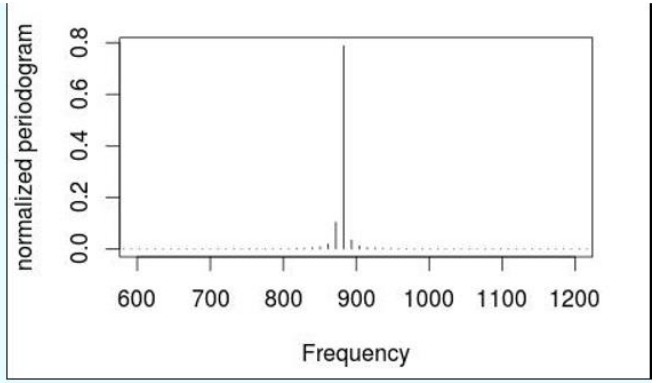
La Voz



El Sonido – Tiempo y Trecuencia







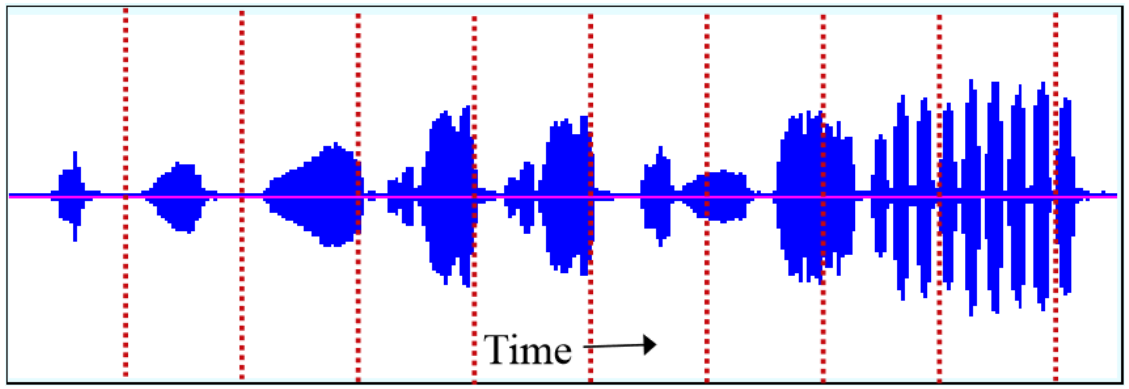
Análisis de Fourier

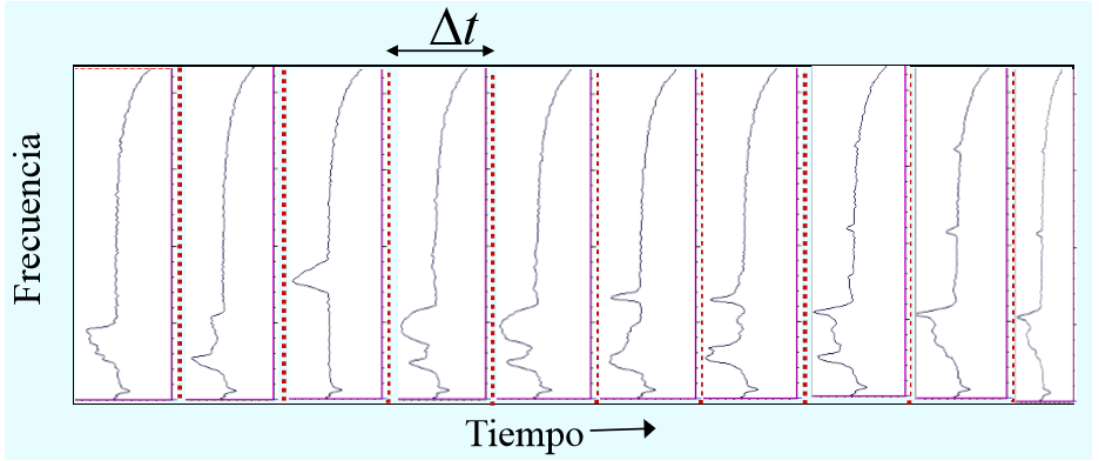
Por suerte!

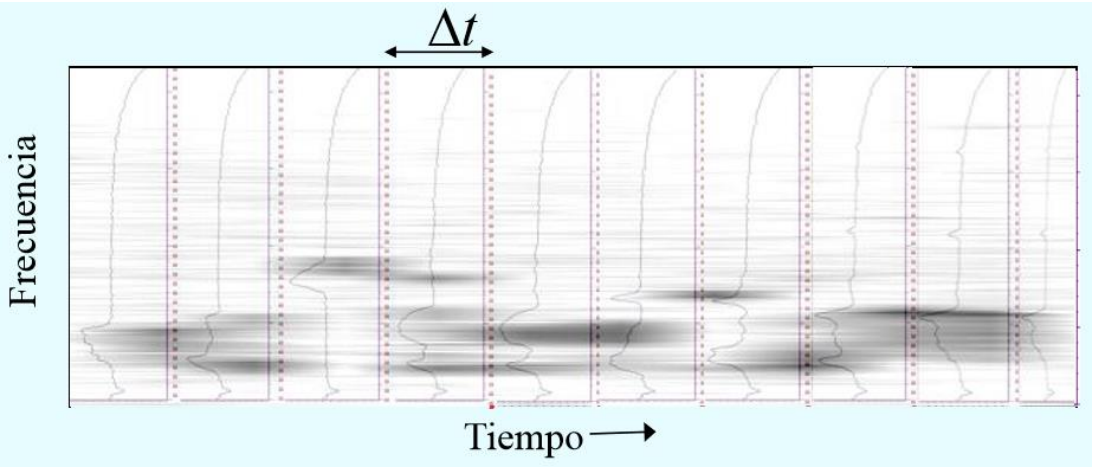
- Cualquier onda continua puede ser descompuesta en ondas sinusoidales puras con frecuencia y amplitud (Análisis de Fourier)
- Los gráficos en el dominio de frecuencia son MUY útiles para comparar sonidos!

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-j2\pi kn/N}, \quad 1 \leq k \leq K$$

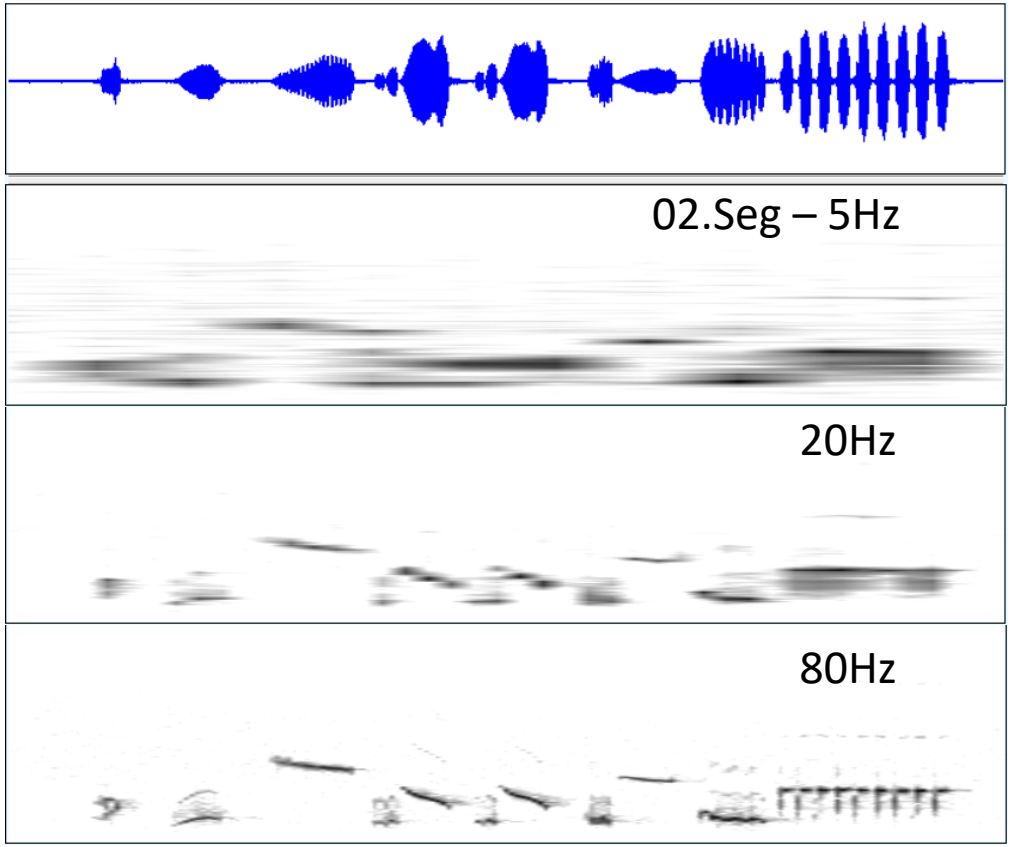
Análisis de Fourier

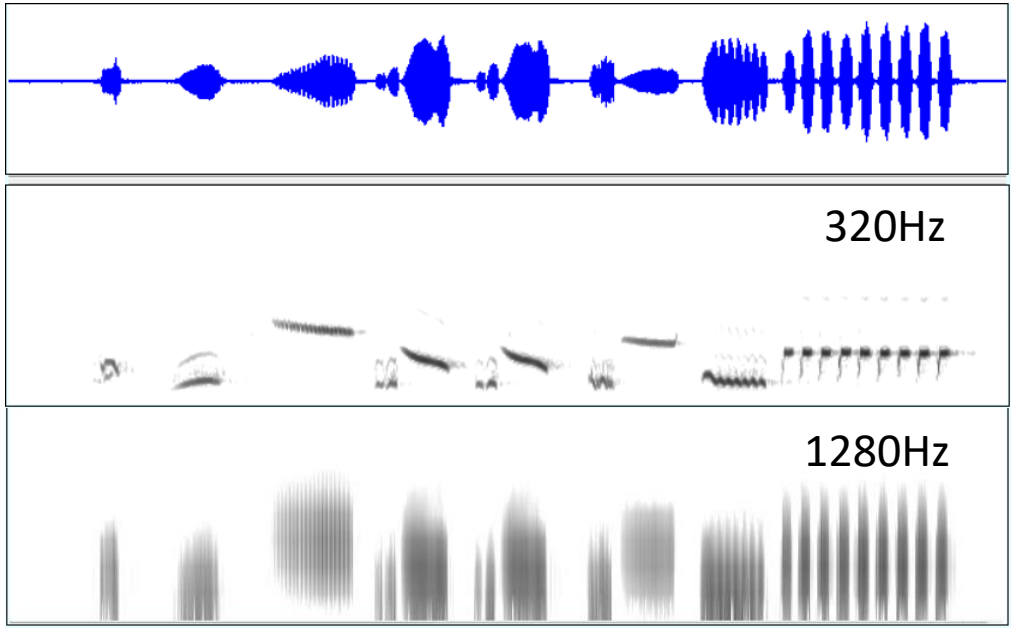






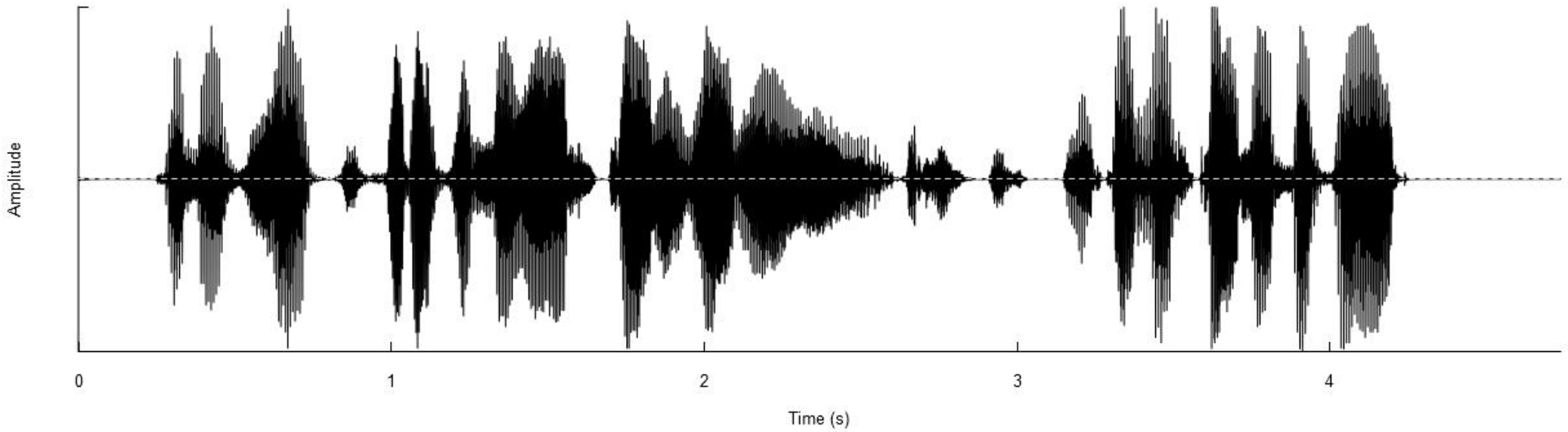
Análisis de Fourier - Espectrograma



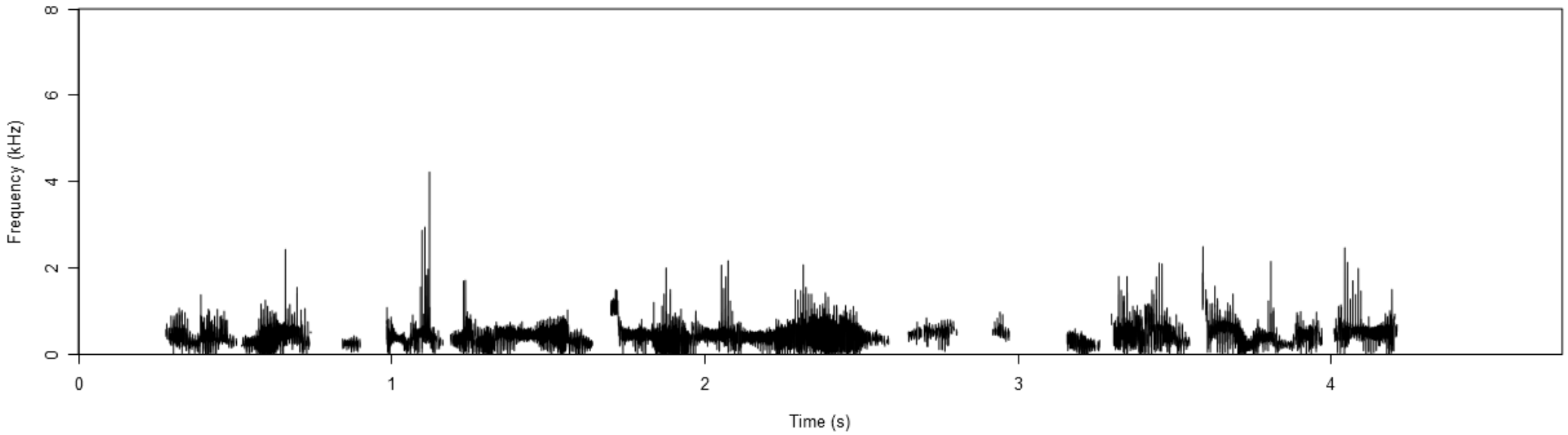


La banda de frecuencia de voz utilizable oscila entre aproximadamente 300 Hz y 3400 Hz. Según el teorema de muestreo Nyquist-Shannon , la frecuencia de muestreo (8 kHz) debe ser al menos el doble del componente más alto de la frecuencia de voz a través del filtrado apropiado antes del muestreo en tiempos discretos (4 kHz) para la reconstrucción efectiva de la señal de voz.

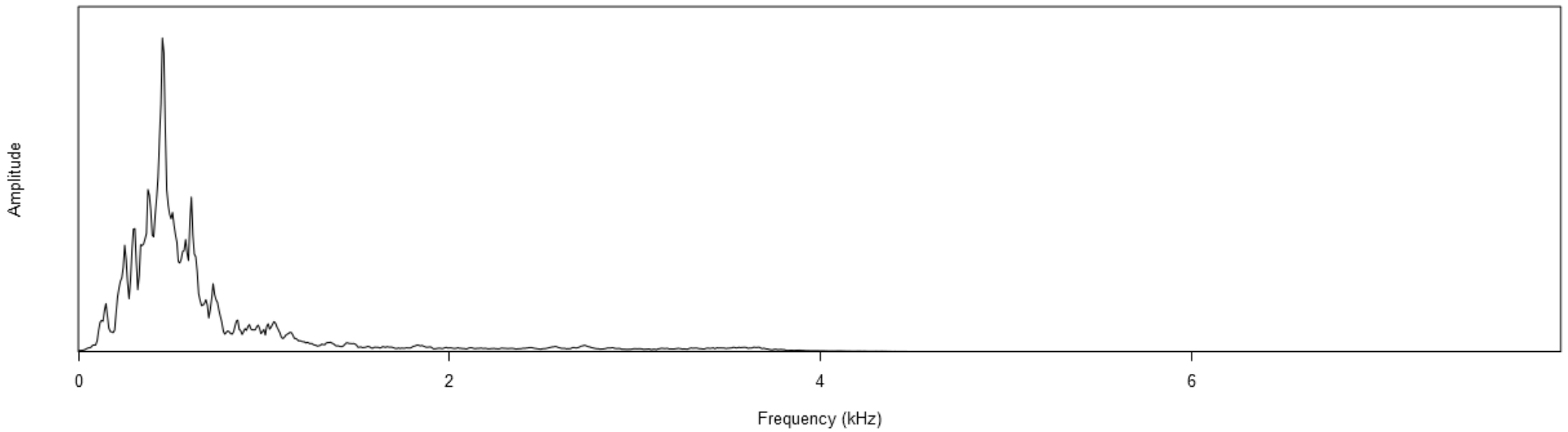
La Voz – un ejemplo en tiempo



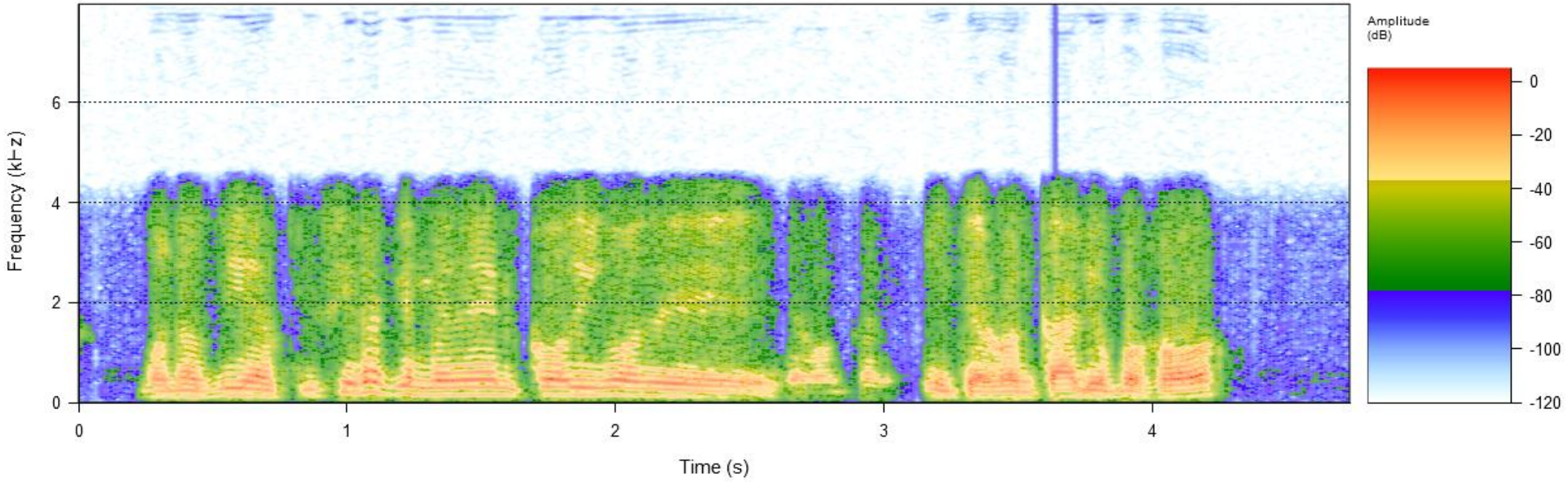
La Voz – un ejemplo en frecuencia



La Voz – Distribución de frecuencias



La Voz – Espectrograma



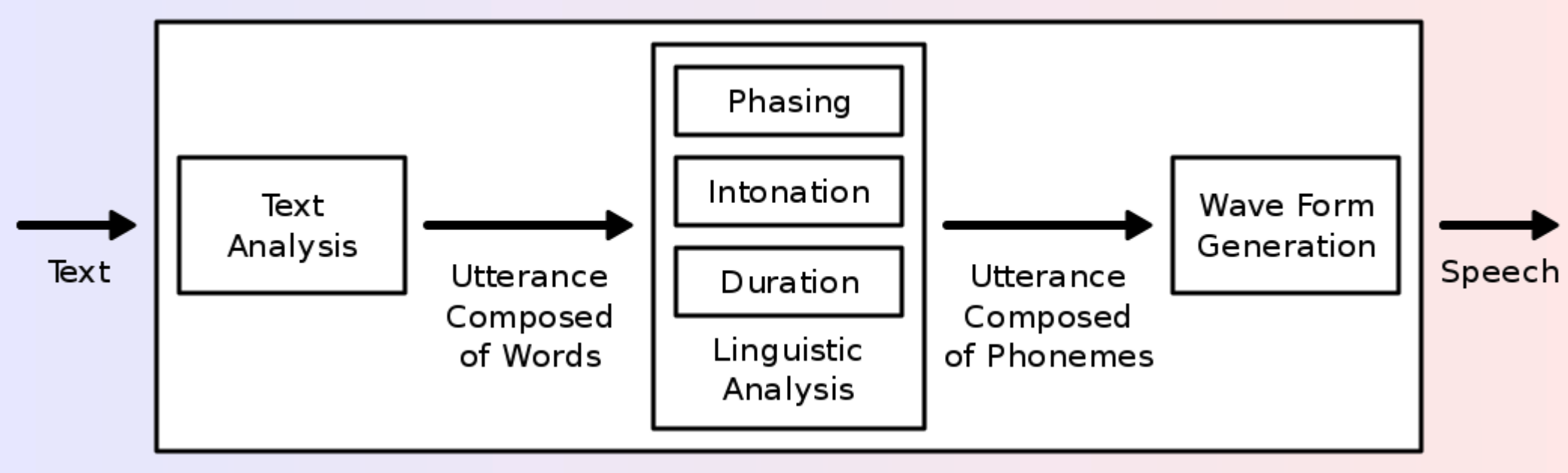
Síntesis de la Voz

PRODUCCIÓN ARTIFICIAL DEL HABLA

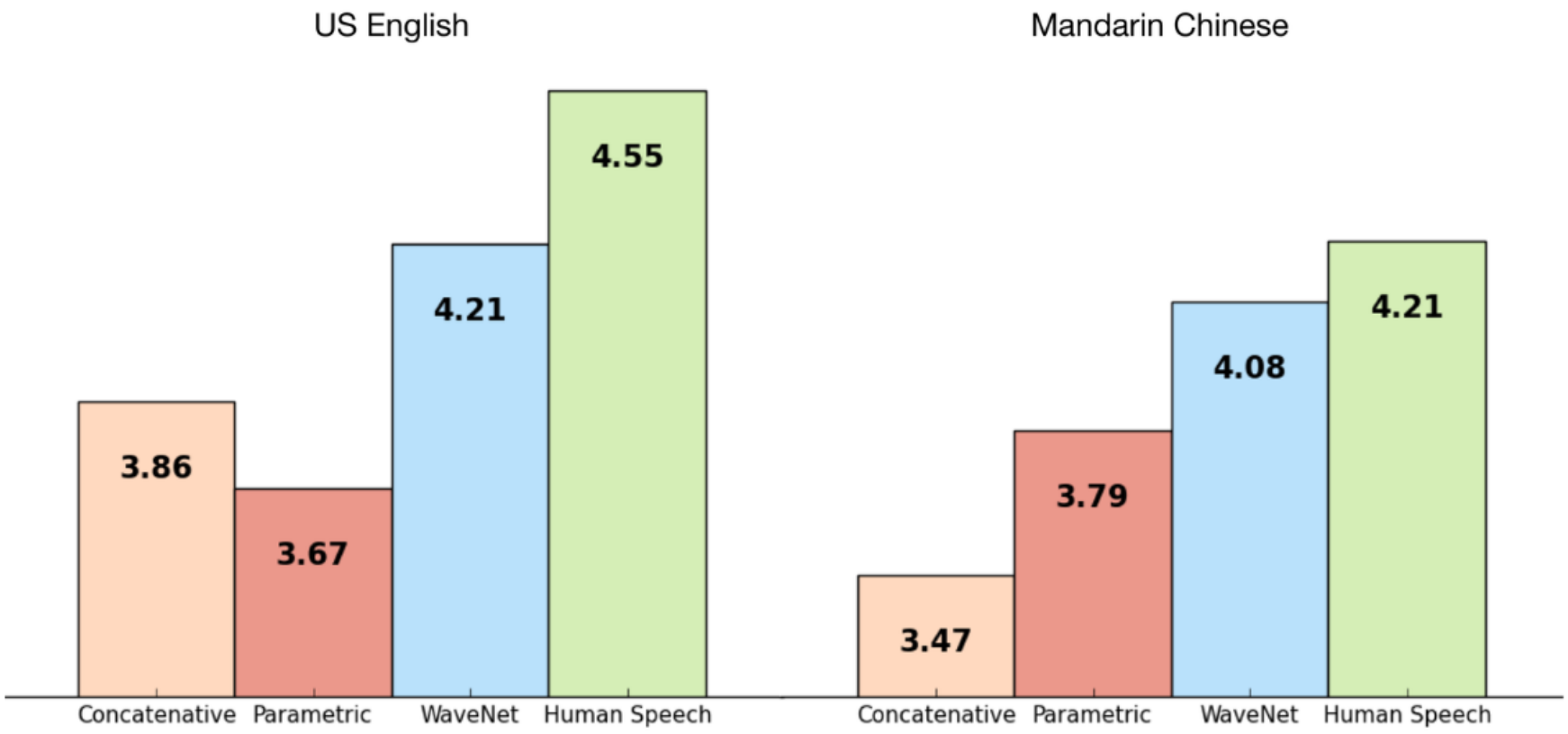
Síntesis de voz (Text to Speech)

- Sintetic voice (ejemplo)
- Parametric TTS 
- Concatenative TTS 
- Deep Learnig (RNN, WaveNet) 

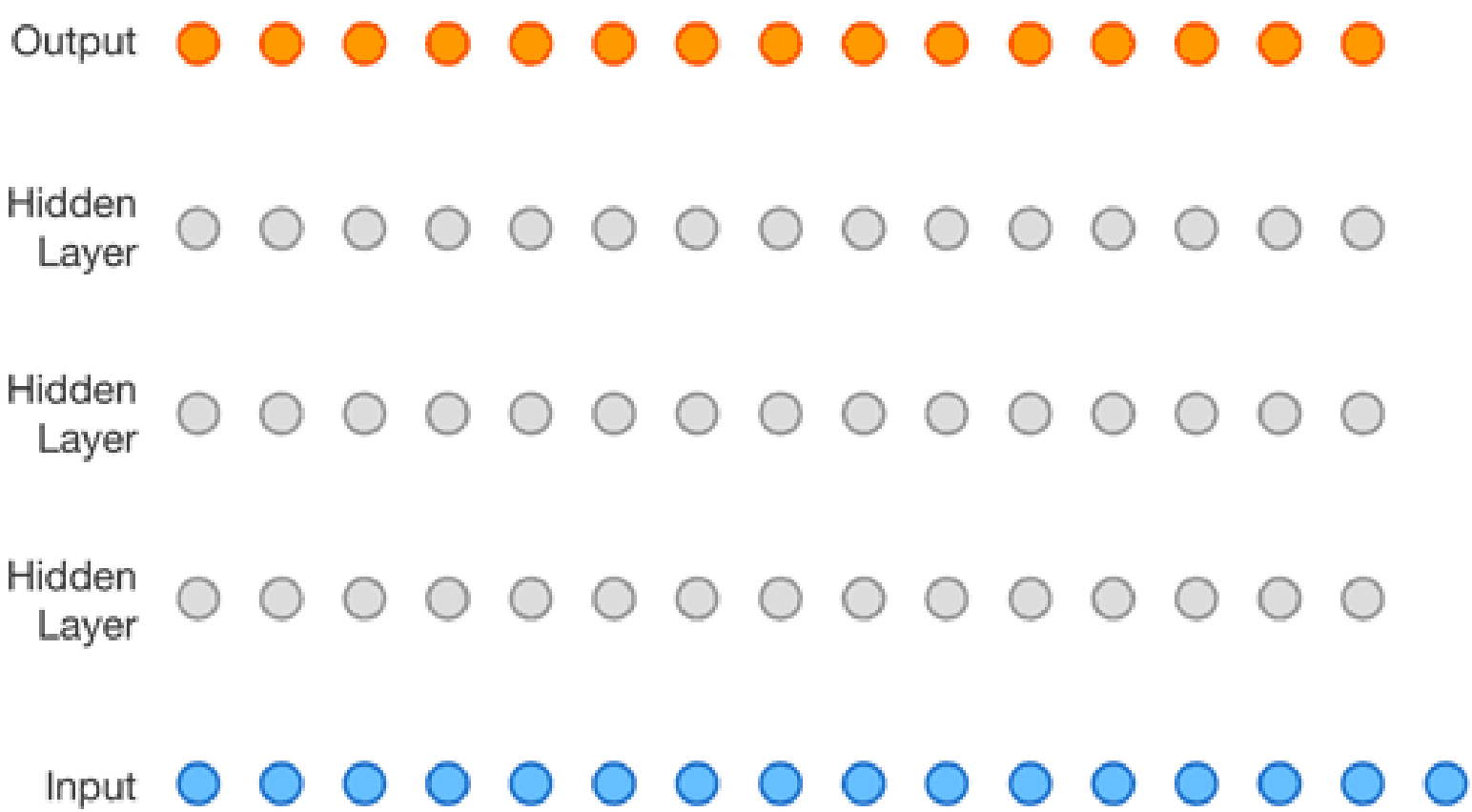
Síntesis de voz - Parametric TTS



Síntesis de voz (Text to Speech)



Síntesis de voz - WaveNet

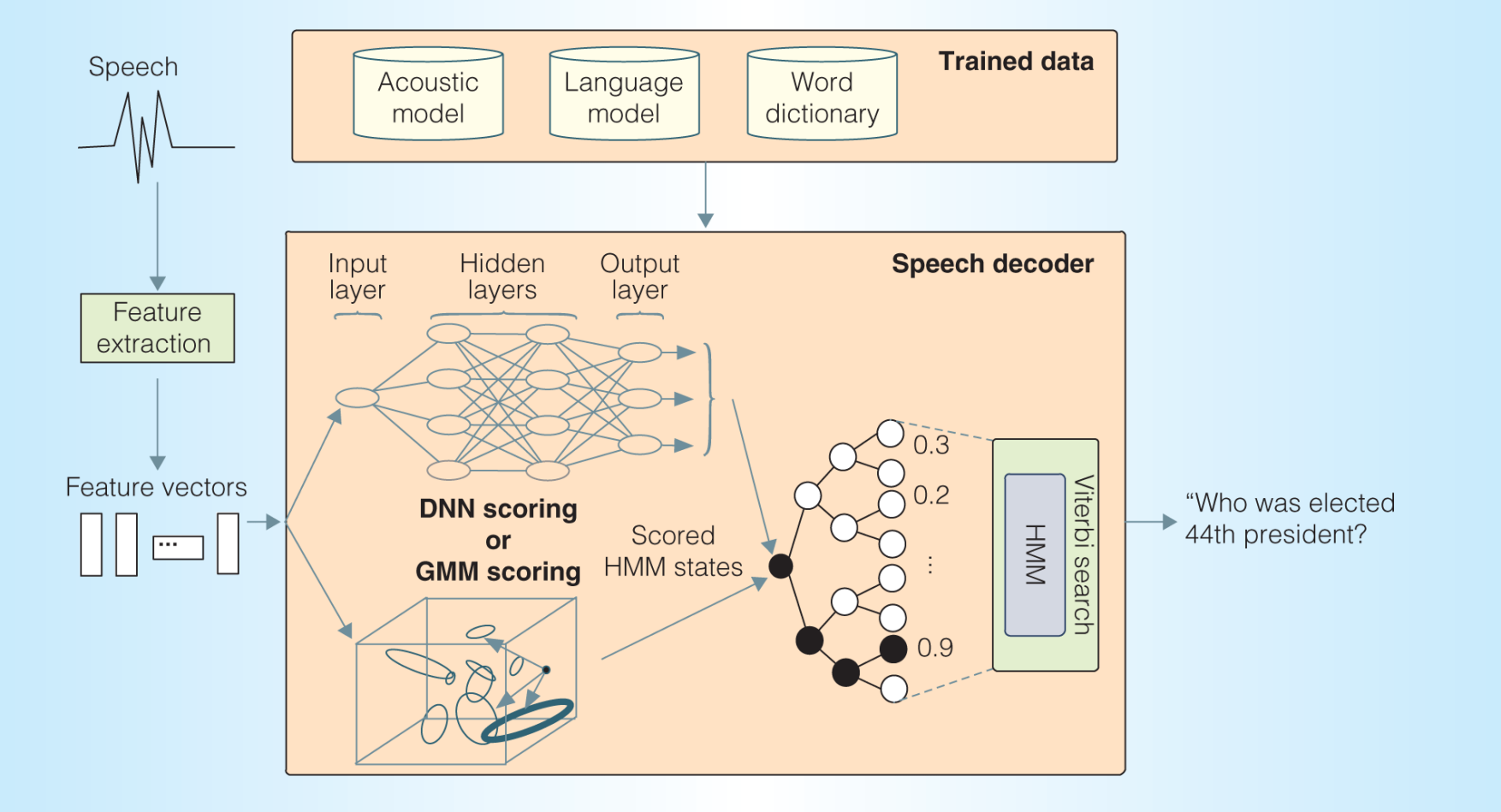


everis
an NTT DATA Company

Reconocimiento del habla

DETECCIÓN DEL CONTENIDO DE UNA CONVERSACIÓN

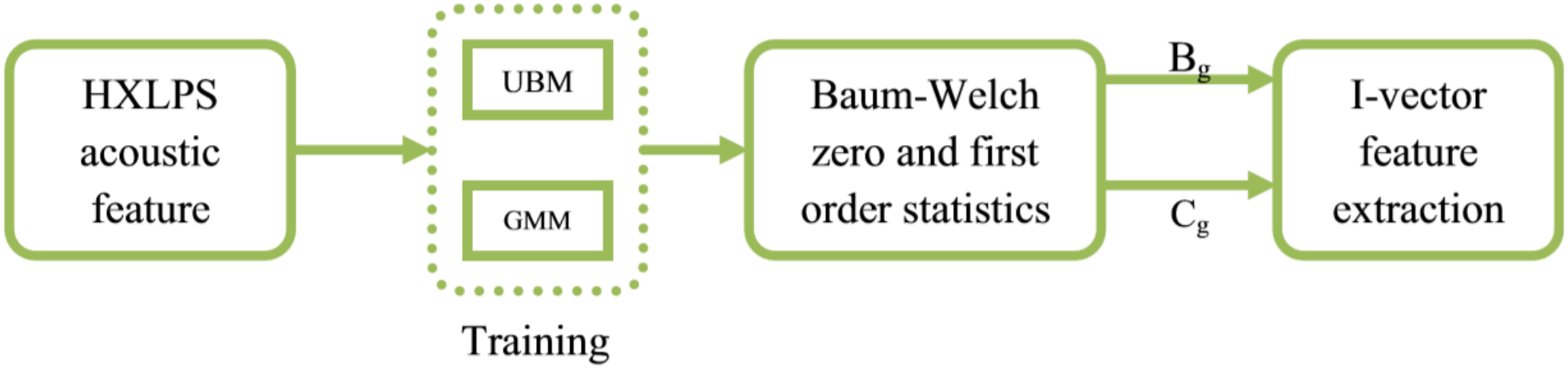
Reconocimiento del habla



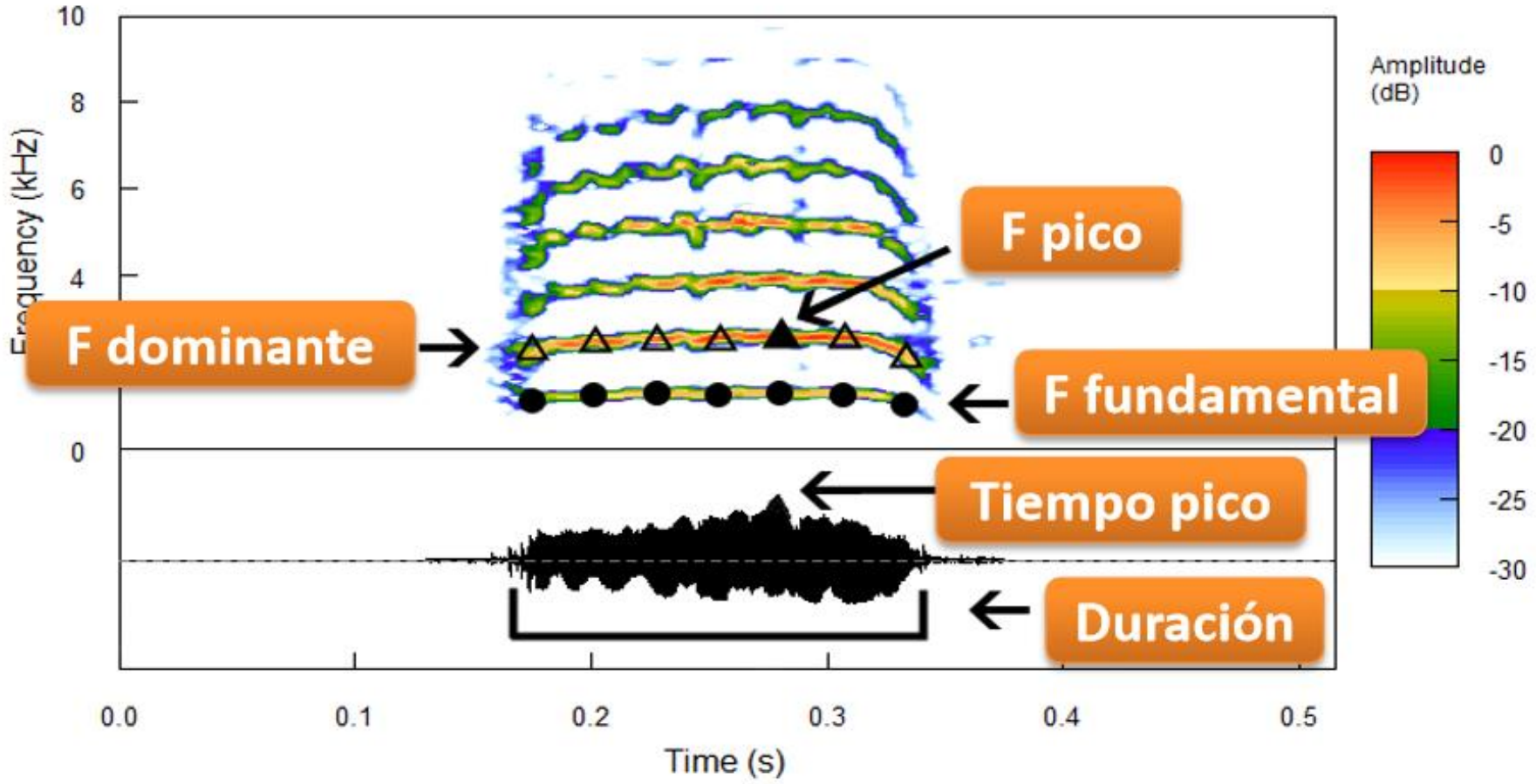
Reconocimiento de Patrones

RECONOCIMIENTO DE PATRONES SONOROS DE UNA CONVERSACIÓN Y SU ENTORNO

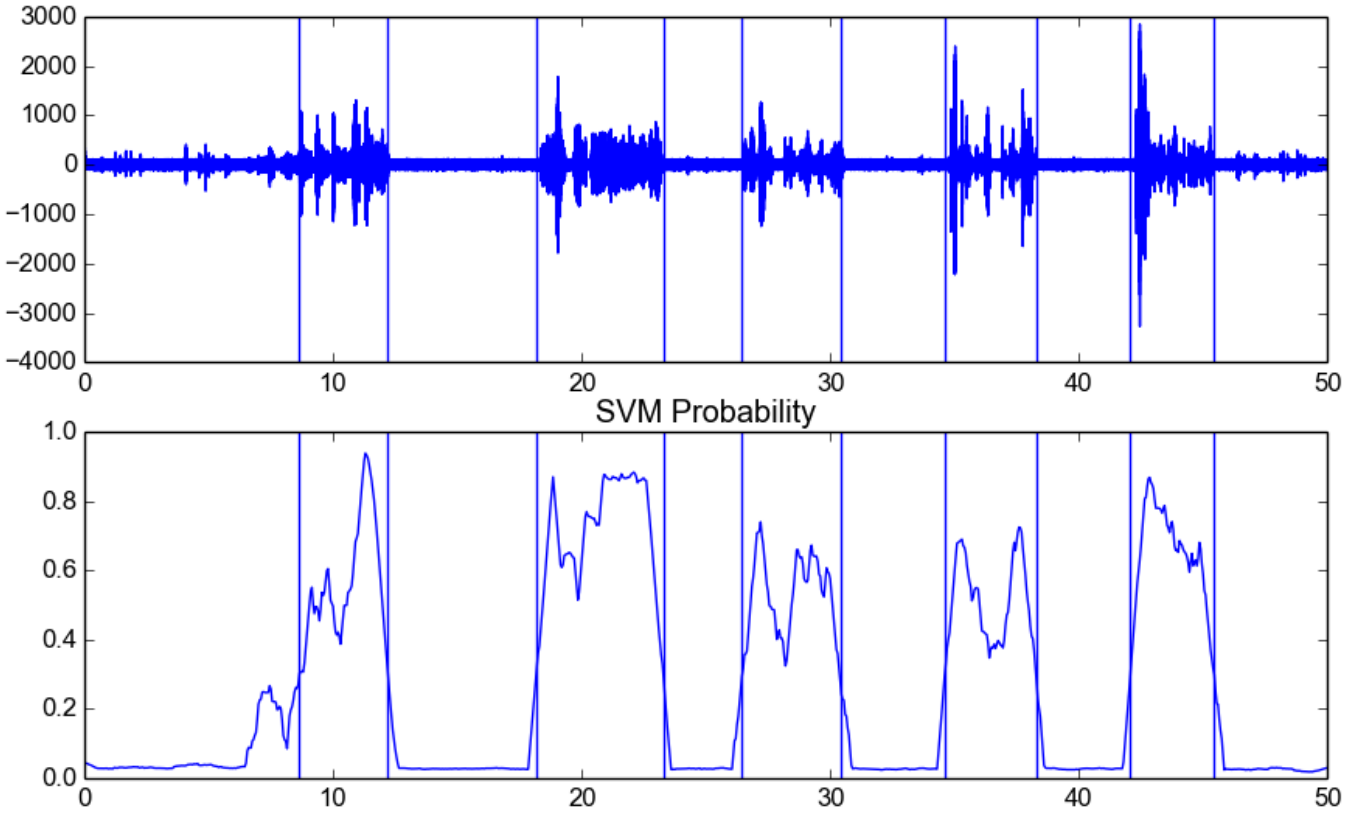
Extracción de características



La Voz – Espectrograma



Detección de silencios - pyAudioAnalysis



Speech Diarization

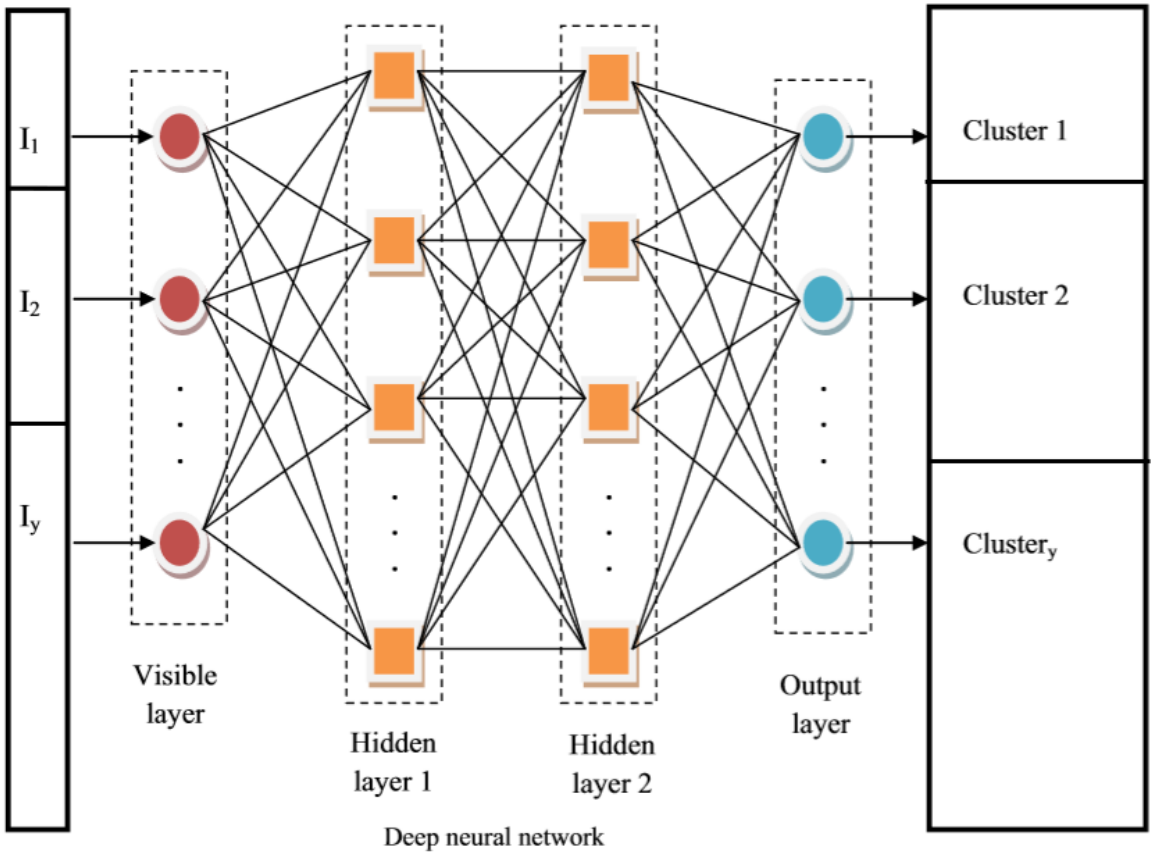
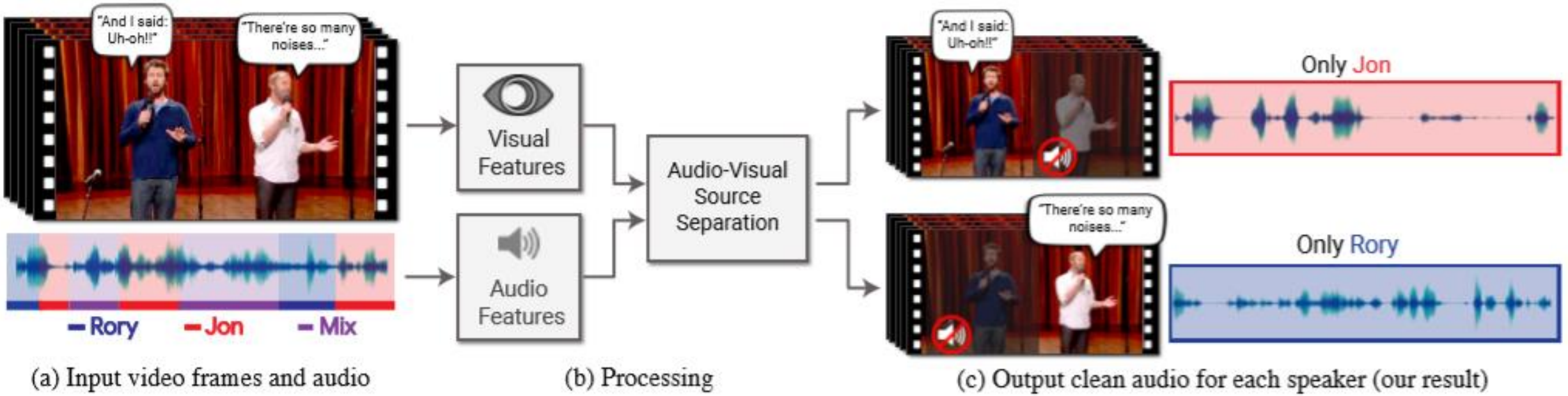


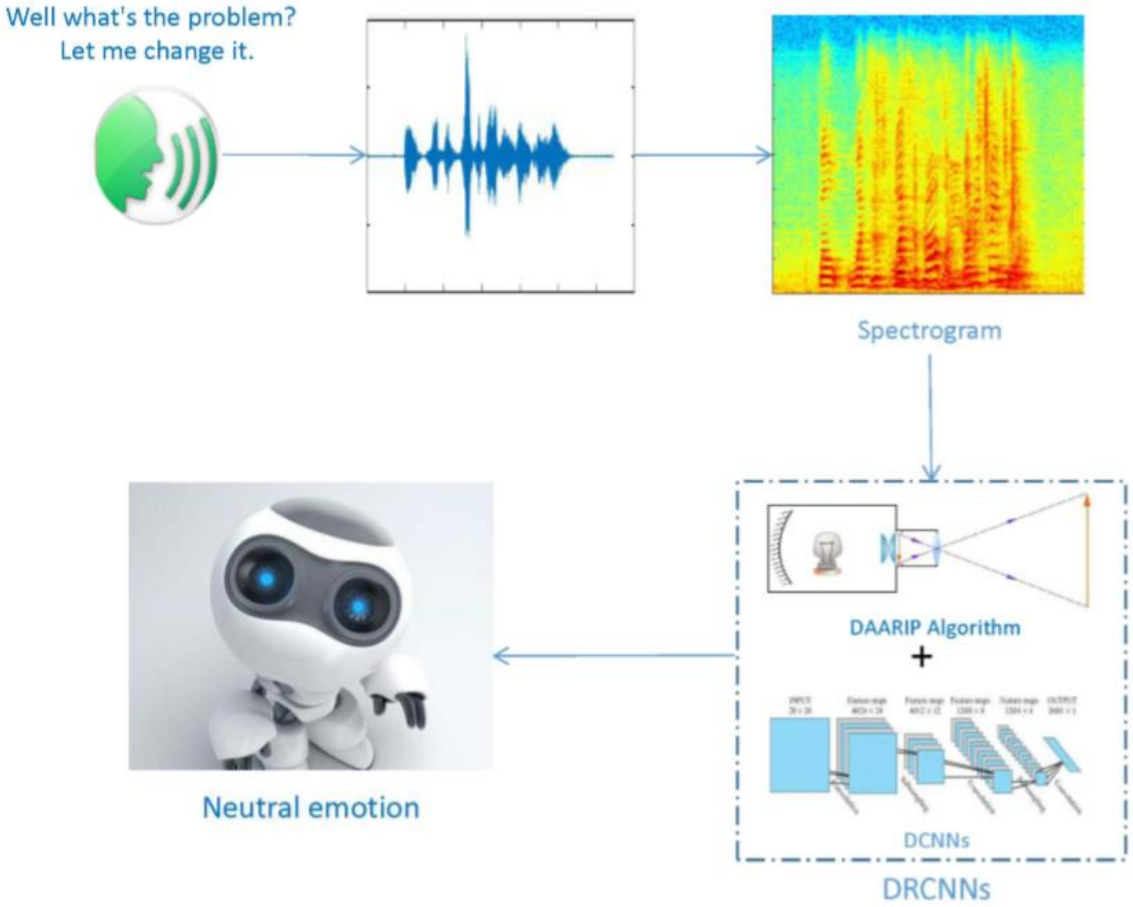
Fig. 3 Processing model for speaker clustering using DNN.

Cocktail_Party



(a) Input video frames and audio (b) Processing (c) Output clean audio for each speaker (our result)

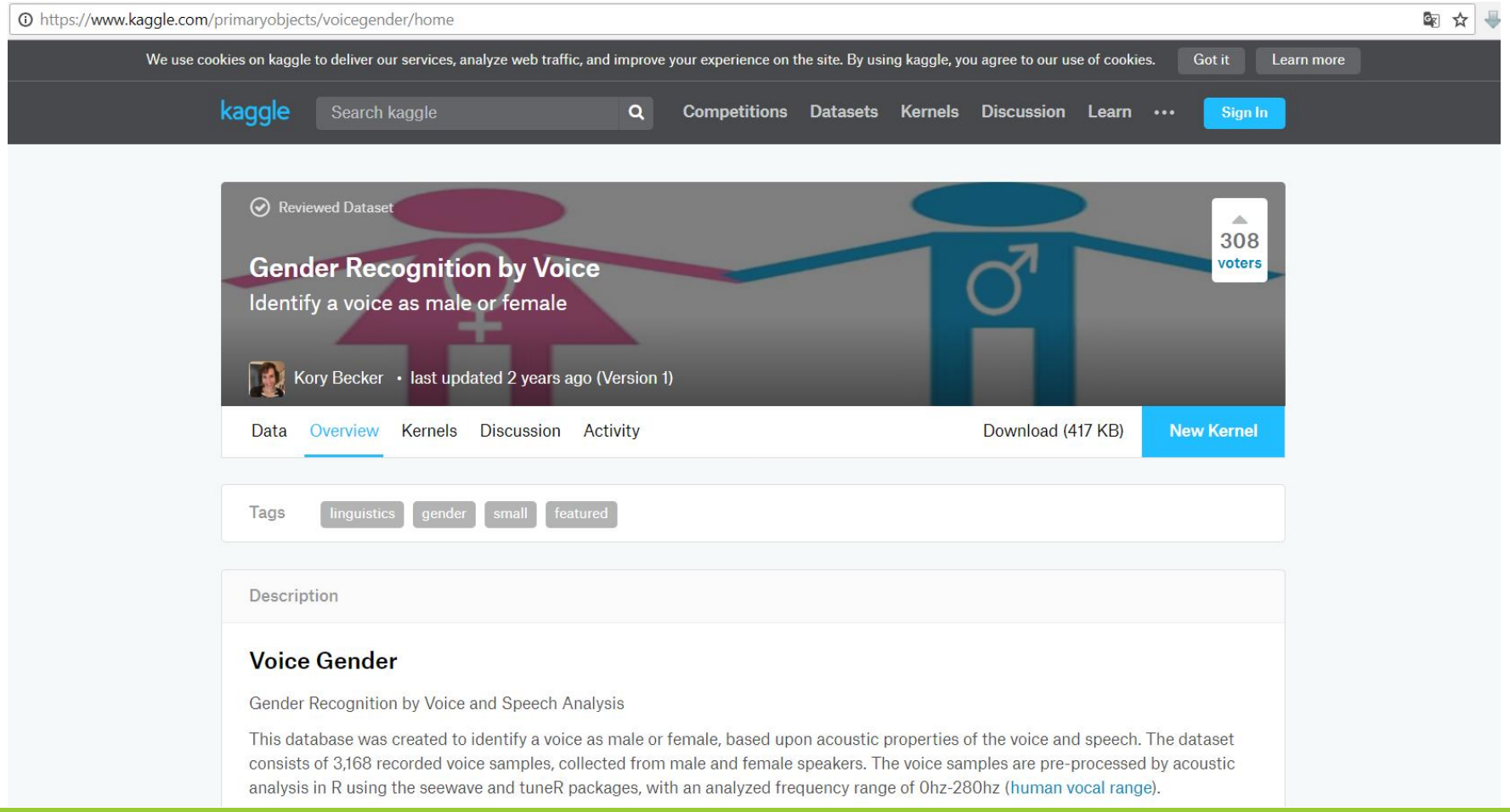
Reconocimiento de emociones



Reconocimiento de genero

UN EJEMPLO DE RECONOCIMIENTO DE PATRONES EN LA VOZ

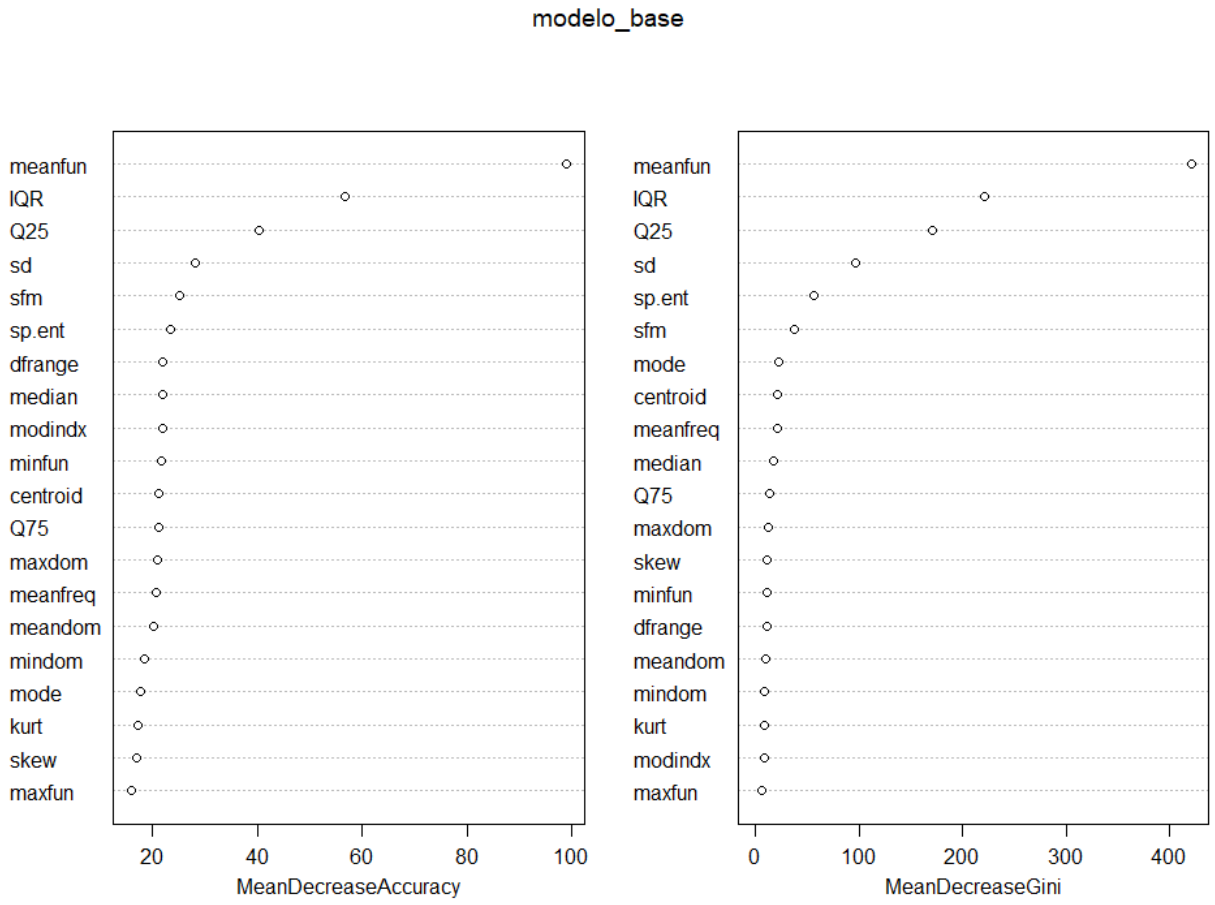
Reconocimiento de género



Reconocimiento de género

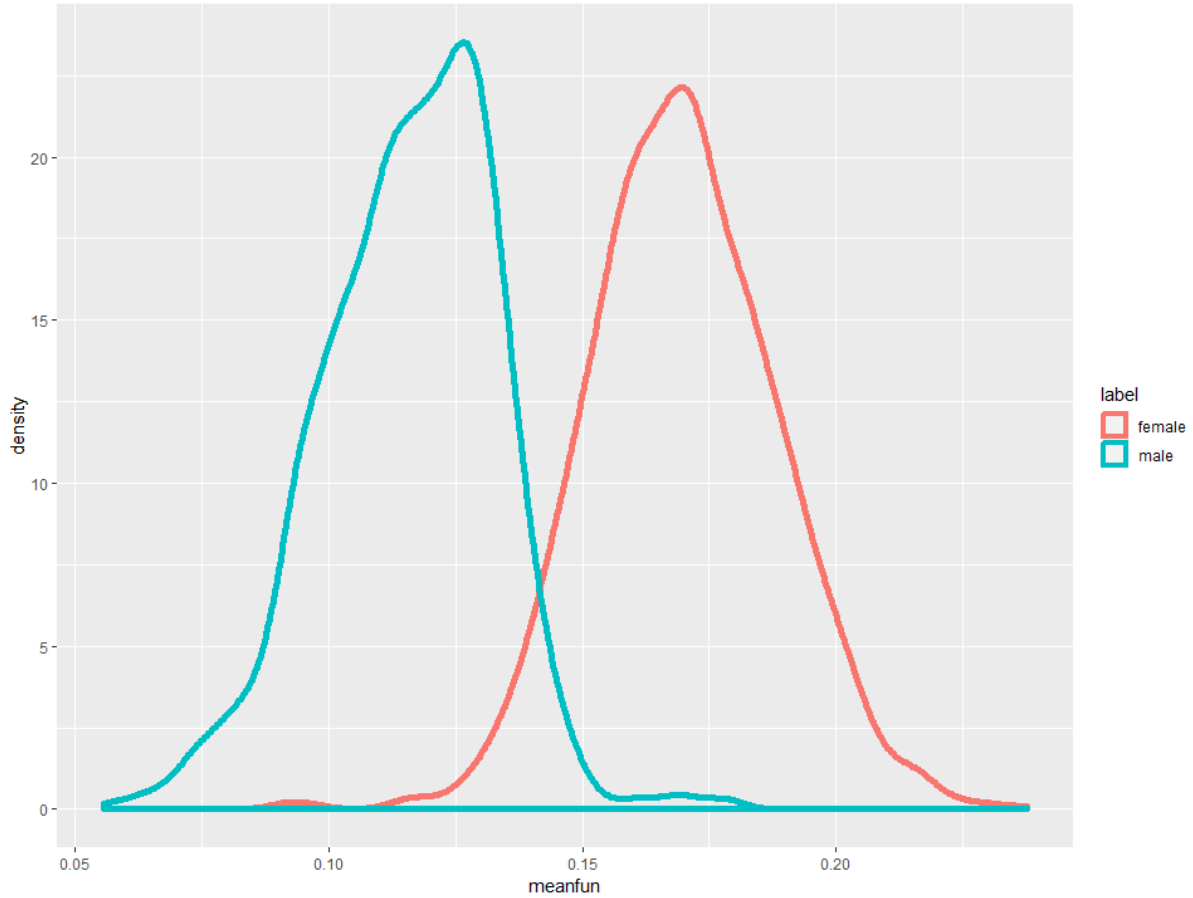
```
> str(datos)
'data.frame':   3168 obs. of  21 variables:
 $ meanfreq: num  0.0598 0.066 0.0773 0.1512 0.1351 ...
 $ sd      : num  0.0642 0.0673 0.0838 0.0721 0.0791 ...
 $ median  : num  0.032 0.0402 0.0367 0.158 0.1247 ...
 $ Q25     : num  0.0151 0.0194 0.0087 0.0966 0.0787 ...
 $ Q75     : num  0.0902 0.0927 0.1319 0.208 0.206 ...
 $ IQR     : num  0.0751 0.0733 0.1232 0.1114 0.1273 ...
 $ skew    : num  12.86 22.42 30.76 1.23 1.1 ...
 $ kurt    : num  274.4 634.61 1024.93 4.18 4.33 ...
 $ sp.ent  : num  0.893 0.892 0.846 0.963 0.972 ...
 $ sfm     : num  0.492 0.514 0.479 0.727 0.784 ...
 $ mode    : num  0 0 0 0.0839 0.1043 ...
 $ centroid: num  0.0598 0.066 0.0773 0.1512 0.1351 ...
 $ meanfun : num  0.0843 0.1079 0.0987 0.089 0.1064 ...
 $ minfun  : num  0.0157 0.0158 0.0157 0.0178 0.0169 ...
 $ maxfun  : num  0.276 0.25 0.271 0.25 0.267 ...
 $ meandom : num  0.00781 0.00901 0.00799 0.2015 0.71281 ...
 $ mindom  : num  0.00781 0.00781 0.00781 0.00781 0.00781 ...
 $ maxdom  : num  0.00781 0.05469 0.01562 0.5625 5.48438 ...
 $ dfrange : num  0 0.04688 0.00781 0.55469 5.47656 ...
 $ modindx : num  0 0.0526 0.0465 0.2471 0.2083 ...
 $ label   : Factor w/ 2 levels "female","male": 2 2 2 2 2 2 2 2 2 2
```

Reconocimiento de género




```
> CrossTable(tprueba$label, results, prop.chisq = FALSE,
+            prop.c = TRUE, prop.r = TRUE, dnn = c("actual gender",
+                                                  "predicted gender"))


   Cell Contents
|-------------------------|
|                       N |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  792


             | predicted gender
actual gender |    female |      male | Row Total |
--------------|-----------|-----------|-----------|
       female |       372 |         5 |       377 |
              |     0.987 |     0.013 |     0.476 |
              |     0.982 |     0.012 |           |
              |     0.470 |     0.006 |           |
--------------|-----------|-----------|-----------|
         male |         7 |       408 |       415 |
              |     0.017 |     0.983 |     0.524 |
              |     0.018 |     0.988 |           |
              |     0.009 |     0.515 |           |
--------------|-----------|-----------|-----------|
 Column Total |       379 |       413 |       792 |
              |     0.479 |     0.521 |           |
--------------|-----------|-----------|-----------|


> cat(sum(diag(mc))/sum(mc) * 100,"% casos correctamente clasificados\n")
98.48485 % casos correctamente clasificados
```

Reconocimiento de género



Servicios

SERVICIOS SOBRE LA VOZ

Servicios cognitivos basados en voz

- Existe otros APIs de servicios a diferentes niveles:

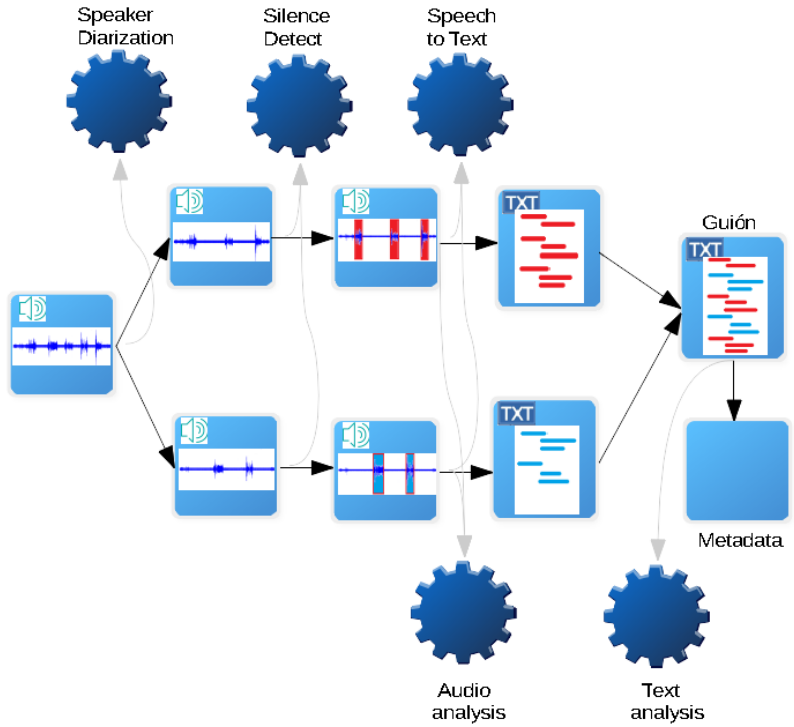
a) CMU Sphinx (works offline)
b) Google Speech Recognition
c) Wit.ai
d) Microsoft Bing Voice Recognition
e) Houndify API
f) IBM Speech to Text
g) Snowboy Hotword Detection (works offline)

Proyectos

ALGUNOS EJEMPLOS DE PROYECTOS DE CIENCIA DE DATOS

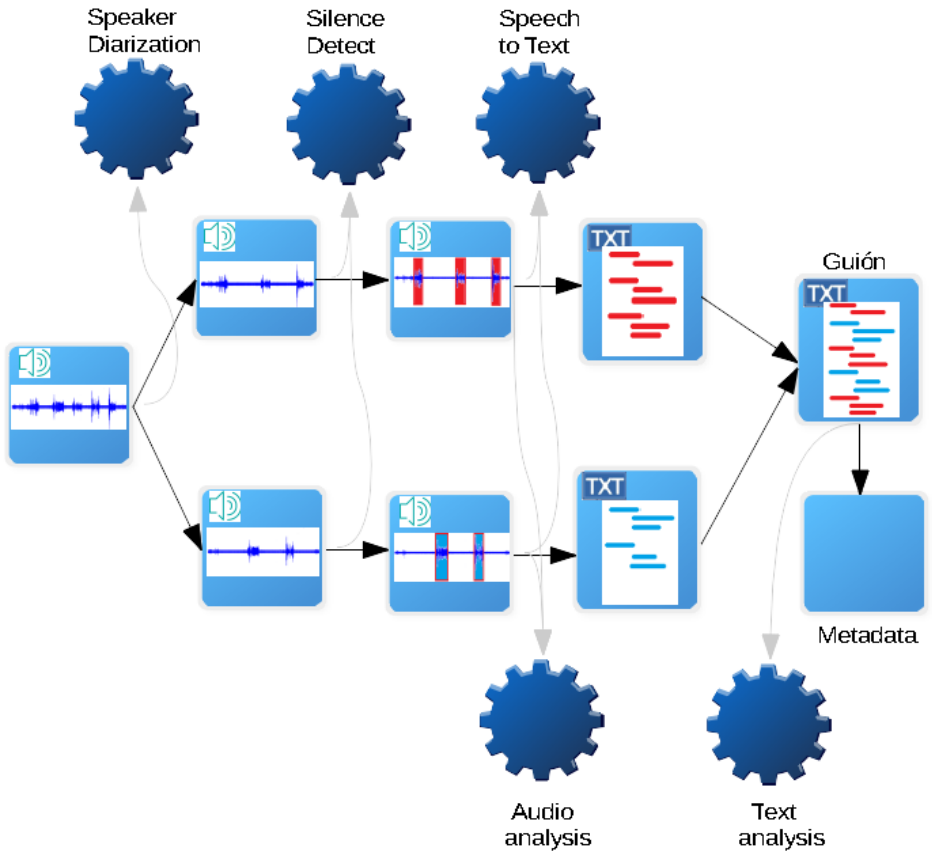
Speech Analytics POC

POC: *análisis de voz*



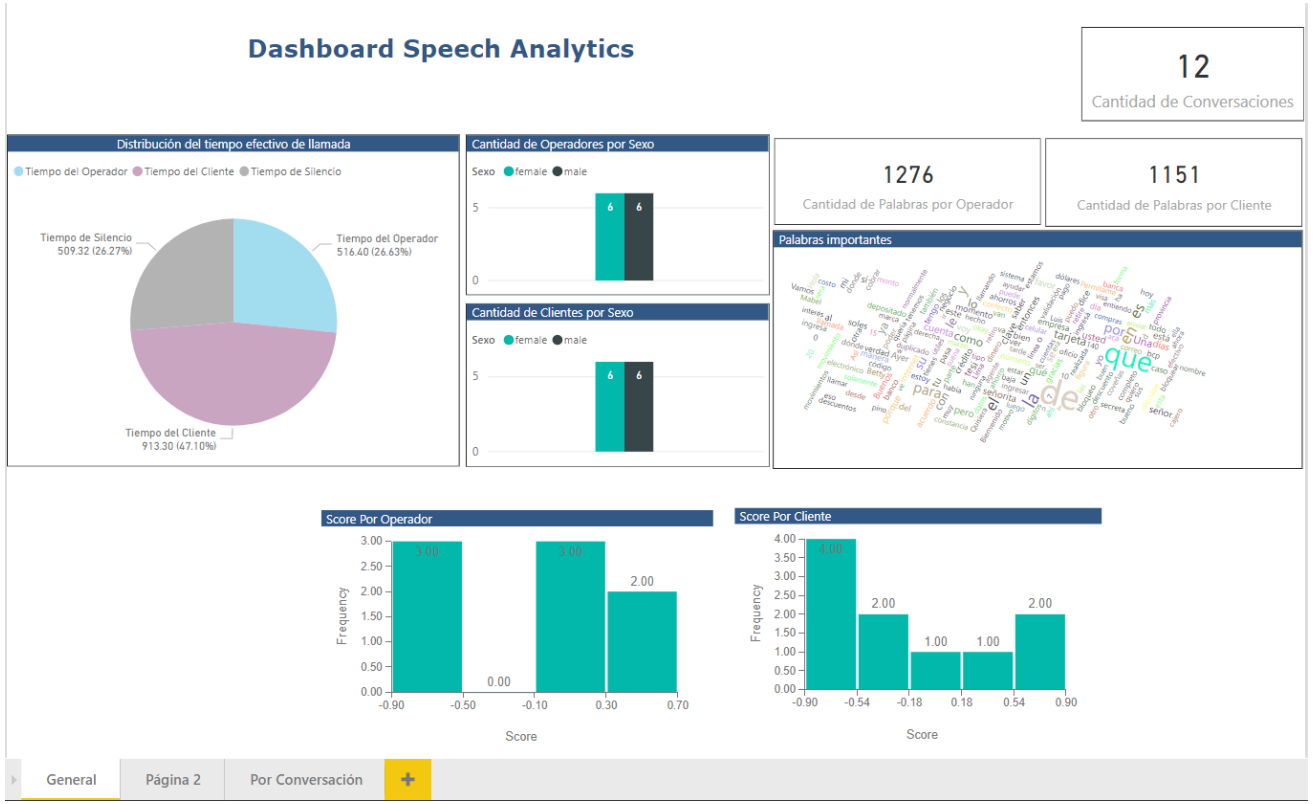
POC – Speech Analytics

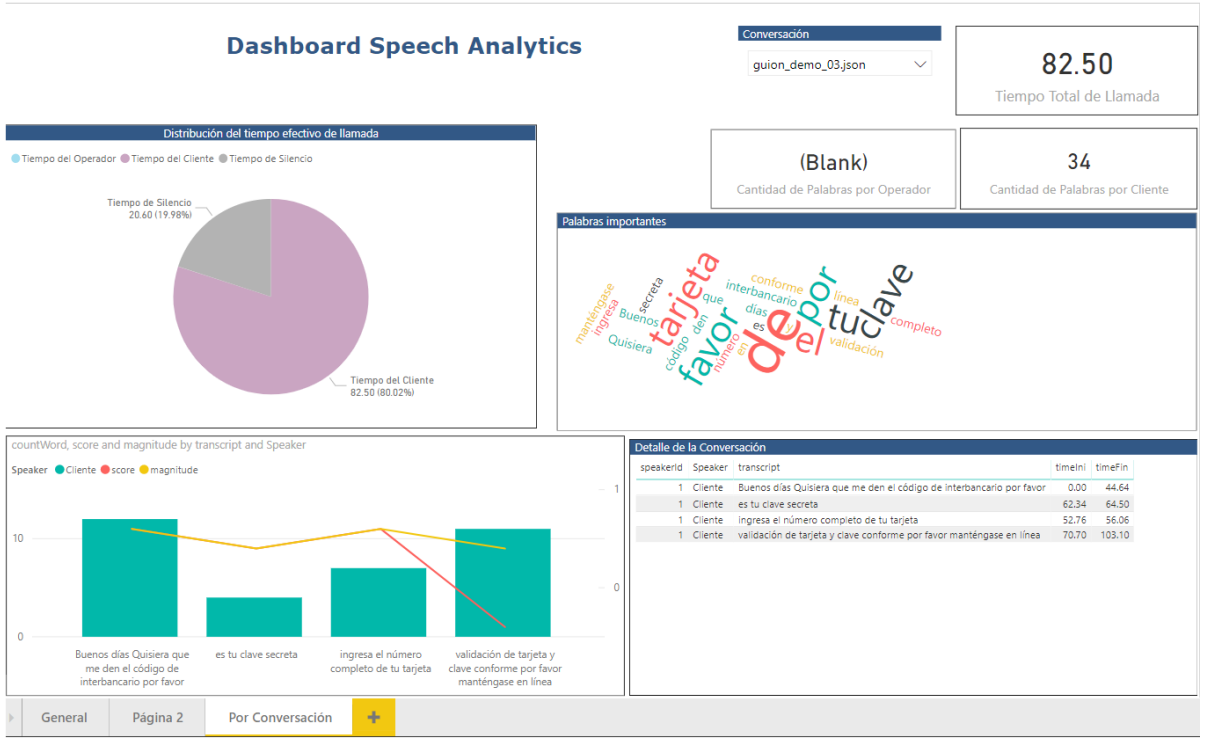
POC: *análisis de voz*



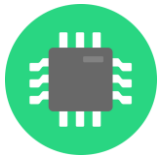

```
{
  "guion":[
    {
      "SentenceNum":0,
      "SentenceNumSpeaker":0,
      "confidence":0.88210535,
      "metadata":{
        "magnitude":0.20000000298023224,
        "score":-0.20000000298023224
      },
      "speakerId":0,
      "time":[
        0.0,
        4.98
      ],
      "transcript":"Buenos días Te estoy llamando para ver si me pueden dar la clave de internet"
    },
    {
      "SentenceNum":1,
      "SentenceNumSpeaker":1,
      "confidence":0.9268964,
      "metadata":{
        "magnitude":0.5,
        "score":0.5
      },
      "speakerId":0,
      "time":[
        11.24,
        15.200000000000001
      ],
      "transcript":"la clave de internet muy bien Vamos a verificarlo momento por favor"
    },
```

POC – Speech Analytics





Tecnología utilizada



- Python 2.7 y 3.6
  - PyDub - 3.6
  - PyAudioAnalysis - 2.7
  - GoogleCloud - 3.6
- R
  - Reconocimiento de género en audios
- Programas para manipulación de audios
  - Sox
  - Ffmpeg

- Google Cloud Platform (GCP)
  - Speech Recognition
  - Sentiment Analysis
- Microsoft Azure
  - Ubuntu Virtual Machine

Everis virtual agent - eVA

eva is transactional service

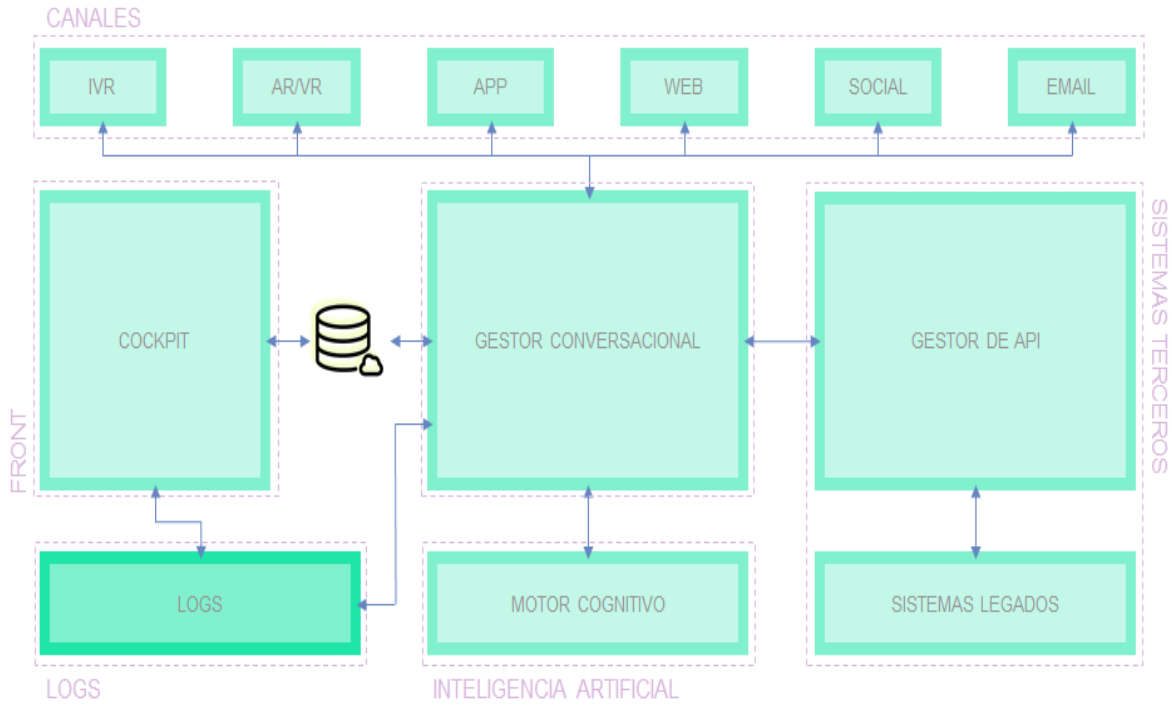
Our AI is capable of:

- Perform **transactions**.
- Provide **relevant information**.
- Walk through **conversational flow**.
- **Memorize** customer information to quickly solve their problems.
- Uses **next best actions** rules.

http://eva.bot/

eva is IVR

- **Enrich your interactions** through voice recognition.
- Multiply your channels using **phone calls** and **voice commands**.
- **Text to Speech** and **Speech to Text** features.
- **Multi-language** supported.
- Our AI understands the **language of your industry**.



GRACIAS

JOSÉ R. SOSA
email: jsosabri@everis.com
Linkedin: https://ve.linkedin.com/in/josersosa
Twitter: http://www.twitter.com/josersosab