



## Introducción



1. El proceso del habla
2. Voz como señal acústica
3. Conceptos de codificación
4. Reconocimiento de Voz
5. Verificación de locutor
6. Síntesis de voz

## 1.- El proceso del habla

Contexto Semántico  
Reglas de Lenguaje

Contexto Semántico  
Reglas de Lenguaje



Señal  
Acústica



## 1.- El proceso del habla

Modelo de capas para la comunicación

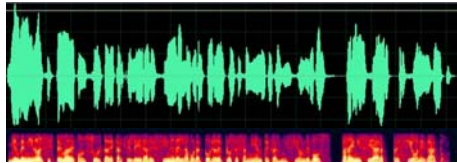
Contexto Semántico <i>Escuela, polola, casa, fútbol, etc....</i>
Reglas de Lenguaje <i>Léxico, Sintaxis, Fonética, Entonación</i>
Señal Acústica (onda mecánica)

## 2.- Voz como señal acústica

- Fonemas sonoros y sordos.
- Periodicidad y cuerdas vocales.
- Formantes y tracto vocal.
- Información fonética-acústica.

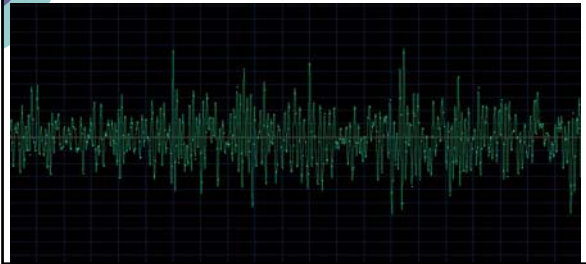
## 2.- Voz como señal acústica

- Señal completa:



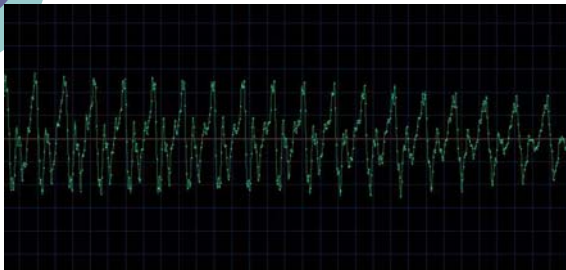
## 2.- Voz como señal acústica

- Señal no periódica "s":



## 2.- Voz como señal acústica

- Señal periódica "a":



## 3.- Codificación y Compresión

- La señal de voz tiene mucha redundancia.
- No es necesario transmitir toda la forma de onda para reproducir la señal de voz manteniendo la inteligibilidad necesaria para la comunicación.
- En una conversación, además de la información acústica, los contextos semántico y lingüístico juegan un papel fundamental.

### 3.- Codificación y Compresión

o Hay diferentes tipos de codificación:

- Por forma de onda temporal.
- Por espectro.
- Con modelos del aparato fonador.

### 4.- Reconocimiento de Voz (RV)

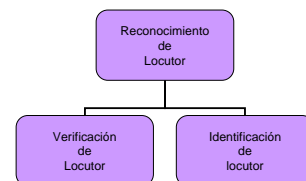
- Reconocer la información lingüística de la señal.
- RV puede ser independiente del locutor o dependiente del locutor.
- RV puede ser de palabra aislada o continua.
- Diferentes capas de modelamiento.
- Word Spotting, aislamiento de palabras llaves de contexto (e.g. document retrieval, espionaje).
- Lenguaje Natural

### 4.- Reconocimiento de Voz (RV)

Diferentes capas de modelamiento

Semántico
Frase
Palabra
Acústica (fonética)

### 5.- Verificación del locutor



La verificación del locutor (VL) puede ser texto dependiente o independiente



## 6.- Síntesis de voz o TTS (text-to-speech)

Generar voz a partir de texto es mas fácil que el reconocimiento de voz o de locutor sin embargo resulta difícil conseguir naturalidad.

La información prosódica es fundamental.