

TDA de Señales EEG Conectividad Funcional Respuesta de Infantes a Audios

Ignacia Gothe
Poryecto 1-b

Pontificia universidad catolica
15 diciembre 2025

Dataset

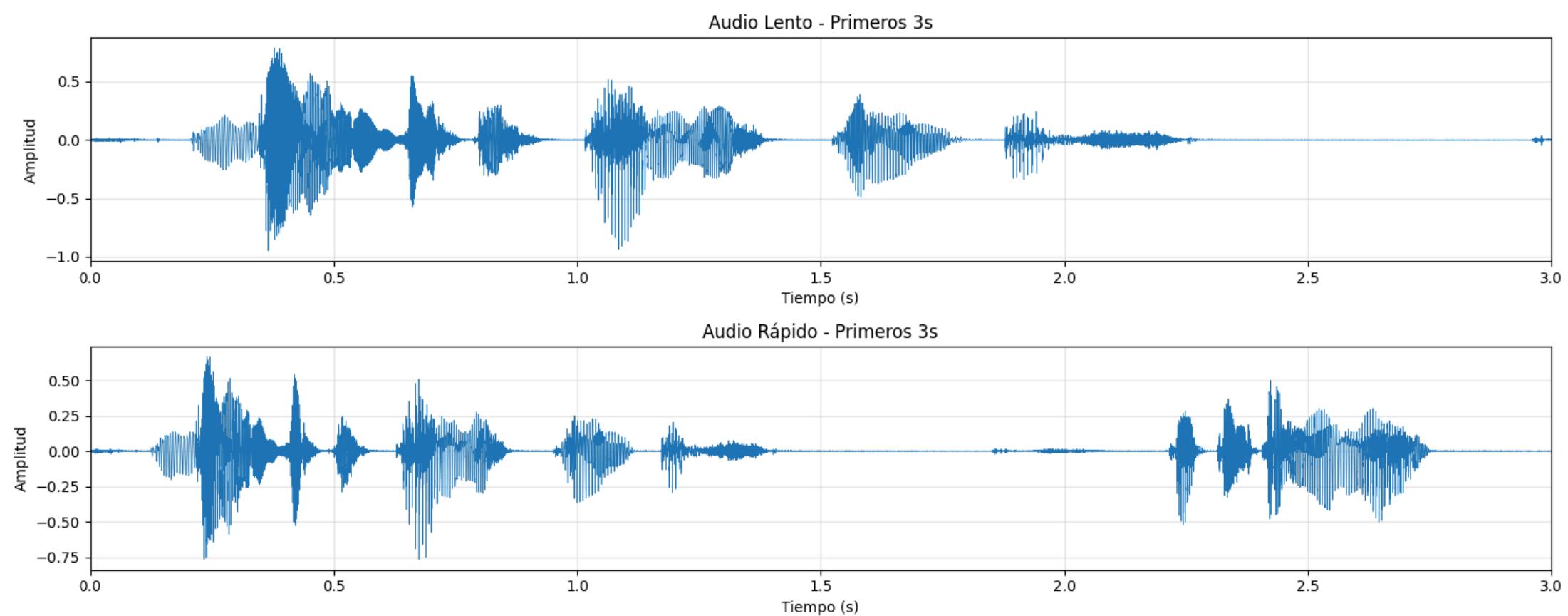
Datos consisten en 1,416 grabaciones EEG de 45 infantes distintos. Para cada uno para se registro respuesta cognitiva bajo dos condiciones experimentales:

- Audio **lento** → 710 grabaciones
- Audio **rápido**→706 grabaciones

EEG de 47 electrodos cerebrales por grabación

Frecuencia de muestreo: 250 Hz

Duración señales: ~20 segundos de grabación promedio por archivo



Condición rápida es el mismo audio, pero acelerado, por lo que tiene menor duración.

Preprocesamiento

1

Filtrado en
bandas de
frecuencia

2

Sliding windows
con overlap para
cada banda

3

Calculo de
correlaciones
entre electrodos

4

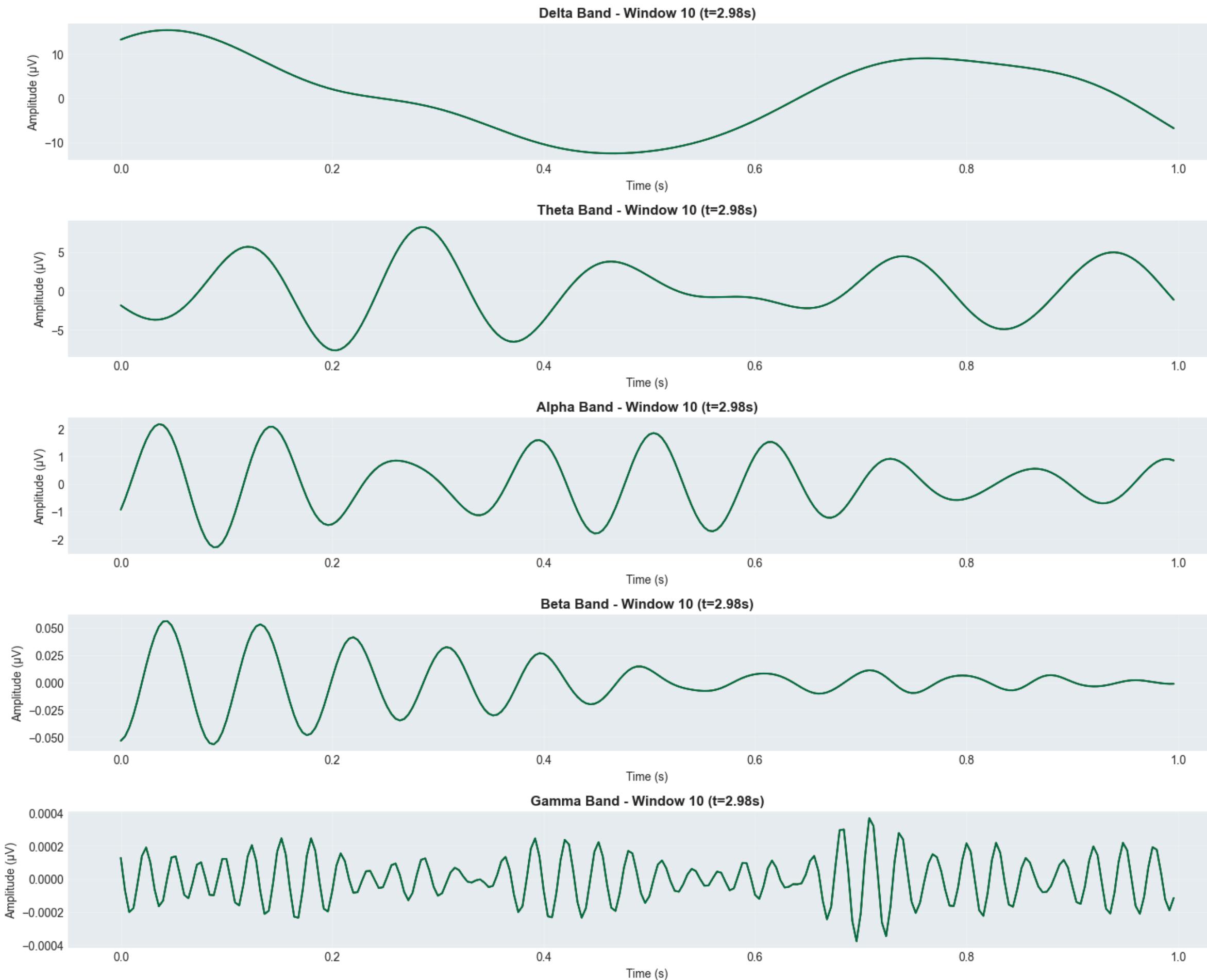
Construccion
de Grafos
pesados

Bandas de frecuencia

Cada señal es filtrada en las bandas de frecuencia, luego cada banda se segmenta en ventanas deslizantes de 1s y 50% overlap

Se calcula la correlación entre las ventanas de cada banda de todos los electrodos.

Grafo pesado: Cada electrodo es un nodo y la arista se le asigna un peso igual a 1-correlación



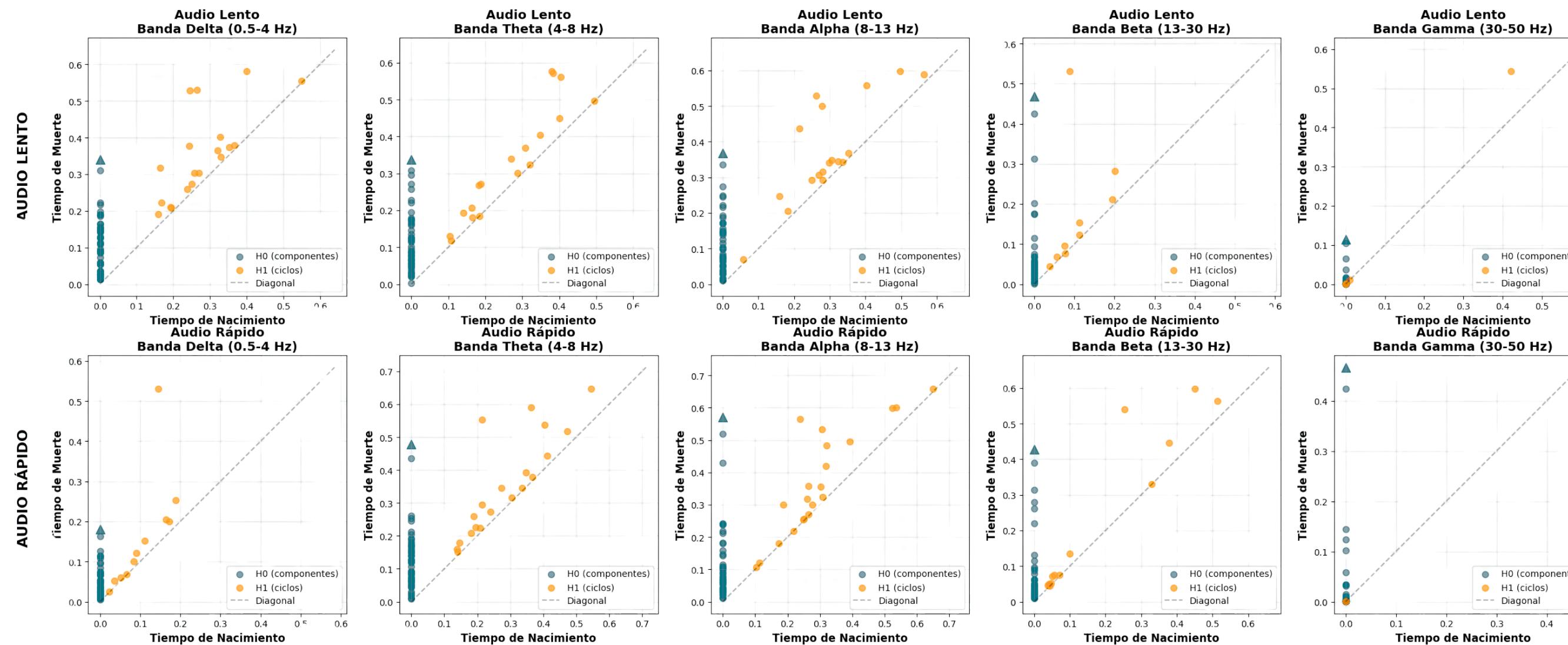
Pipeline

- 1** Filtración complejo de Vietoris–Rips
- 2** Homología persistente
- 3** Extracción de Features
- 4** Vectorización
- 5** Entrenamiento Random Forest
- 6** Clasificación

Filtración vietoris-rips

Inicialmente cada electrodo se considera un punto aislado. Al aumentar el umbral de distancia, agregas aristas entre puntos y luego simplices de orden superior.

→ Genera **familia de grafos anidados** que describe la red a múltiples escalas.



Homología persistente

Sobre la filtración, se calcula la homología persistente en H0 (componentes conexas) e H1 (ciclos de conectividad funcional.).
Salida: diagramas de persistencia (pares nacimiento–muerte) por dimensión.

Sobre la filtración, se calcula la homología persistente en H0 (componentes conexas) e H1 (ciclos de conectividad funcional.).

Extracción y agregación

De cada diagrama de persistencia se extrajen 11 features para H0 y otras 11 para H1: counts, medias, desvios de birth/death, entropía de persistencia, etc.

Salida: Ventanas con 22 features. *Cada banda tiene muchas ventanas.*

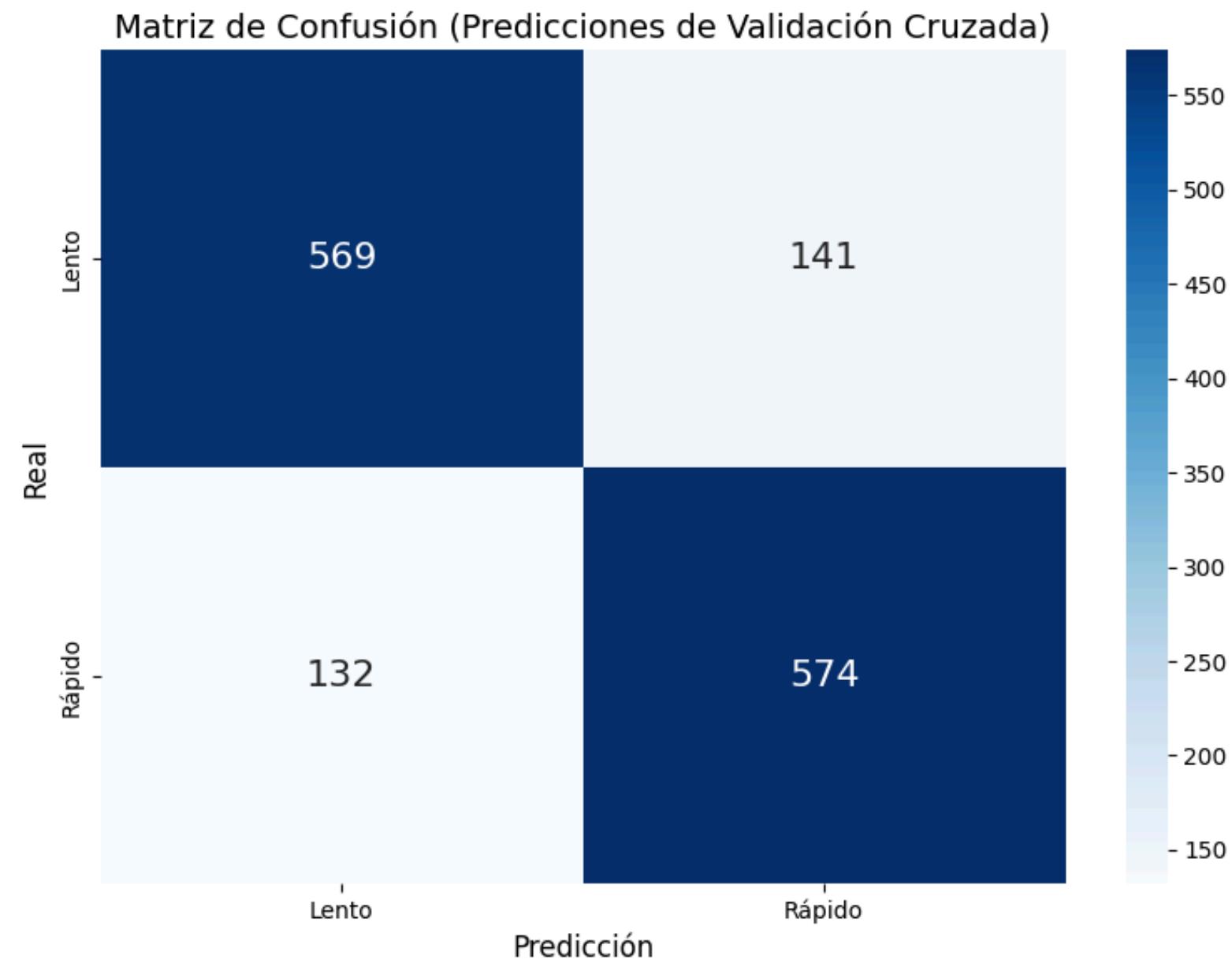
Vectorización

Para cada banda, se calculan la **media y desviación estándar** de las 22 features de sus ventanas. Con cinco bandas por grabacion y 44 features por banda → 44×5 por grabacion.

Salida: Cada grabación quedó representada por un vector de 220 características.

Este vector captura tanto las tendencias promedio de la topología de conectividad (mediante las medias) como su variabilidad temporal (mediante las desviaciones estándar)

Clasificación



Random Forest

Luego los vectores pueden usarse junto con las etiquetas de la grabacion (slow/fast) para entrenar un modelo ML. En este caso; Random Forest (100 árboles, profundidad máx 10)

GroupK-Fold

Para validar el modelo, en cada fold todas las ventanas de un mismo infante quedan juntas (en train o test). Evita *data leakage* y mide capacidad de **generalización real a infantes no vistos**.

Grid-Search

max_depth: 15
min_samples_leaf: 4
min_samples_split: 10
n_estimators: 200

Comparación:
Modelo original: 80.7%
Modelo optimizado: 81.6%

Resultados

- ROC-AUC: 0.887
- F1 Score (*weighted*): 0.816
- Accuracy: 0.82

0.80

Precision tomando
clase **Rapido** como la
clase positiva

0.84

Recall tomando clase
Rapido como la clase
positiva

0.83

Precision tomando
clase **Lento** como la
clase positiva

0.79

Recall tomando clase
Lento como la clase
positiva

Estudio estadístico

no sólo decir “accuracy alta”, sino mostrar que es muy improbable por azar y estimar incertidumbre.

Prueba de permutación

La prueba de permutación reveló que la precisión observada es significativamente superior al azar:

- Media de la distribución nula: 49.6%
- Precion observada CV: 80.7%
- Valor $p < 0.001$

Bootstrap confianza

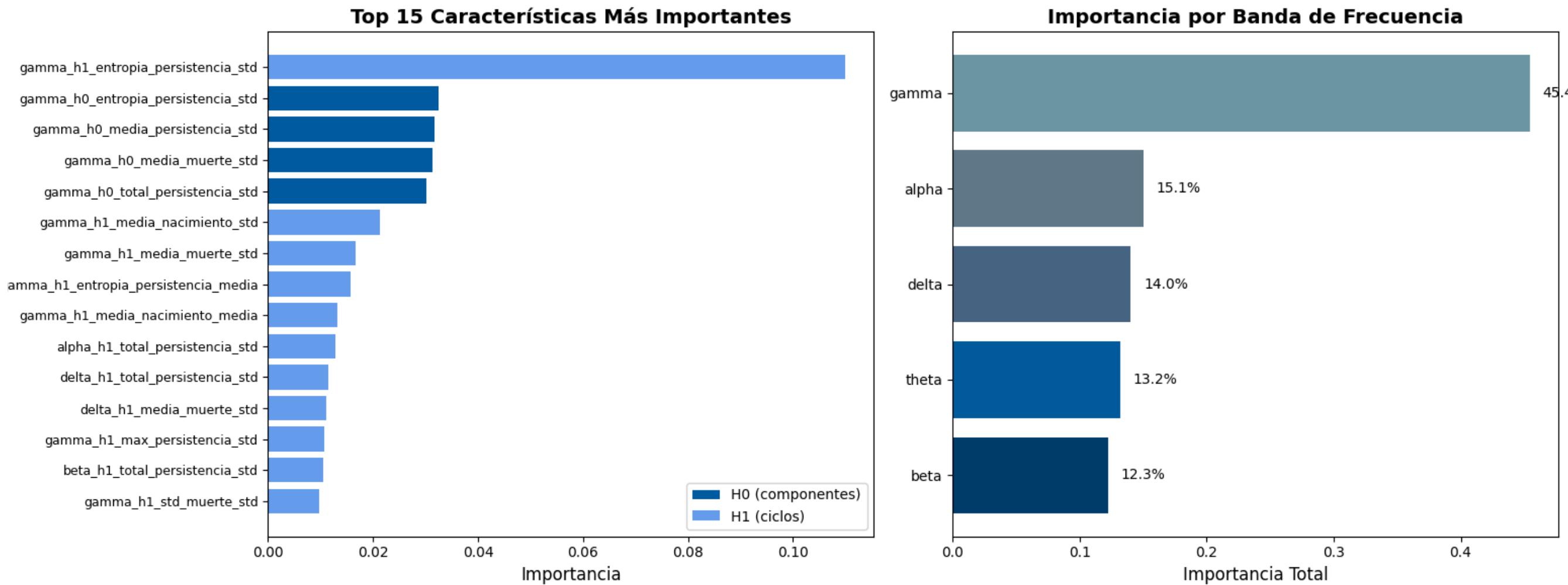
El intervalo de confianza bootstrap del 95%

IC 95%: [0.7387, 0.8180]

Valor observado: 0.8157

IC incluye al observado y excluye el azar 50%, confirmando la robustez del resultado.

Importancia features



Delta (0.5–4 Hz): Sueño profundo, actividad lenta global
Theta (4–8 Hz): Somnolencia, memoria y atención
Alpha (8–13 Hz): Relajación, inhibición y control atencional
Beta (13–30 Hz): Alerta, actividad cognitiva y sensorimotora
Gamma (30–100 Hz): Procesamiento sensorial rápido, integración y cognición

Conclusiones

La velocidad del audio (lento vs rápido) escuchado en infantes induce patrones de conectividad cerebral topológicamente distintos por lo que se puede discriminar desde la topología de la conectividad EEG

La banda gamma mostró ser la señal más informativa para el modelo, especialmente en la entropía H1 y después features de H0, muy sobre las otras. Esto era esperable ya que gamma se asocia con el **procesamiento cognitivo**

Que dominen features de H1 y su std temporal sugiere que lo discriminativo no es un patrón fijo, sino la dinámica de la organización en gamma (ciclos).

Los resultados establecen que la velocidad del audio escuchado en infantes induce patrones de conectividad cerebral topológicamente distintos.

Gracias

