

# CAMPUS CIENTÍFICOS DE VERANO 2020

¡VIVE LA CIENCIA!

Publicado en: CCV (<https://www.campuscientificos.es>)

[Inicio](#) > Tratamiento Digital de Voz

## Tratamiento Digital de Voz

Campus:  
UAM-CSIC  
Departamento:  
Teoría de la señal y comunicaciones  
Área:  
Tecnología electrónica y de las comunicaciones (COM)  
Provincia:  
Madrid

### Resumen:

El proyecto presentará al alumnado los fundamentos y las principales aplicaciones del tratamiento digital de la voz. Se pretende que los/as alumnos/as entiendan de una forma didáctica e intuitiva las características de la señal de voz y cómo éstas se pueden explotar para dar lugar a aplicaciones como el reconocimiento automático de habla o la síntesis de voz. De esta forma, con este conjunto de aplicaciones, se podrá resolver el enigma planteado en el proyecto: descubrir al/la secuestrador/a.

### 1ª Sesión:

Se aprenderá lo necesario del software MATLAB para desarrollar las prácticas, en particular aprender a grabar y reproducir audio en MATLAB y ejecutar programas (los facilitados por los/as profesores/as) sobre dicho audio. Experimentar con la señal de voz propia gracias a los programas facilitados por los/as profesores/as que permitirán analizar el espectro (la representación en frecuencia) de la señal, así como los formantes y el tono o frecuencia fundamental.

Para la realización de esta práctica se requiere, como en todas las demás de este curso, un PC con MATLAB instalado y el Signal Processing Toolkit, cascos con micrófono incorporado para cada pareja y los programas desarrollados por el profesorado como apoyo a este curso.

Las actividades a desarrollar consistirán en lo siguiente:

1. Introducción a MATLAB como herramienta básica sobre la que desarrollar todo el trabajo práctico del curso. Se explicará cómo grabar la propia voz, representarla y reproducirla utilizando MATLAB.
2. El segundo bloque va dirigido a excitar la curiosidad del alumnado mediante el análisis de su propia voz. La representación habitual de la señal de voz ofrece muy poca información de la señal de voz, así que se presentarán (con el apoyo de programas desarrollados por los profesores) representaciones alternativas como la representación en frecuencia y la representación frecuencial-temporal mediante espectrogramas 3D y 2D. La representación frecuencial-temporal introducirá de forma natural los conceptos de formante y frecuencia fundamental.

### 2ª Sesión:

Los objetivos de esta sesión son que los/as estudiantes conozcan los conceptos básicos de la producción de voz humana y que sean capaces de simular ese sistema de producción humana en MATLAB (con ayuda de programas y datos proporcionados por los profesores). Un objetivo intermedio para conseguir esto es aplicar varios filtrados a la señal de voz y observar su efecto acústico y en la representación tiempo-frecuencia.

Para la realización de esta práctica se requiere, como en todas las demás de este curso, un PC con MATLAB instalado y el Signal Processing Toolkit, cascos con micrófono incorporado para cada pareja y los programas desarrollados por el profesorado como apoyo a este curso. También se necesitará un cañón proyector para mostrar las presentaciones más teóricas y los ejemplos sobre una pantalla a todos/as los/as alumnos/as.

Las actividades a desarrollar en esta sesión consistirán en lo siguiente:

1. En primer lugar se realizará una breve exposición teórica sobre el funcionamiento del aparato fonador humano y cómo simularlo en un

ordenador.

2. En segundo lugar los/as alumnos/as aprenderán a filtrar la voz y a observar el efecto que produce el filtrado en la percepción de la voz, así como en la representación tiempo-frecuencia.

3. En tercer y último lugar, los/as estudiantes realizarán (con ayuda de programas y datos facilitados por el profesorado) síntesis de vocales (idealmente a partir de datos extraídos de su propia voz) y jugarán con la variación de parámetros como el tono (para simular voz masculina o femenina) y la vibración de las cuerdas vocales (para simular voz susurrada). También podrán crear sencillas canciones musicales a partir de la propia síntesis de vocales.

3ª Sesión:

Los objetivos fundamentales de esta práctica son que los/as estudiantes entiendan los mecanismos de percepción de la voz humana y los apliquen para entender cómo funcionan los sistemas automáticos de reconocimiento de voz. Como objetivo concreto se pretende que el alumnado compruebe empíricamente la importancia de los formantes en el reconocimiento de las vocales y que lleguen a representar su propio triángulo vocálico.

Para la realización de esta práctica se requiere, como en todas las demás de este curso, un PC con MATLAB instalado y el Signal Processing Toolkit, cascos con micrófono incorporado para cada pareja y los programas desarrollados por el profesorado como apoyo a este curso. También se necesitará un cañón proyector para mostrar las presentaciones más teóricas y los ejemplos sobre una pantalla a todos/as los/as alumnos/as.

Las actividades a desarrollar en esta sesión consistirán en lo siguiente:

1. En primer lugar se realizará una breve exposición teórica sobre el funcionamiento del aparato auditivo humano y cómo simularlo en un ordenador.

2. En segundo lugar los/as estudiantes analizarán (gracias a los programas facilitados por los profesores) las posiciones de los formantes en grabaciones de sus propias vocales, dibujando el triángulo vocálico que representa las vocales en función de la posición de dos formantes.

3. Como continuación de esta práctica y proyecto final se propondrá a los/as alumnos/as realizar un reconocedor de vocales que permita clasificar una grabación como una de las cinco vocales del castellano.

4ª Sesión:

Esta sesión se plantea como una sesión de repaso, aclaración de dudas y preparación del trabajo final.

Para la realización de esta práctica se requiere, como en todas las demás de este curso, un PC con MATLAB instalado y el Signal Processing Toolkit, cascos con micrófono incorporado para cada pareja y los programas desarrollados por el profesorado como apoyo a este curso. También se necesitará un cañón proyector para mostrar las presentaciones más teóricas y los ejemplos sobre una pantalla a todos/as los/as alumnos/as.

Dado que esta práctica es una práctica de repaso, no se puede precisar de antemano las actividades a desarrollar, pues éstas dependerán de las dudas que surjan y del desarrollo de las sesiones previas.

5ª Sesión:

Esta quinta sesión irá enfocada a la presentación del trabajo final por parte de los/as estudiantes, a fin de resolver el enigma que se ha presentado al inicio del proyecto: descubrir, a partir del tratamiento digital de la voz, quién o quiénes son los/as responsables del secuestro.

Referencias recomendadas:

- MATLAB Signal Processing Toolbox.
- Hugo L. Rufiner, Diego H. Milone, "Sistema de reconocimiento automático del habla", Ciencia, Docencia y Tecnología, vol. XV, núm. 28, mayo, 2004, pp. 151-177

Lugar donde se desarrollará el proyecto:

Universidad Autónoma de Madrid  
Escuela Politécnica Superior  
Ciudad Universitaria de Cantoblanco  
28049 Madrid