

# **Análisis de una base de datos de fNIRS realizada en bebés de 4 meses en reposo**

Naomi Marianne Argüello Calderón

## **Introducción**

El estudio de la conectividad funcional en estado de reposo se ha convertido en una herramienta clave para comprender cómo se organiza el cerebro cuando no está involucrado en tareas específicas. Este tipo de conectividad se basa en la observación de fluctuaciones espontáneas en la actividad hemodinámica que, al sincronizarse entre diferentes zonas corticales, revelan la existencia de redes funcionales que emergen incluso en ausencia de estimulación externa. En el caso de los bebés, este enfoque resulta especialmente valioso: permite investigar la arquitectura funcional temprana sin requerir atención sostenida, interacción con estímulos o cumplimiento de instrucciones, y puede aplicarse mientras duermen de manera natural.

Aunque la resonancia magnética funcional ha sido el método predominante para caracterizar estas redes, su uso con infantes implica retos prácticos importantes como el ruido del escáner, la necesidad de inmovilidad estricta y la incomodidad del ambiente pueden comprometer la viabilidad o la calidad de los registros. Frente a estas limitaciones, la espectroscopía funcional del infrarrojo cercano (fNIRS) ofrece una alternativa más adecuada para la investigación del desarrollo. Esta técnica es silenciosa, portable y tolerable para los bebés, y permite obtener medidas de cambios en oxi- y desoxihemoglobina que reflejan la dinámica de la actividad neuronal.

Los datos utilizados en este proyecto provienen de un estudio diseñado para explorar la organización funcional del cerebro en infantes de cuatro meses empleando fNIRS de alta densidad.

## **Descripción del experimento**

El estudio incluyó inicialmente 104 infantes, de 4 meses de edad, sanos y nacidos a término. Los participantes pertenecían a tres contextos lingüísticos:

- 36 bilingües español-vasco
- 30 monolingües español
- 33 monolingües vasco

Las familias completaron un cuestionario que estimaba el porcentaje de exposición del bebé al español y al vasco.

## **Actividad realizada**

No se presentó ninguna tarea: las mediciones fueron en estado de reposo durante el sueño natural.

Los bebés se colocaban en el regazo de sus padres, dentro de un cuarto silencioso y con iluminación mínima. Los padres mantenían quietud para evitar desplazamientos del gorro fNIRS.

Duración de los ensayo: Cada participante fue registrado entre 9 y 25 minutos de actividad espontánea.

Los infantes usaron un gorro fNIRS NIRx NIRScout con:

- 16 emisores
- 24 detectores
- 52 canales por cromóforo (HbO y HbR)
- Longitudes de onda: 760 nm y 850 nm

### Descripción del dataset

Los datos provienen de un estudio longitudinal de conectividad funcional en estado de reposo medido con fNIRS en infantes de **4 meses**. El objetivo del dataset es permitir el análisis de la organización funcional cortical en reposo y comparar posibles efectos de la exposición temprana a ambientes bilingües frente a monolingües.

Formato de los registros

Archivos individuales: archivos MATLAB (\*.mat) preprocesados con sufijo \_preprocessed.mat.

Cada .mat contiene una estructura rsData con múltiples campos (p. ej. clean\_data, conc, filt, wav, GSR\_oxy, etc.).

Metadata: archivo metadata.csv (separador ;) con las columnas: ID (entero): identificador numérico del sujeto, usado para casar con los archivos .mat. Gender (string/categorico): sexo del bebé. Language (numérico o categorico): código de exposición lingüística (mapeado a grupos). % SP (float): porcentaje de exposición al español (columna expresada originalmente con coma decimal).% BQ (float): porcentaje de exposición al euskera/basque (columna expresada originalmente con coma decimal).Age (days) (entero): edad del bebé en días.

### Metodología

1. Definición de rutas e importación de librerías
2. Carga y limpieza del metadata (.csv)--> selección de columnas, limpieza de porcentajes, mapeo de la columna "Language" con "Group" mediante un diccionario mapping = {1: "Bilingual", 2: "Spanish-monolingual", 3: "Basque-monolingual"}.
3. Listar archivos .mat en una variable y extraer el ID del nombre del archivo, se hace un filtrado para juntar a cada sujeto con el ID en el metadata que le corresponde.
4. Se extrae la estructura de los archivos .mat
5. Se extraen las señales HbO mediante una heurística multinivel.
6. Se calculan las matrices de conectividad para cada sujeto mediante una correlación de Pearson entre canales. El pipeline imprime advertencias si falta un archivo para algún ID.
7. Se concatenan todos los valores de correlación individuales, se calculan los percentiles para sugerir un umbral.
8. Inicialmente se seleccionó una muestra de solo 5 bebés por grupo aleatoriamente para sintetizar el proceso de hacer una matriz por sujeto, sin embargo finalmente decidí hacerlo para los 104 sujetos y por separado implementar una función que muestra la matriz de correlación de un sujeto de interés, sin necesidad de mostrar todas en el código.
9. Para cada grupo se apilan las matrices individuales y se calcula una matriz promedio, se guarda en una nueva variable.

10. Umbralización grupal: inicialmente se utilizó el umbral sugerido según el percentil 75, que era 0.57, sin embargo decidí utilizar 0.30 ya que con con percentil 75 las redes quedaban muy desconectadas.
11. Visualización mediante heatmaps de la matriz promedio por grupo, además de una matriz binaria umbralizada.
12. Plots estadísticos: histogramas por sujeto superpuestos, ECDFs por sujeto para comparar distribuciones y series.
13. Grafos 2D con coordenadas definidas según una aproximación basada en el sistema 10-20, tamaño de nodos proporcional al grado, color por comunidad y grosor de las aristas proporcional al peso medio.
14. Medidas topológicas calculadas para cada grupo: clustering, camino más corto, small world, comunidades, modularidad, centralidad (closeness, bewteenness, degree), eficiencias globales y locales, y rich club (no hubo).

### Justificación de las decisiones metodológicas

Métrica de conectividad: Pearson sobre señales HbO. Es una medida simple de sincronía lineal y adecuada para un primer análisis de fNIRS. No capta relaciones en fase/frecuencia (coherencia/PLV), por lo que interpretaciones deben limitarse a sincronía temporal.

Se seleccionó un umbral fijo de  $r = 0.30$  con el objetivo de:

- Garantizar la comparabilidad entre sujetos

- Usar un mismo umbral para todos los participantes permite comparar topologías sin introducir sesgos asociados a umbrales individuales.

- Conservar al menos el 50% de los sujetos debido a la alta variabilidad interindividual en lactantes —en especial por inmadurez cortical, movimientos frecuentes y patrones hemodinámicos menos definidos— se buscó un umbral que preservara un número representativo de conexiones presentes en al menos la mitad de los participantes. Investigaciones de redes funcionales en bebés suelen emplear umbrales entre 0.2 y 0.3 para evitar la desconexión excesiva de la red y maximizar la estabilidad topológica ante datos ruidosos.

Un umbral demasiado bajo produciría redes casi completas, restando valor descriptivo a métricas como grado, clustering o eficiencia.

Manejo de sujetos faltantes: el pipeline alerta y salta archivos faltantes; se implementó realineamiento de `df_meta` con `lista_corr_dfs` para evitar `IndexError`.

También algunas métricas (rich-club, modularidad) requieren redes con suficientes nodos y densidad para resultar estables; al umbralizar fuertemente pueden obtenerse grafos muy fragmentados que complican interpretación, en mi caso, `rich_club` no se pudo analizar, ya que no había suficientes nodos que pasaran el umbral gampoco hubo modularidad.

### Resultados y discusión

Para los tres grupos, se calcularon diversos estadísticos descriptivos derivados de las distribuciones de correlación funcional (rsFC). Estos valores resumen la organización global de la conectividad en reposo antes de aplicar umbrales y construir grafos binarios. Los indicadores incluidos fueron el percentil 25 (th25), el percentil 50 o mediana (th50), el percentil 75 (th75), el promedio general y el índice  $\mu/\sigma$ , que contextualiza la fuerza de conectividad promedio respecto a la dispersión de la distribución.

Los resultados permiten contrastar la estructura general de la conectividad espontánea entre los grupos y sirven como referencia para justificar el umbral de conectividad utilizado

posteriormente ( $t = 0.3$ ), lo cual es especialmente relevante en estudios con fNIRS en infantes, donde la conectividad suele ser baja y altamente variable debido a la inmadurez cortical.

El grupo monolingüe español presenta un patrón de conectividad de magnitud intermedia dentro del conjunto de grupos evaluados.

Los valores de la mediana ( $th50 = 0.3879$ ) y el rango superior ( $th75 = 0.5612$ ) muestran que la mayoría de las conexiones funcionales se concentran en un nivel moderado de correlación positiva. La media ( $0.4034$ ) coincide estrechamente con la mediana, lo cual sugiere una distribución relativamente simétrica, sin sesgos fuertes hacia valores extremos.

El índice  $\mu/\sigma$  ( $0.6498$ ) revela una conectividad más estable que la del grupo bilingüe, aunque menos robusta que la del grupo vasco, lo que concuerda con los heatmaps de conectividad promedio: se observan áreas conectadas, pero con menor densidad y menos consistencia entre sujetos que en el grupo Basque-monolingual.

El mínimo negativo ( $-0.3661$ ) indica que existen algunas conexiones no correlacionadas, aunque no predominan en la distribución global. Esto puede reflejar la variabilidad propia del reposo en infantes y la inmadurez de la coherencia interregional.

En conjunto, el grupo Spanish-monolingual presenta un perfil de conectividad relativamente homogéneo, sin valores especialmente altos, pero suficiente como para que al aplicar un umbral de  $0.3$  aproximadamente la mitad de las conexiones positivas se mantengan.

Los infantes expuestos a dos lenguas muestran una distribución de conectividad más amplia y variable. El mínimo ( $-0.6364$ ) es el más bajo entre los grupos, lo cual señala la mayor dispersión y la presencia de un subconjunto de conexiones con fuerte anticorrelación. De igual manera, el máximo ( $0.7230$ ) sugiere que algunas parejas de canales alcanzan conectividades considerablemente altas.

La mediana ( $0.4159$ ) es levemente mayor que la del grupo Spanish-monolingual, pero la media ( $0.4136$ ) y el percentil  $75$  ( $0.5674$ ) indican que la parte superior de la distribución no es tan elevada cuando se compara con el grupo Basque-monolingual. En otras palabras, aunque el grupo Bilingual presenta conectividades altas en ciertos nodos, la organización general es menos uniforme.

El valor  $\mu/\sigma$  ( $0.6522$ ), similar al del grupo monolingüe español, confirma esta interpretación: la fuerza media de conexión no se separa ampliamente de la variabilidad interna. Esta mayor dispersión se observa también en tus heatmaps, que mostraban una distribución más heterogénea y menos concentrada en patrones fronto-temporales específicos.

Este patrón coincide con reportes previos en infantes expuestos a dos lenguas, donde la variabilidad funcional en reposo puede ser mayor debido a un desarrollo simultáneo de dos sistemas fonológicos y mayor flexibilidad funcional a nivel cortical.

El grupo monolingüe vasco destaca consistentemente como el que presenta las conectividades más altas y estables. El percentil  $75$  ( $0.5563$ ), la media ( $0.4206$ ), y especialmente el valor máximo ( $0.8429$ ) indican que este grupo exhibe conexiones de mayor magnitud, superando a los otros grupos en la región superior de la distribución.

La relación  $\mu/\sigma$  ( $0.6401$ ) es comparable a la de los grupos monolingües español y bilingüe, pero el componente crucial es el desplazamiento global hacia correlaciones más fuertes. El hecho de que incluso el mínimo sea menos negativo ( $-0.0502$ ) evidencia que la distribución de correlaciones en estos infantes está más concentrada y menos afectada por anticorrelaciones extremas.

En cuanto a las métricas:

En primer lugar, los coeficientes de *clustering* mostraron una diferencia clara entre grupos. El grupo bilingüe presentó el valor más elevado ( $0.530$ ), lo que indica una mayor tendencia de sus

nodos a formar triadas y, por tanto, una red más localmente cohesionada. Los dos grupos monolingües presentaron valores ligeramente inferiores: Spanish-monolingual (0.507) y Basque-monolingual (0.490). Esta diferencia sugiere que, en los bilingües, las asociaciones entre nodos tienden a organizarse en configuraciones más densas y mutuamente reforzadas, mientras que, en los monolingües, las conexiones locales se encuentran distribuidas de manera menos compacta.

En cuanto a la eficiencia global, evaluada mediante el *average path length*, se observó un patrón convergente con el clustering. La red bilingüe mostró el menor valor promedio de distancia entre nodos (3.06), reflejando un acceso más eficiente y directo entre distintas partes de la red. La red Spanish-monolingual registró un valor mayor (3.32), lo que indica una estructura menos eficiente desde el punto de vista global. Para la red Basque-monolingual no fue posible calcular esta métrica, ya que el grafo no resultó completamente conectado tras la aplicación del umbral; este hecho es indicativo de una fragmentación estructural que impide la comunicación global entre todos los nodos.

El análisis de la distribución del grado confirmó las diferencias en la estructura organizativa de cada red. En Spanish-monolingual se observaron grados entre 4 y 17, con una variabilidad notable, lo que sugiere que la conectividad se concentra en ciertos nodos muy dominantes. En el grupo bilingüe, los grados oscilaron entre 3 y 14, con una conectividad más homogénea entre nodos y sin la presencia de un “hiper-hub” aislado. El grupo Basque-monolingual mostró la mayor amplitud, llegando a grados de hasta 18, lo que indica la existencia de uno o varios nodos altamente conectados que sostienen buena parte de la arquitectura de la red, mientras otros nodos permanecen más periféricos, contribuyendo así a la mencionada fragmentación.

Las medidas de *betweenness* proporcionaron información adicional sobre el papel de los nodos como puentes estructurales. En la red Spanish-monolingual se observó un nodo extremadamente dominante (nodo 36, 0.265), cuya centralidad es muy superior al resto, reflejando una dependencia estructural marcada: gran parte del flujo de conexión pasa por este punto crítico. El grupo bilingüe, en contraste, distribuyó sus nodos puente de manera más equitativa. Aunque contó con nodos con valores elevados (como los nodos 13, 19 y 17, entre 0.138 y 0.170), ninguno alcanzó la dominancia extrema observada en el grupo español, lo que indica una red más resiliente y menos vulnerable ante la pérdida de un nodo individual. En Basque-monolingual también se identificaron nodos centrales (especialmente el nodo 10, con 0.192), pero nuevamente con menor extremo que en el grupo Spanish-monolingual.

El análisis de *closeness* confirmó estos patrones. En cada grupo apareció un nodo especialmente eficiente para acceder al resto de la red (nodo 36 en Spanish-monolingual, nodo 17 en Bilingual y nodo 10 en Basque-monolingual). Sin embargo, en el grupo bilingüe la eficiencia nodal se encontró más repartida entre varios nodos, con múltiples valores superiores a 0.40. Esto contrasta con lo observado en los monolingües, donde la eficiencia se concentró en un número más limitado de nodos, consistente con una estructura más jerárquica o dependiente de hubs específicos.

## **Conclusión**

En conjunto, los resultados muestran que el grupo bilingüe presenta una organización de la red más cohesiva, eficiente y distribuida, tanto a nivel local como global. Su estructura evita la dependencia excesiva de nodos hipercentrales y favorece una comunicación rápida y flexible entre regiones de la red. En contraste, la red Spanish-monolingual se caracteriza por una mayor centralización y por su dependencia de nodos críticos, lo que la convierte en una estructura menos eficiente y potencialmente más vulnerable. Finalmente, la red Basque-monolingual muestra signos de fragmentación, con la presencia de nodos altamente conectados pero también

con sectores desconectados, lo que impide el cálculo de la eficiencia global y sugiere un desarrollo o una organización menos integrada en comparación con los otros grupos. Estos hallazgos adquieren sentido en el marco del desarrollo temprano, redes funcionales en bebés de 4 meses son altamente sensibles a la variabilidad del entorno lingüístico, y la exposición bilingüe temprana puede introducir una mayor diversidad en la organización funcional, mientras que entornos monolingües consistentes tienden a generar patrones más uniformes.

## Referencias

Del dataset utilizado: Blanco, B., Molnar, M., Carreiras, M., & Caballero-Gaudes, C. (2022). *Open access dataset of task-free hemodynamic activity in 4-month-old infants during sleep using fNIRS*. *Scientific Data*, 9, 102. <https://doi.org/10.1038/s41597-022-01216-0>

Homae, F. (2014). *A developmental perspective on functional connectivity in the human infant brain*. *Neuroscience Research*, 86, 20–27. <https://doi.org/10.1016/j.neures.2014.06.006>

White, B. R., Snyder, A. Z., Cohen, A. L., Petersen, S. E., Raichle, M. E., Schlaggar, B. L., & Culver, J. P. (2009). *Resting-state functional connectivity in the human newborn brain*. *NeuroImage*, 48(2), 350–358. <https://doi.org/10.1016/j.neuroimage.2009.06.029>

Taga, G., Watanabe, H., Homae, F., & Nakano, T. (2018). *Development of functional cortical networks in early infancy: A fNIRS study*. *Frontiers in Human Neuroscience*, 12, 400. <https://doi.org/10.3389/fnhum.2018.00400>