



*1821 Universidad de Buenos Aires*

## **Resolución Consejo Directivo**

**Número:**

**Referencia:** EX-2024-00077963- -UBA-DMESA#FCEN - POSTGRADO - Sesión  
11/03/2024

---

**VISTO:**

La nota presentada por la Dirección del Departamento de Computación,  
mediante la cual eleva la información del curso de posgrado Procesamiento de Audio  
con Redes Neuronales para el año 2024,

**CONSIDERANDO:**

lo actuado por la Comisión de Doctorado,

lo actuado por este Cuerpo en la sesión realizada el día 11 de marzo de 2024,

en uso de las atribuciones que le confiere el Artículo 113° del Estatuto  
Universitario,

**EL CONSEJO DIRECTIVO DE LA FACULTAD**

## DE CIENCIAS EXACTAS Y NATURALES

### RESUELVE:

**ARTÍCULO 1º:** Aprobar el nuevo curso de posgrado **Procesamiento de Audio con Redes Neuronales** de 64 horas de duración, que será dictado por el Dr. Diego Fernández Slezak con la colaboración del Dr. Pablo Riera.

**ARTÍCULO 2º:** Aprobar el programa del curso de posgrado **Procesamiento de Audio con Redes Neuronales** que como anexo forma parte de la presente Resolución, para su dictado en el primer cuatrimestre de 2024.

**ARTÍCULO 3º:** Aprobar un puntaje máximo de tres (3) puntos para la Carrera del Doctorado.

**ARTÍCULO 4º:** Establecer un arancel de **CATEGORÍA BAJA**, estableciendo que dicho arancel estará sujeto a los descuentos y exenciones estipulados mediante la Resolución CD N.º 1072/19. Disponer que los fondos recaudados ingresen en la cuenta presupuestaria habilitada para tal fin, y sean utilizados de acuerdo a la Resolución 072/03

**ARTÍCULO 5º:** Disponer que, de no mediar modificaciones en el programa, la carga horaria y el arancel, el presente Curso de Posgrado tendrá una vigencia de cinco (5) años a partir de la fecha de la presente Resolución.

**ARTÍCULO 6º:** Comuníquese a todos los Departamentos Docentes, a la Dirección de Estudiantes y Graduados, a la Biblioteca de la FCEyN y a la Secretaría de Posgrado con copia del programa incluida. Cumplido, pase COMPUTACION#FCEN y resérvese.

## ANEXO

### PROGRAMA

#### Programa:

Esta materia busca introducir temáticas del procesamiento de habla, desde el punto de vista del procesamiento de señales clásico y también del de las redes neuronales. Se busca que los participantes logren llevarse los conceptos básicos del procesamiento de señales clásico (filtros, análisis de Fourier) para luego pasar a métodos actuales basados en redes neuronales (autoencoders, self-supervised learning) teniendo en cuenta el uso de estas herramientas para tareas de análisis de audio y tareas de reconocimiento del habla.

La materia busca una mezcla balanceada de teoría y ejercitación práctica.

#### Temario:

Elementos de acústica y procesamiento digital de señales. Elementos de fonología y producción de habla. Análisis de Fourier y filtros. Elementos de estadística y aprendizaje automático. Elementos de procesamiento del lenguaje natural. Reconocimiento del habla. Modelos ocultos de Markov (HMM) y redes neuronales profundas. Aprendizaje de representaciones. Temas avanzados: reconocimiento de eventos sonoros y géneros musicales; extracción de información del hablante; detección de emociones;

- Programa (tentativo, pueden ser más o menos clases, sumado a las clases de laboratorio):

#### Clase 1: Introducción al Procesamiento de Señales

- ¿Qué es el procesamiento de señales? Señales continuas vs. señales discretas. Muestreo y cuantización.
- Convolución y correlación. Auto-correlación y cross-correlación. Filtrado en el dominio del tiempo.

#### Clase 2: Análisis en el Dominio de la Frecuencia

- Análisis de Fourier: series de Fourier y transformadas de Fourier. Transformada

de Fourier Discreta (DFT) y Transformada Rápida de Fourier (FFT).

### Clase 3: Análisis en el Dominio de la Frecuencia

- STFT
- Filtrado y convolución en el dominio de la frecuencia.
- Análisis espectral

### Clase 4: Procesamiento Estadístico de Señales

- Modelado de señales con distribuciones de probabilidad. Modelos gaussianos y no gaussianos. Detección y estimación en señales ruidosas. Relación señal-ruido (SNR).
- Introducción al filtrado óptimo y el filtro de Wiener.

### Clase 5: Introducción al Procesamiento de Audio y Habla

- Características acústicas de las señales de audio y habla.
- Fundamentos de la producción y percepción de sonidos ambientales, musicales y habla.
- Atributos para el reconocimiento de música y habla. Chromagramas, MFCC.

### Clase 5: Descubrimiento de unidades

- Clustering. GMMs
- Segmentación de señales
- Identificación de anomalías

### Clase 6: Modelado de series y secuencias

- Modelos ARMA
- Modelos Ocultos de Markov (HMMs)

### Clase 7: Reconocimiento de Habla Clásico

- GMM-HMMS
- Decodificación con transductores de estados finitos (FSTs)

#### Clase 8: Introducción al Aprendizaje Profundo

- Arquitectura de redes neuronales. Backpropagation y descenso de gradiente. Funciones de activación y funciones de pérdida.
- Introducción a redes neuronales profundas para señales. Redes Neuronales Convolucionales (CNNs) y Redes Neuronales Recurrentes (RNNs).
- Atención. Transformers
- Modelado de secuencias: seq2seq, Encoder-Decoder

#### Clase 9: Aprendizaje Profundo en el Procesamiento de Señales

- Autoencoders profundos para el aprendizaje de características.
- Aprendizaje por transferencia para el procesamiento de señales. Aprendizaje auto-supervisado.

#### Clase 10: Modelos generativos para síntesis de audio

- Compresión de audio
- Supresión de ruido
- Separación de fuentes

#### Clase 11: Reconocimiento de Habla 2

- DNN-HMM
- Reconocimiento de habla de extremo a extremo con aprendizaje profundo.

#### Clase 12: Temas extras

- Reconocimiento de eventos sonoros
- Reconocimiento géneros musicales
- Extracción de información del hablante (identidad, edad, género)
- Detección de emociones

## **Bibliografia:**

- Smith, J. O. (2007). Introduction to digital filters: with audio applications (Vol. 2). Julius Smith.
- Daniel Jurafsky & James H. Martin, "Speech and Language Processing" (3er edition). Prentice Hall, 2023. (<https://web.stanford.edu/~jurafsky/slp3/>)
- Keith Johnson, "Acoustic and Auditory Phonetics" (2nd edition). Blackwell, 2003.
- Jacob Benesty, M. Mohan Sondhi & Yiteng Huang (Eds.), "Springer Handbook of Speech Processing". Springer-Verlag, 2008.
- "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016
- Purwins, H., Li, B., Virtanen, T., Schlüter, J., Chang, S. Y., & Sainath, T. (2019). Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13(2), 206-219.

Mohamed, A., Lee, H. Y., Borgholt, L., Havtorn, J. D., Edin, J., Igel, C., ... & Watanabe, S. (2022). Self-supervised speech representation learning: A review. *IEEE Journal of Selected Topics in Signal Processing*