

Mentoría Data Science aplicado a BCI

Consignas TP2: Continuación Análisis y Visualización y Curación de Datos

INTEGRANTES:

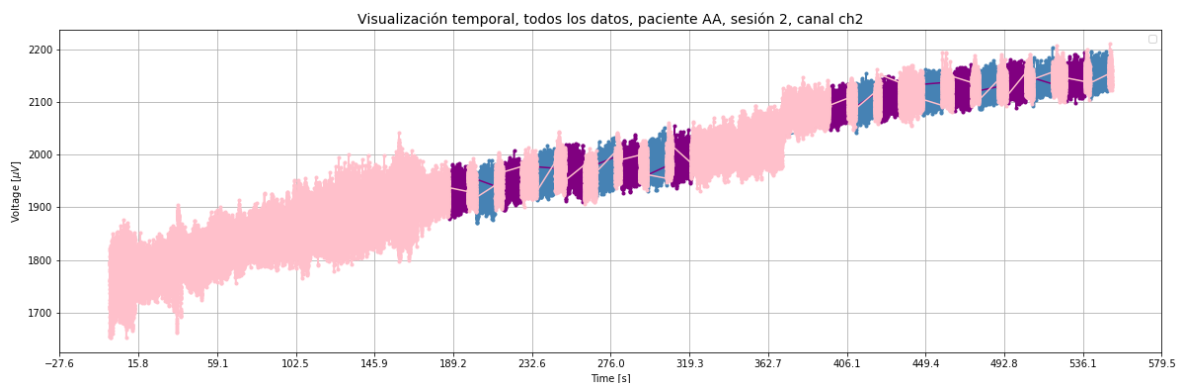
- Antonello Adriano
- Bergamasco Paula

MENTORES:

- Juan López
- Joaquín Seia

DOMINIO DEL TIEMPO

En el trabajo práctico número 1 se evaluó el 'dominio del tiempo' es decir, evaluar la evolución de la señal de voltaje con respecto al tiempo, en donde se llegó a la conclusión de que se observa que no hay una diferencia significativa en el valor medio del voltaje entre los estados, es decir que el análisis en el dominio del tiempo, no parece contener información que permita una diferenciación entre los estados, es decir que no se obtuvieron resultados concluyentes. Más bien se presentan muchas complejidades y particularidades de cada sujeto/registro, que vuelven el análisis poco generalizable, por lo que se continuó el análisis a través del 'dominio de la frecuencia'.



DOMINIO DE LA FRECUENCIA

En el 'dominio de la frecuencia', se evalúa la señal, según la frecuencia en la que oscilan dentro de un rango determinado.

- **A) Nivel Segmento/Estado:**

Para efectuar este análisis se escogió al paciente AA, sesión 2, canal ch2.

a) La frecuencia de Nyquist es la frecuencia más alta que posiblemente podamos esperar observar en los datos.

La fórmula para la frecuencia de Nyquist es:

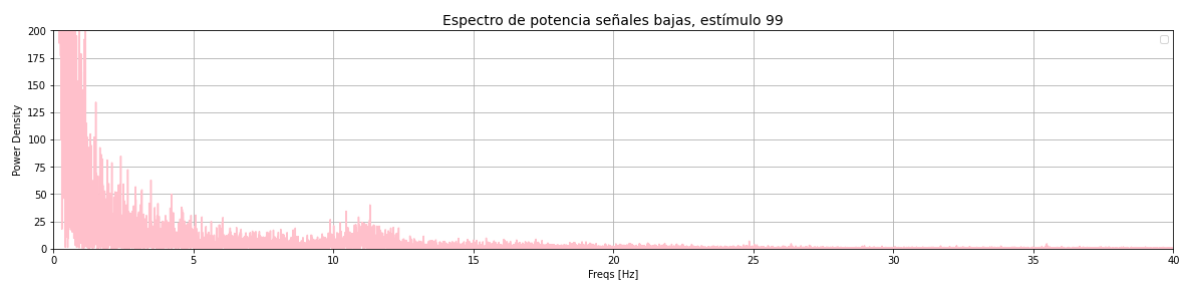
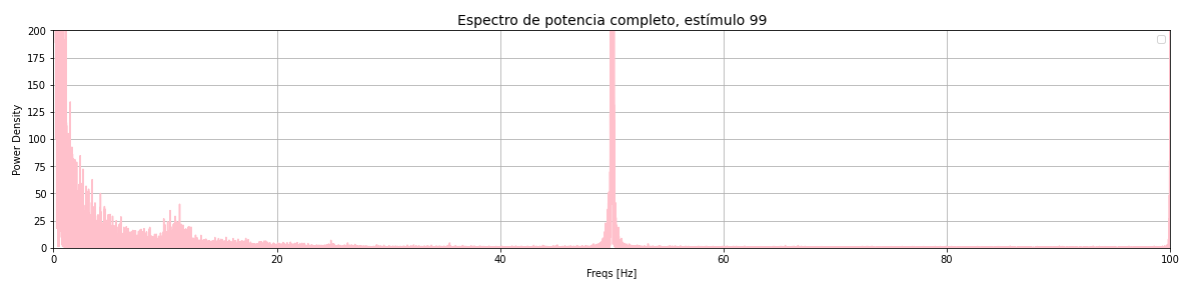
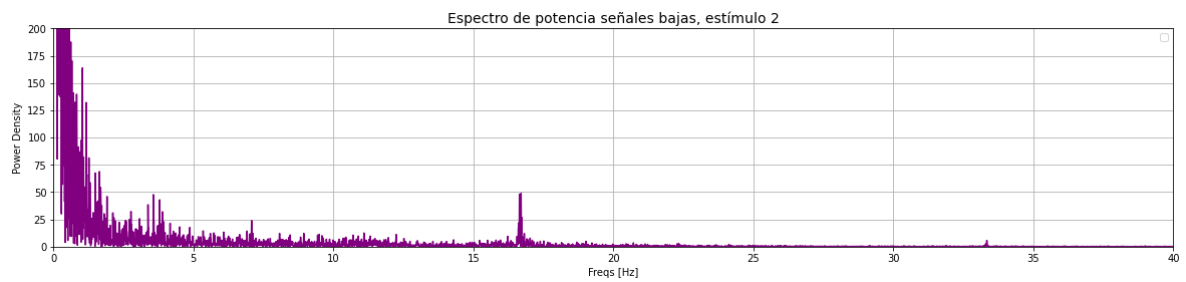
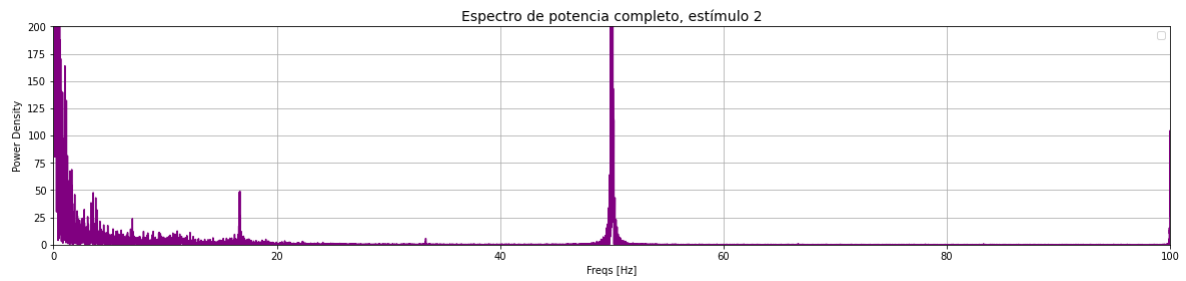
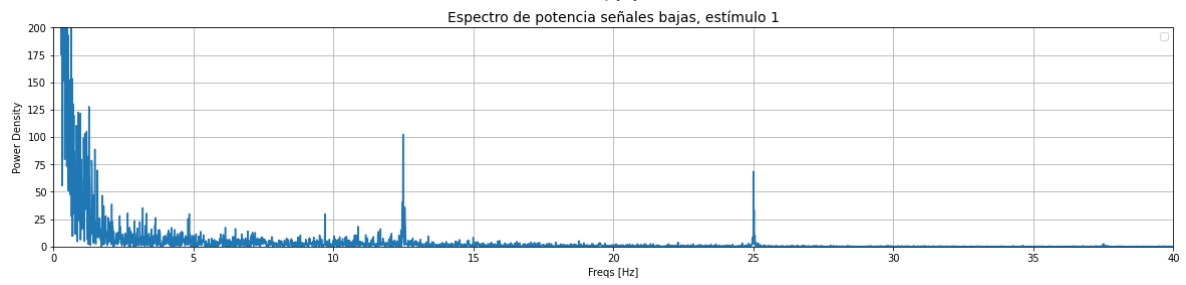
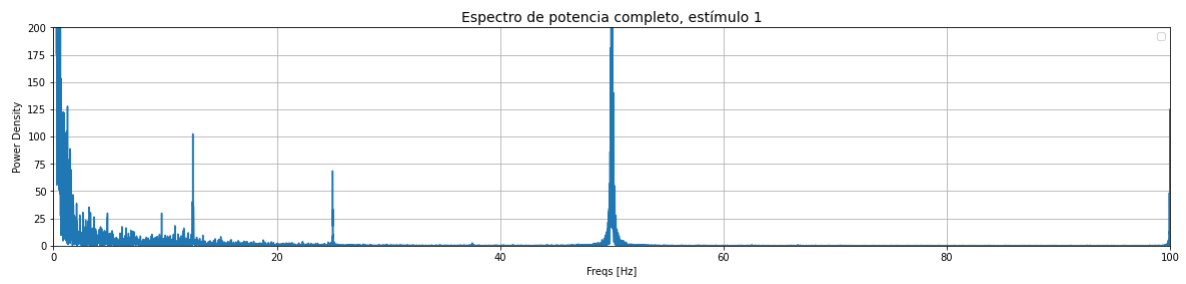
- $f_{NQ} = 1/2 \cdot dt \text{ ó } f_0/2$

Donde f_0 es la frecuencia de muestreo.

En nuestros datos la frecuencia máxima que pudo ser registrada es igual a 100Hz, ya que:

- $f_{NQ} = 1 / (2 \cdot 0.005) = 100\text{Hz}.$

b) La Densidad Espectral de Potencia de una señal es una función matemática que da a conocer la distribución de la potencia de dicha señal sobre las distintas frecuencias en donde está formada. Así, se puede establecer el rango de frecuencias donde se concentran las variaciones de potencia. Se calcula la densidad espectral de potencia, o espectro, para poder evaluar la actividad rítmica en los datos del EEG. Esta densidad espectral de potencia es calculada utilizando la transformada de Fourier, ya que el espectro de los datos es la magnitud al cuadrado de la transformada de Fourier. El espectro indica la amplitud de la actividad rítmica (potencia) en función de la frecuencia (Hz).



Se puede observar en los gráficos representados que hay una actividad rítmica a la frecuencia de **50Hz**, para todos los casos (estímulo 1, estímulo 2 y estímulo 99), que es generado por la interferencia electromagnética, y que exhibe la mayor potencia. Pero cuando se analiza en más detalle las frecuencias más bajas, donde se encuentran nuestras frecuencias de interés, se puede observar que existe actividad rítmica a la frecuencia de **12,5Hz** en el caso del estímulo 1 y a la frecuencia de **16,5Hz** en el caso del estímulo dos, que predomina sobre el resto de las frecuencias, y para ambos casos también se llegan a observar los primeros armónicos de ambas frecuencias, **25 Hz** y **33 Hz**, respectivamente, aunque con menor potencia. Para el caso en el que el paciente no está recibiendo ningún estímulo, cuando se analiza en más detalle las frecuencias más bajas, no se observa ninguna frecuencia predominante, estando ausentes los picos observados en el caso de los otros estados.

Una técnica para enfatizar los ritmos de menor amplitud ocultos por oscilaciones de gran amplitud es cambiar la escala del espectro a **decibelios**, el decibel es una escala logarítmica. En estos casos el ritmo de 50 Hz sigue siendo predominante al igual que en los casos previos, y se amplían las magnitudes de señales bajas, pero agregan ruido a las señales de interés, por lo que el gráfico de potencia en decibelios no brinda información relevante al análisis, sino que lo obstaculiza, ya que atenúa las señales de interés, y resalta las señales irrelevantes.

c) En las tablas siguientes se representan las 10 frecuencias con mayor potencia, en el rango de los 10-40Hz.

ESTADO 1

POTENCIA	FRECUENCIA (Hz)
102.486055	12.49
68.529200	25.01
40.331470	12.46
35.833815	12.55
33.391327	25.03
31.318758	12.47
21.990421	12.56
18.273238	10.88
17.806832	25.02
17.610610	12.57

ESTADO 2

POTENCIA	FRECUENCIA (Hz)
48.921245	16.70
48.221734	16.69
48.018051	16.65
26.822886	16.72
23.703989	16.67
21.978595	16.63
17.105962	16.61
12.585201	16.71
12.559075	11.09
12.077687	16.82

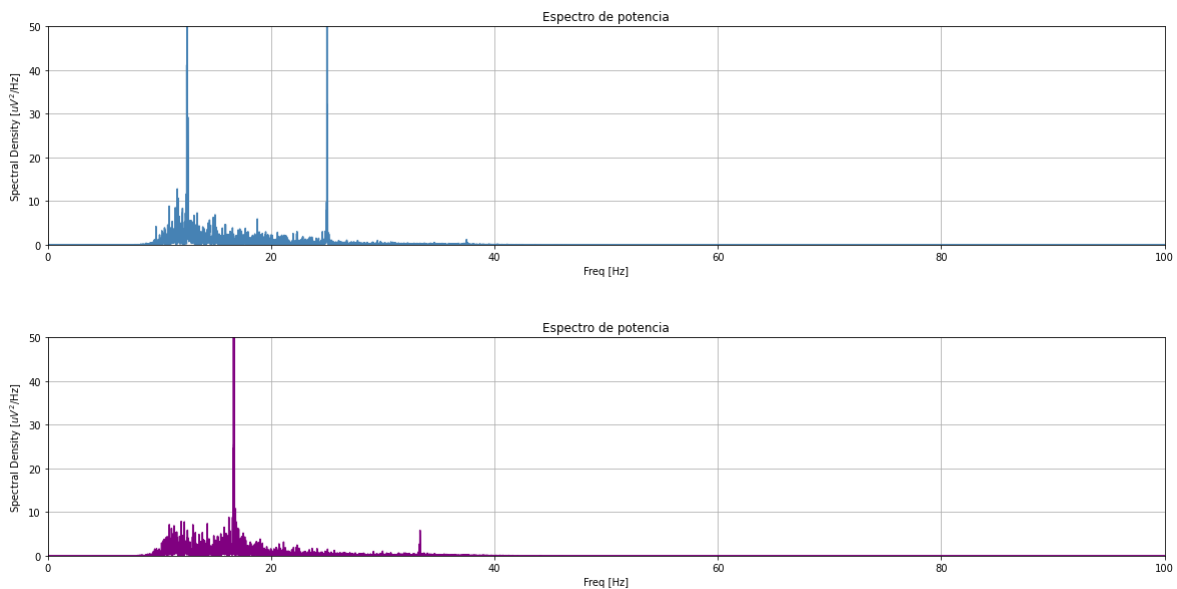
ESTADO 99:

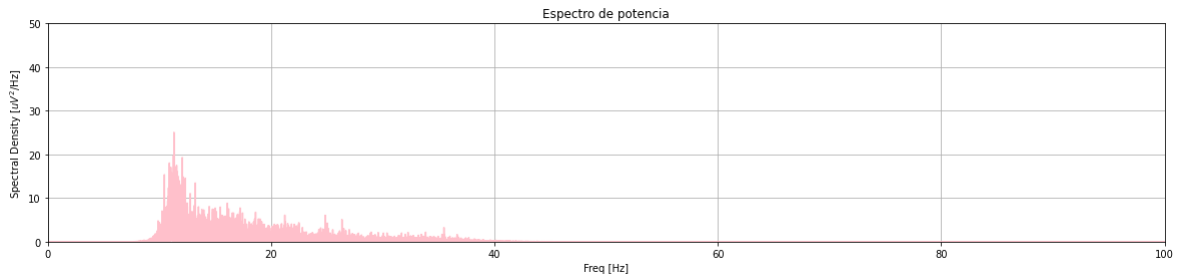
POTENCIA	FRECUENCIA (Hz)
40.085558	11.31
28.951460	10.45
26.451324	10.90
24.543120	11.30
24.323121	10.45
24.272156	10.90
23.170345	11.02
22.177012	11.27
22.091449	10.24
22.066018	11.10

Se pudo observar a través de los datos del espectro de frecuencias que existía una frecuencia con la mayor amplitud que correspondía a las cercanías al pico de 50 Hz, que se genera, como se mencionó previamente, por las interferencias electromagnéticas, y además se podía observar una frecuencia con gran amplitud correspondiente a la componente lineal a lo largo de la línea temporal que contienen los datos en las cercanías de 0Hz. Ignorando estas señales (señales inferiores a 10Hz, y superiores a 50Hz) se puede concluir que para el estado 1 las frecuencias con mayor amplitud son las frecuencias cercanas a 12,5 Hz, siguiéndole en magnitud la correspondiente a la armónica de la señal a 12,5 Hz, es decir 25 Hz; para el estado 2, se puede observar que las frecuencias con mayor amplitud

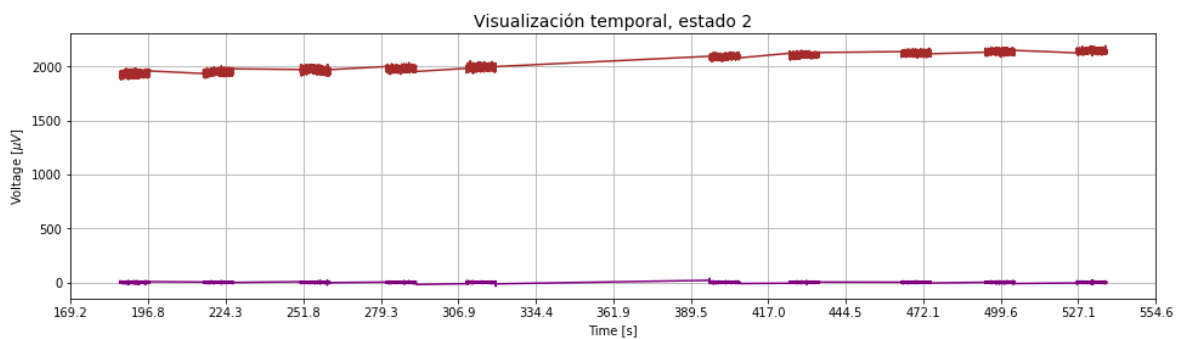
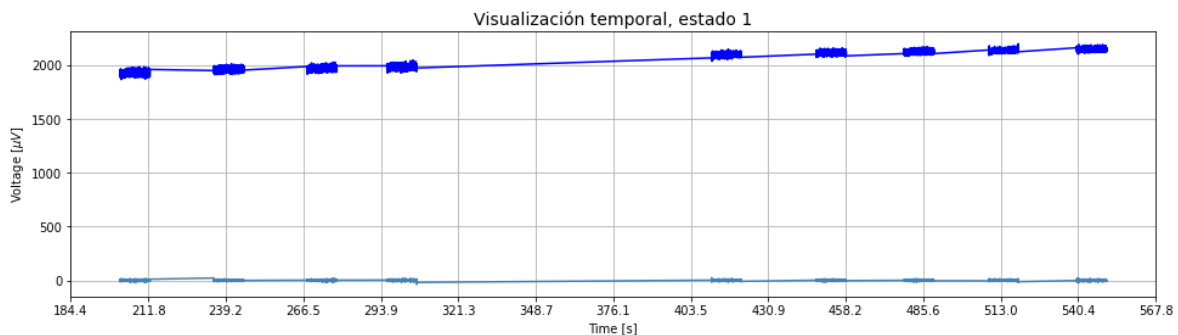
corresponden a las frecuencias cercanas a 16,5 Hz, y si bien el armónico correspondiente a esta señal aparece, este lo hace con una magnitud muy baja, por lo que no figura dentro de las frecuencias con mayor presencia. Por otro lado, se puede observar que para el estado 99, se observan picos con gran amplitud en las cercanías a 11 Hz, lo cual se debe a ruido en la señal. Por lo tanto, se puede concluir que a través del estudio de las frecuencias se encuentra una diferenciación clara entre los distintos estados.

d) Para filtrar las señales obtenidas se utilizó el criterio de descartar aquellas señales que se consideraron como interferentes, es decir que, en primer lugar se eliminó el pico de 50Hz, correspondiente a las señales provenientes de la interferencia electromagnética, y en segundo lugar otros picos interferentes o ruido en la señal, y que se corresponden con la componente lineal observada en el estudio del dominio del tiempo, lo que involucra frecuencias inferiores a 10Hz, estos son los valores de señales que se deben descartar, ya que no son relevantes para el análisis y corresponden a señales interferentes. Los espectros de potencias representados a continuación se corresponden con los espectros de potencias de las señales filtradas correspondientes a cada estímulo, en primer lugar el estímulo 1, en segundo lugar el estímulo 2, y por último el estímulo 99.





e) Al comparar en el dominio del tiempo las señales con y sin filtrado de frecuencias indeseadas se llega a la conclusión de que al filtrar por frecuencias se logra eliminar esa componente lineal que se visualizaba en la serie temporal, y se logra eliminar el ruido de la señal.



- **B) Nivel Paciente - un canal:**

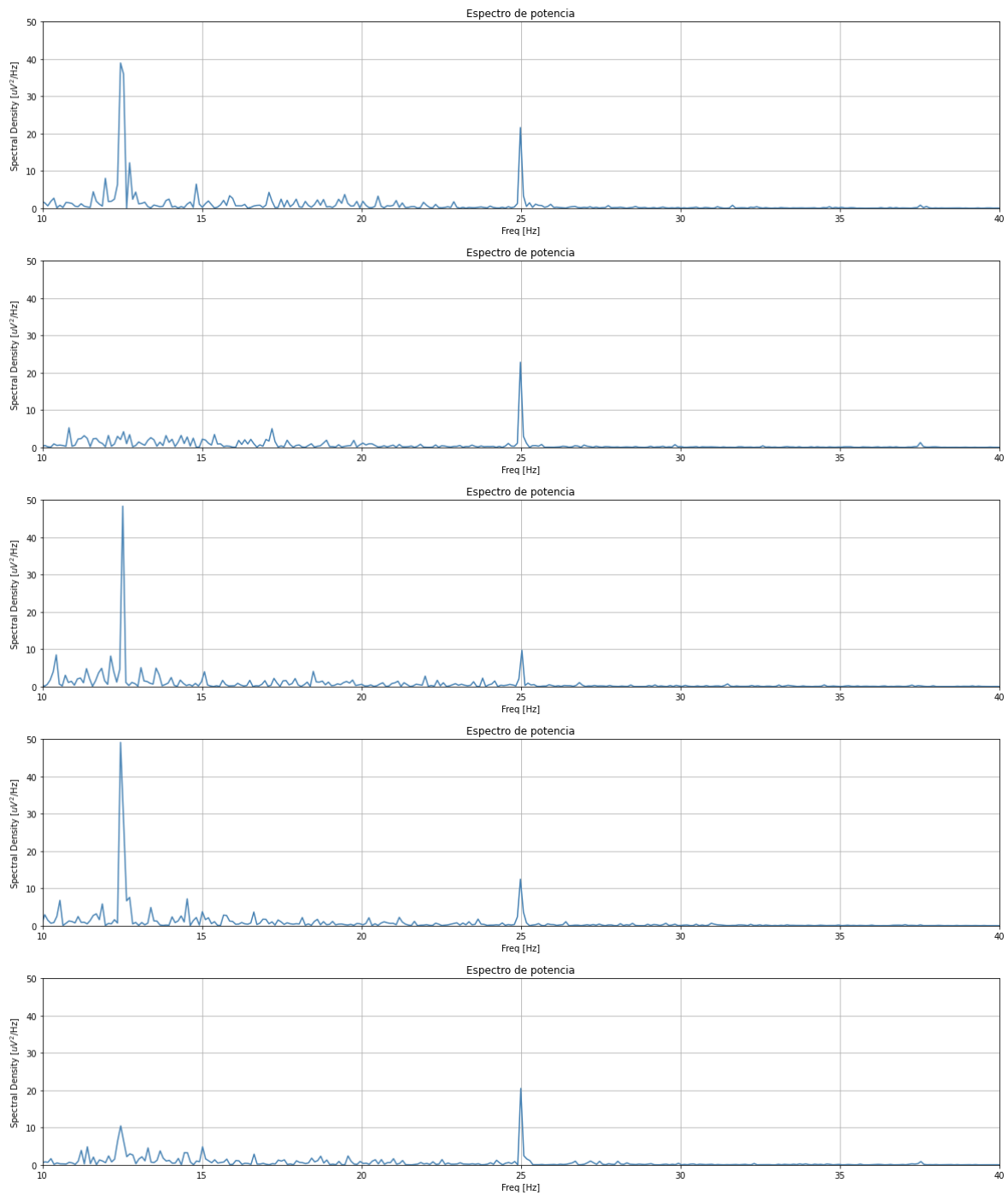
Para efectuar este análisis se escogió al paciente AA, sesión 2, canal ch2.

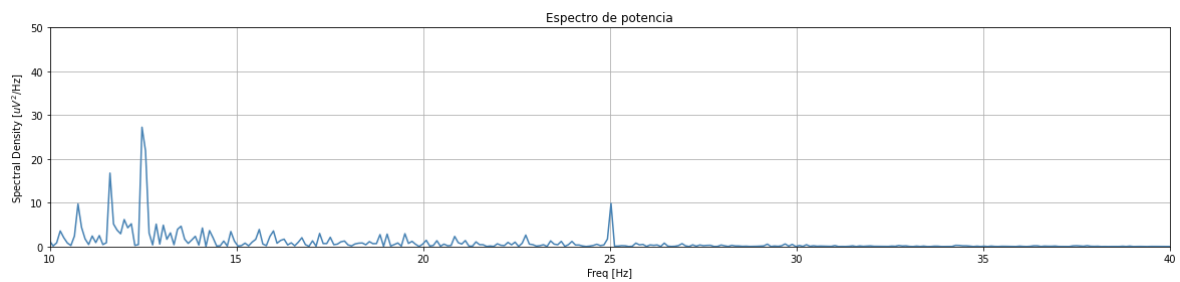
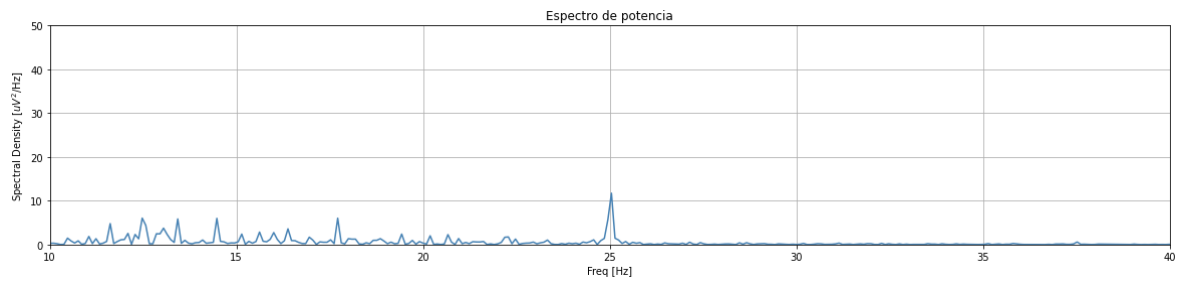
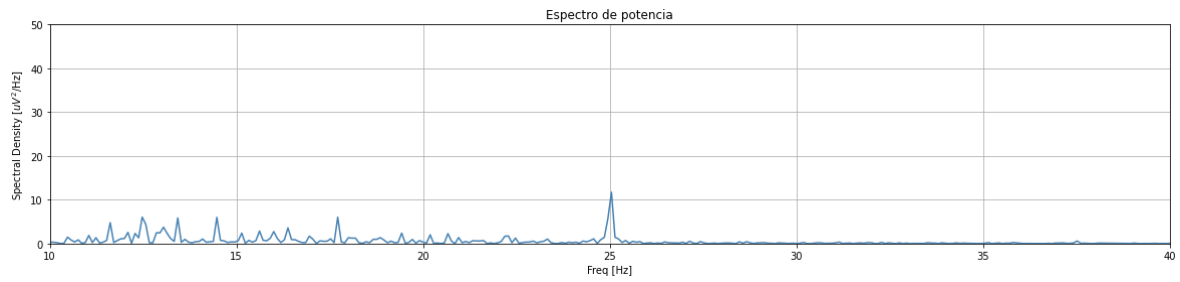
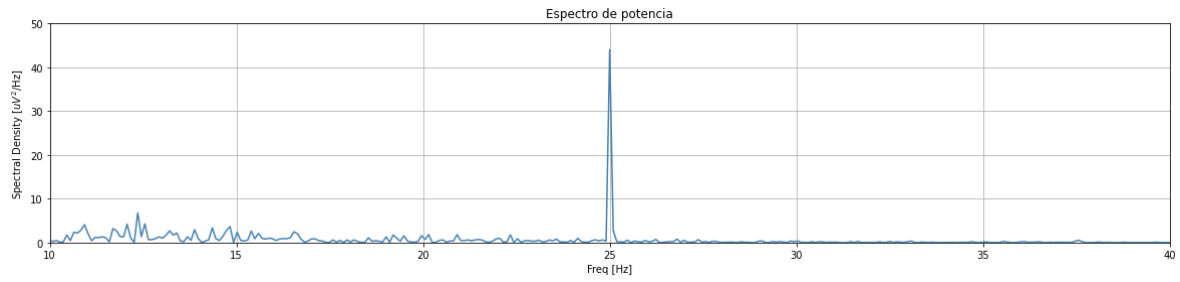
a) Para cada estado, correspondientes al canal 2, del paciente AA, sesión 2, existen los siguientes intervalos:

- **Estado 1:** 9 intervalos.
- **Estado 2:** 10 intervalos.

- **Estado 99:** 20 intervalos.

ESTADO 1:

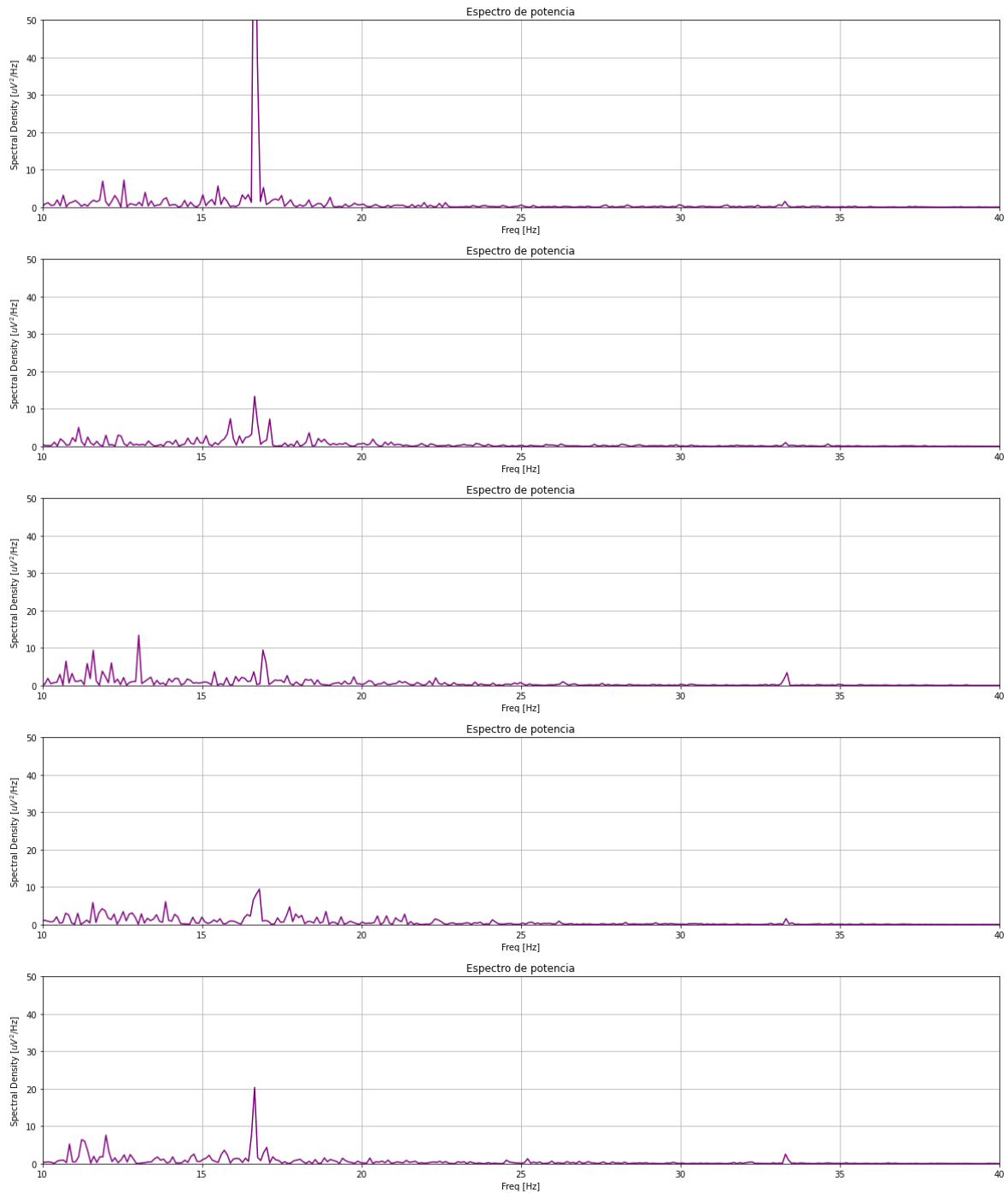


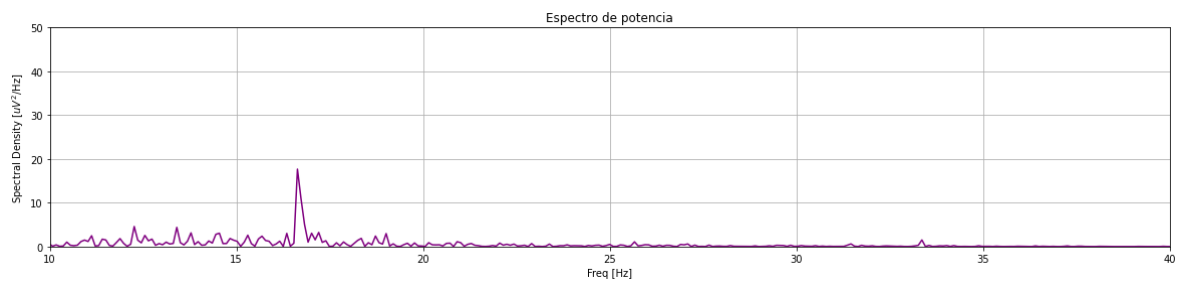
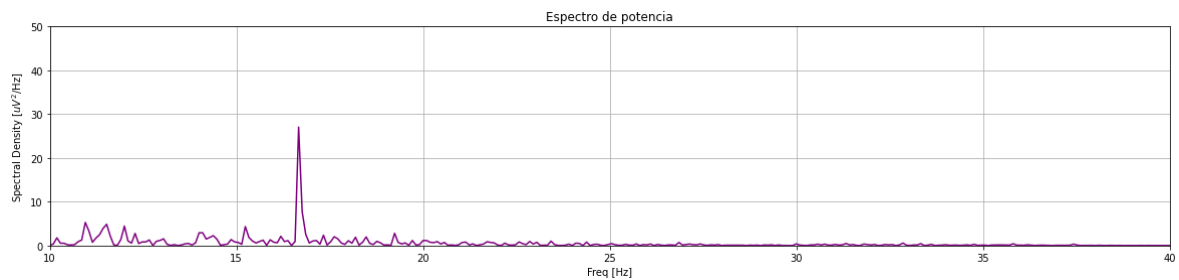


INTERVALO	POTENCIA	FRECUANCIA MAX(Hz)
1	38.938411	12.45
2	22.819252	24.99
3	48.37122	12.52
4	49.156842	12.45
5	20.479103	25.0
6	44.028361	25.0
7	11.720254	25.05
8	11.720254	25.05
9	27.244617	12.47

Estado1: Para cada uno de los intervalos correspondientes al estado 1, la frecuencia máxima oscila alrededor de 12.5Hz (12.45-12.52), y 25.0Hz (29.99 – 25.05) correspondiente al primer armónico de la frecuencia de 12.5Hz.

ESTADO 2:





INTERVALO	POTENCIA	FRECUENCIA MAX (Hz)
1	103.618846	16.64
2	13.326250	16.65
3	13.386154	13.02
4	9.430882	16.80
5	20.345153	16.65
6	10.674092	16.71
7	27.025733	16.74
8	17.668211	16.67
9	25.150473	16.63
10	18.093060	16.67

Estado2: Para cada uno de los intervalos correspondientes al estado 2 la frecuencia máxima oscila alrededor de 16.67Hz (16.63-16.80Hz), exceptuando un intervalo que contiene una frecuencia máxima de 13hz.

ESTADO99:

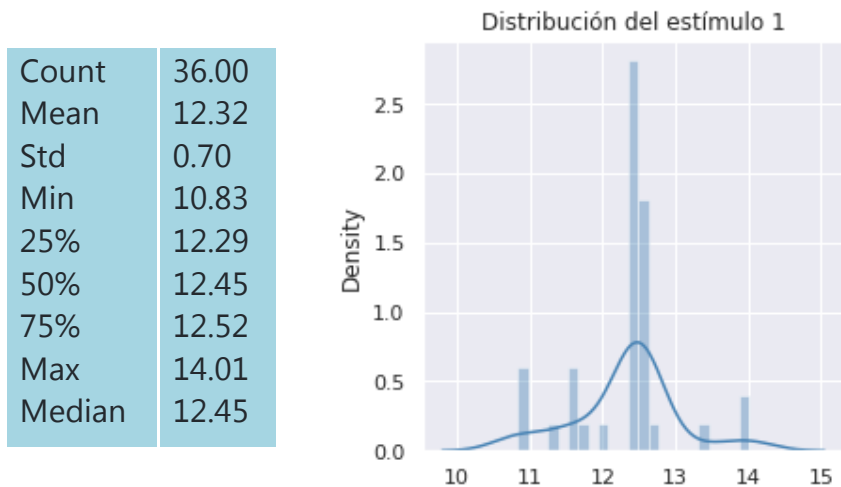
INTERVALO	POTENCIA	FRECUENCIA MAX (Hz)
1	20.705722	11.44
2	5.409377	16.70
3	4.774217	11.81
4	5.057203	15.74
5	6.155306	12.89
6	5.511417	16.46
7	5.946500	10.96
8	3.662531	15.02
9	5.579150	11.95
10	31.757443	10.90
11	4.529304	16.14
12	5.840695	12.93
13	10.571699	14.88
14	9.521564	12.52
15	3.497943	18.29
16	7.631103	12.30
17	8.512773	11.74
18	7.335316	16.56
19	5.637761	11.68
20	8.905695	11.0

Estado99: Para cada uno de los intervalos correspondientes al estado 99 la frecuencia máxima oscila alrededor de un rango variable (10.96-18.29Hz), es decir involucrando las frecuencias de interés, pero todas ellas a una potencia menor que las correspondientes a las de los estados anteriores, por lo que se puede distinguir perfectamente de dichos estados.

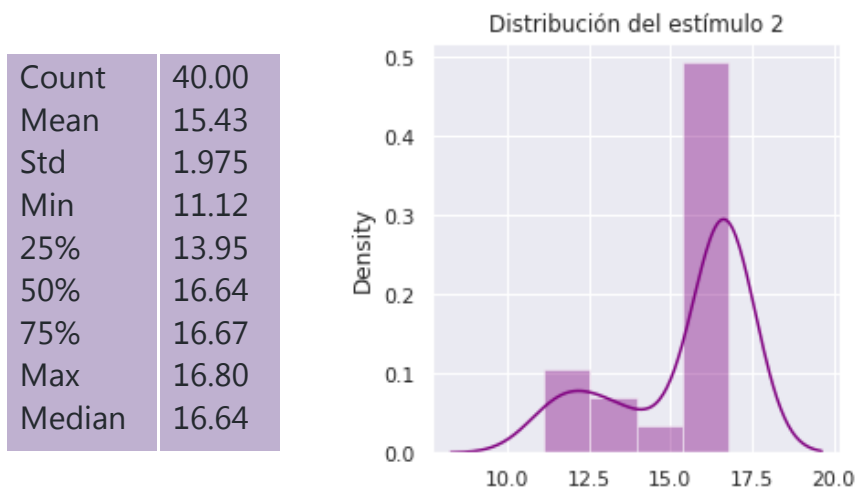
b) Resumen estadístico:

Para este análisis se consideró las frecuencias con mayor presencia de cada intervalo de todos los canales, para poder tener la cantidad necesaria de datos para realizar el análisis estadístico.

- Resumen estadístico estado 1:

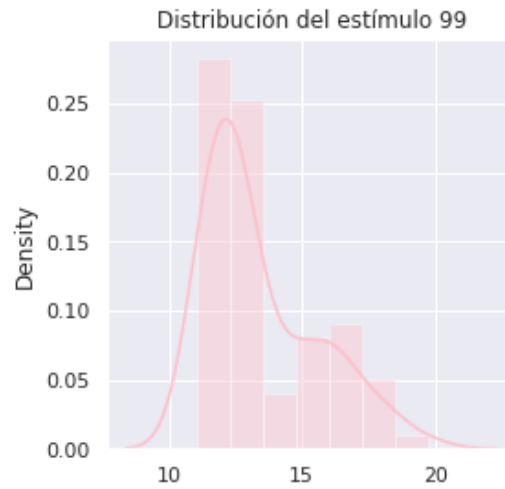


- Resumen estadístico estado 2:

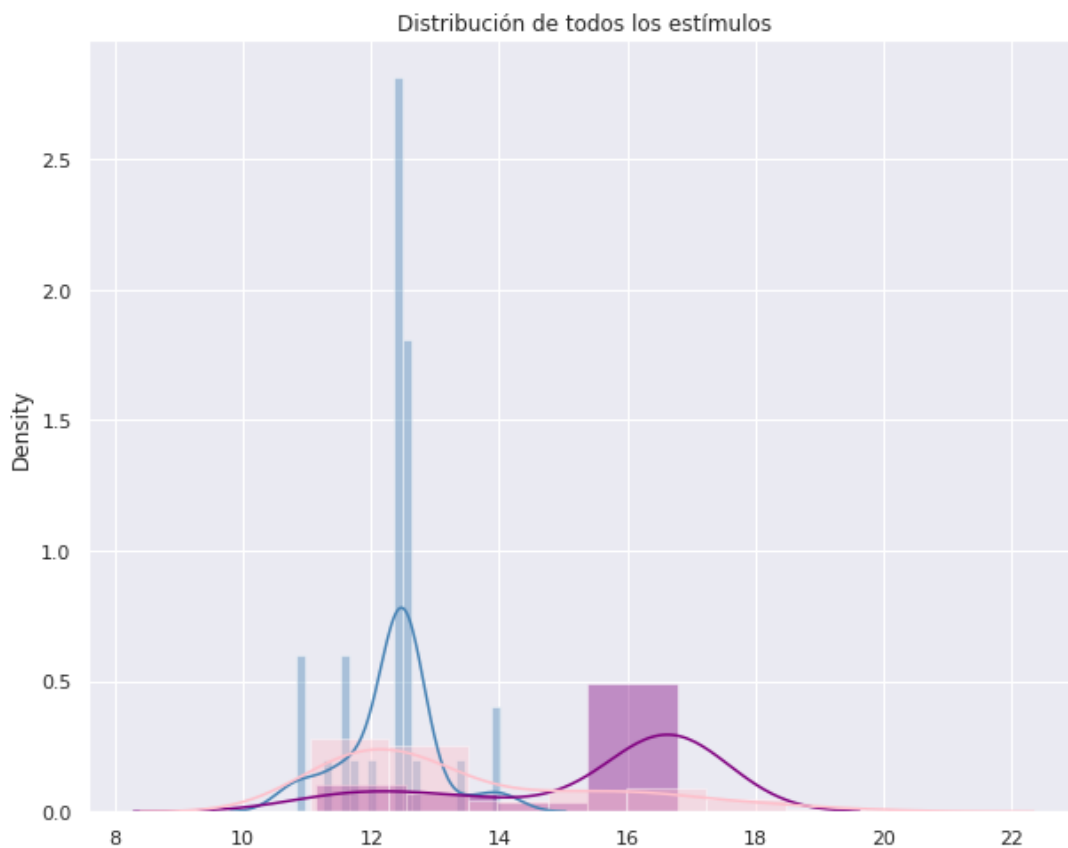


- Resumen estadístico estado 99:

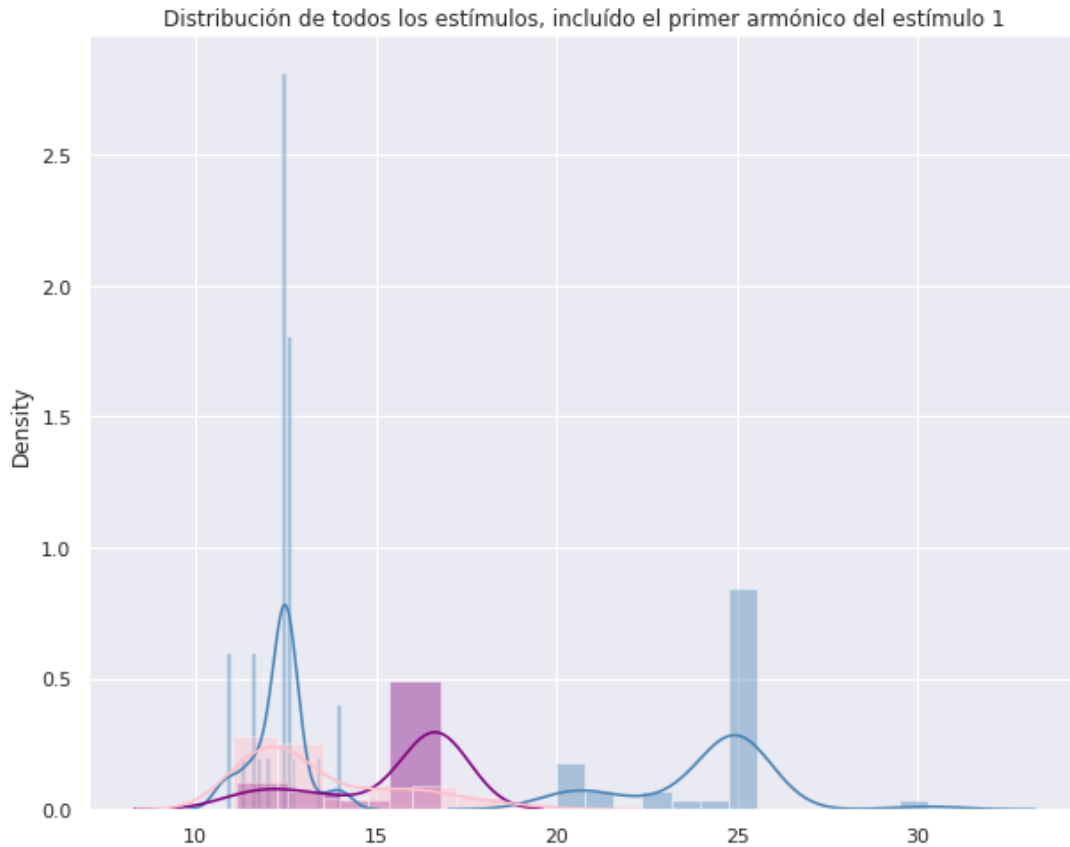
Count	80.00
Mean	13.44
Std	2.12
Min	11.04
25%	11.79
50%	12.67
75%	14.88
Max	19.71
Median	12.67



- Distribuciones conjuntas de todos los estímulos:



- Distribuciones conjuntas de todos los estímulos, incluido el primer armónico del estímulo 1:



d) Diferencia estadística:

Test de diferencia de medias entre el estímulo 1 y el estímulo 2 (stats.ttest_ind(i, n)):

- statistic=-8.950245499569903
- pvalue=2.065183958283648e-13

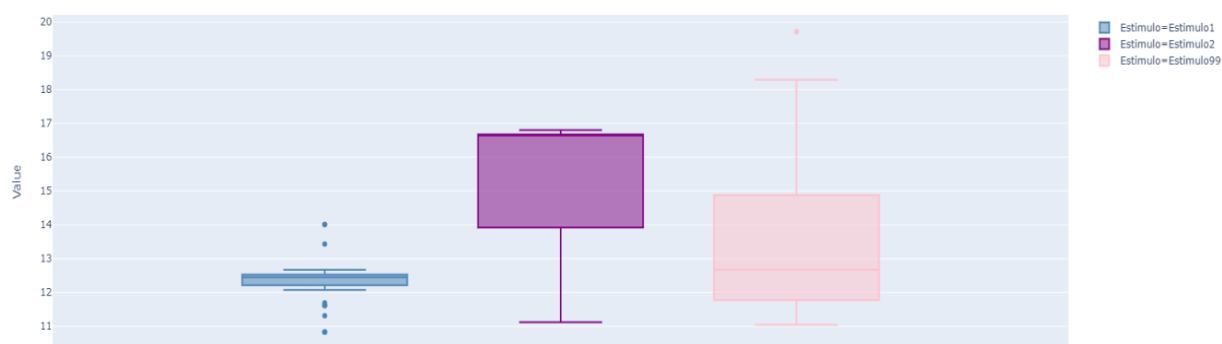
Test de diferencia de medias entre el estímulo 99 y el estímulo 2 (stats.ttest_ind(x, n)):

- statistic=-4.9689898576519465
- pvalue=2.300443298041385e-06)

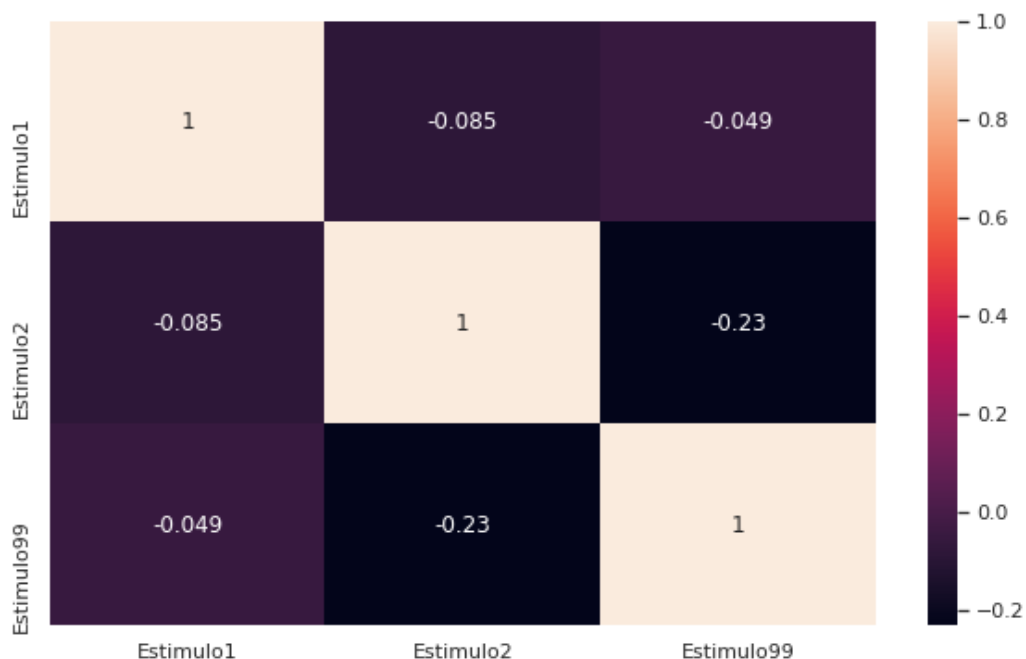
Test de diferencia de medias entre el estímulo 99 y el estímulo 1 (stats.ttest_ind(x, i)):

- statistic=-3.0784896208249504
- pvalue=0.002606694345127086

Al realizar el test de diferencia de medias entre los tres estados, se puede determinar que como el p-valor es inferior a 0.005 en todos los casos, hay suficiente evidencia para decir que existe una diferencia estadísticamente significativa entre los valores centrales de frecuencias para los estados existentes.



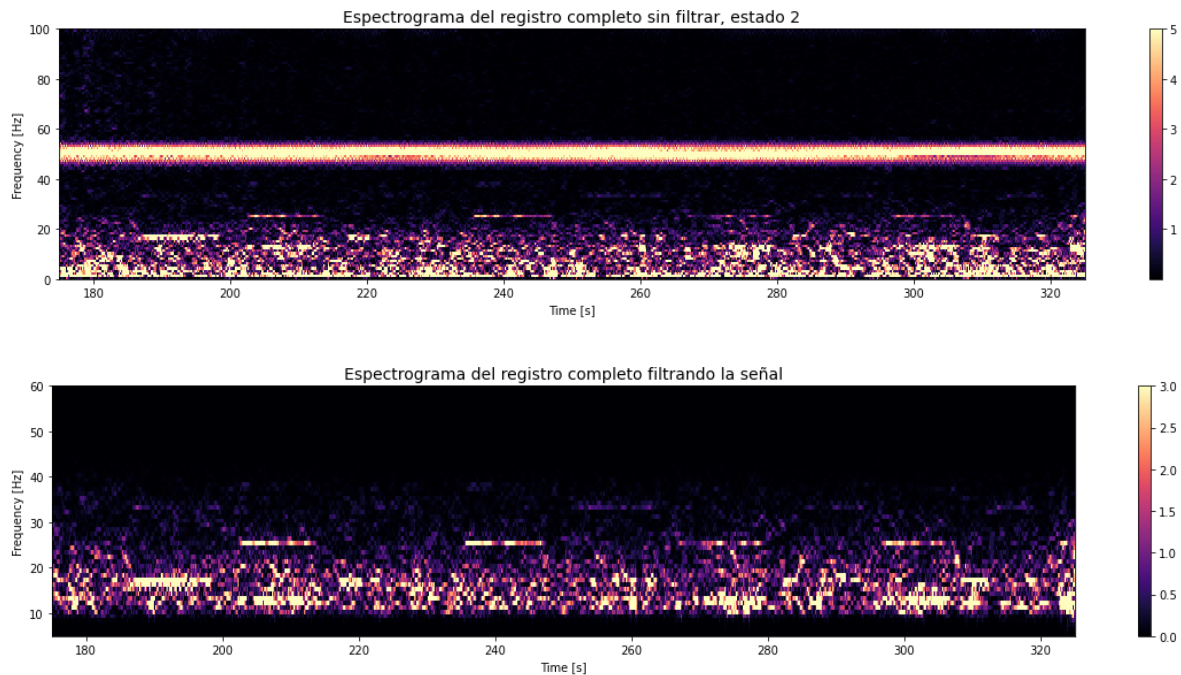
d) Independencia de variables:



Se puede determinar que las variables, estado registrado de la señal y la frecuencia mayoritaria presente en el registro son variables independientes, ya que no existe una correlación entre ellas.

e) Espectrograma: El espectrograma es el resultado de calcular el espectro de una señal por ventanas de tiempo de la misma. Resulta una gráfica tridimensional que representa la energía del contenido frecuencial de la señal según va variando está a

lo largo del tiempo. La idea del espectrograma es dividir la serie de tiempo en intervalos de datos más pequeños y luego calcular el espectro en cada intervalo. Estos intervalos pueden ser bastante pequeños e incluso superponerse. El resultado es el espectro en función de la frecuencia y el tiempo.



Si comparamos los espectrogramas de la señal filtrada con respecto a la señal sin filtrar se puede observar que en el primero predomina la intensidad de la frecuencia a 50Hz, la cual desaparece al filtrar la señal, al igual que la señal ruidosa entre 0 y 10 Hz. Además, en el espectrograma con la señal filtrada se pueden distinguir las frecuencias en las que predomina cada estado y sus armónicos.

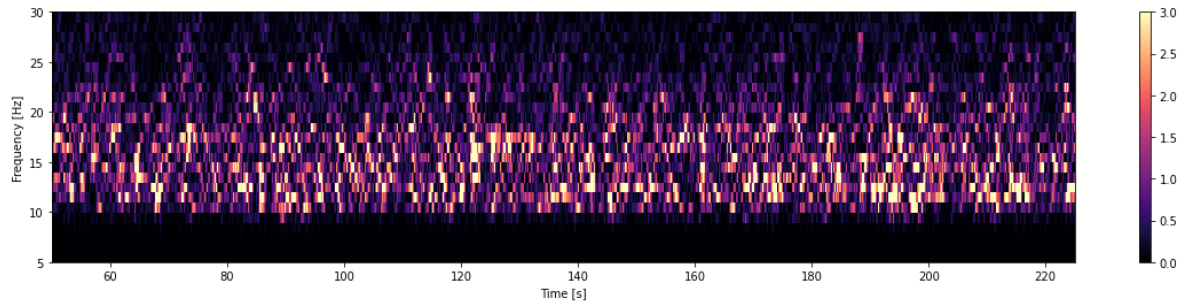
f) Conclusiones:

- En el dominio de la frecuencia existe una diferenciación clara entre los estados, en base a las frecuencias de mayor presencia observadas en cada estado de estimulación.
- Cuando se eliminan las señales interferentes en el dominio de la frecuencia se elimina la componente lineal que se manifiesta en el dominio del tiempo.
- Si bien entre las medias de cada estado de estimulación existe una diferencia estadísticamente significativa, el estado 99, posee una media muy cercana con respecto al estado 1, pero con menor amplitud, por lo que se pueden diferenciar; de todos modos, se puede tener en cuenta el primer armónico, a 25Hz, en donde la señal es más limpia, ya que en el estímulo 99 no presenta frecuencias en las cercanías a ese valor.

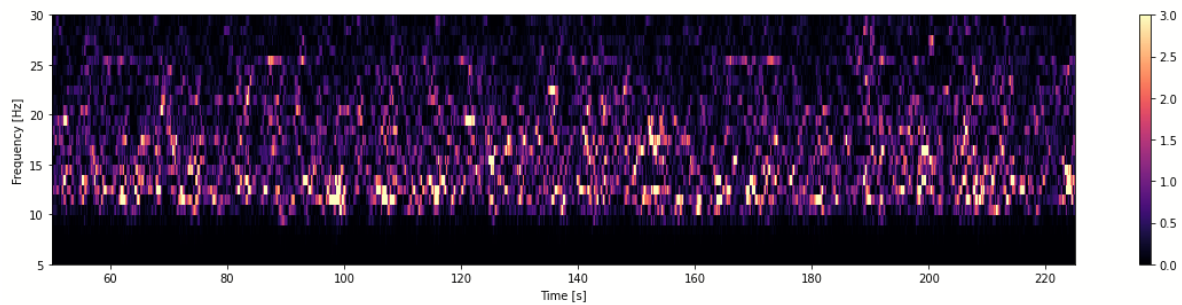
- **C) Nivel Paciente - multi canal:**

Para este análisis se consideró al paciente AA, incluyendo todas sus sesiones.

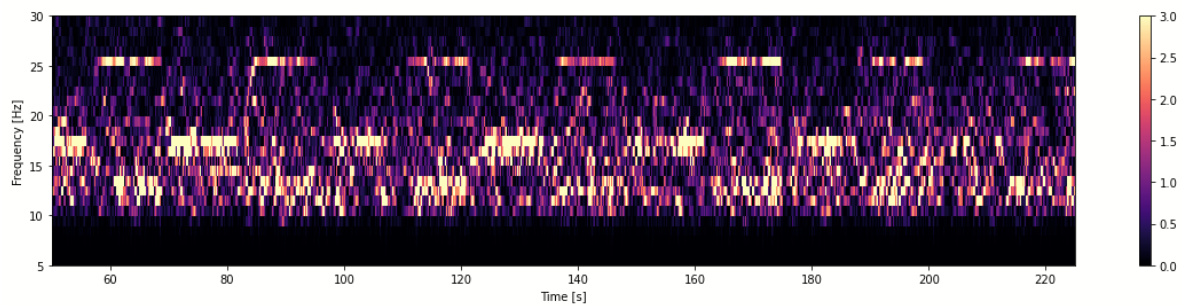
- Espectrograma canal CH0:



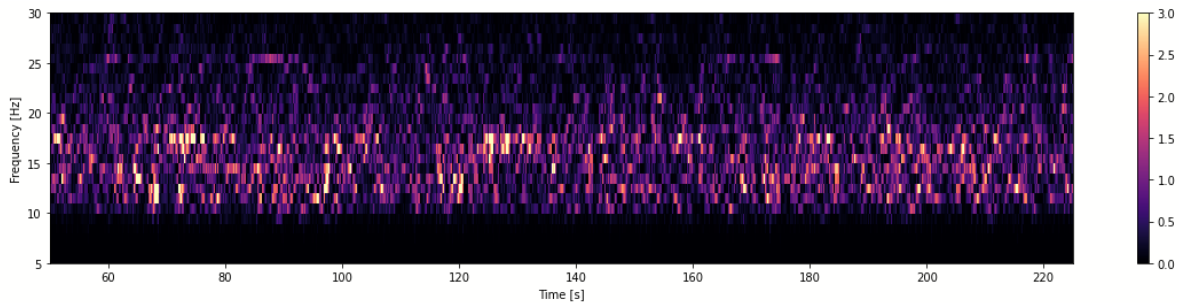
- Espectrograma canal CH1:



- Espectrograma canal CH2:



- Espectrograma canal CH3:



En los gráficos anteriores se puede observar que no existe una clara correlación de la señal respecto a los distintos canales, ya que sus espectrogramas dan imágenes diferentes, algunos con espectrogramas más confusos y otros con espectrogramas más concluyentes, siendo el espectrograma del canal 2 (ch2), el que proporciona una señal más clara y definida, diferenciándose claramente las frecuencias predominantes entre los diferentes estados y sus armónicos. Los espectrogramas de los otros canales proporcionan imágenes de señales más confusas, por lo que se podría seleccionar solamente el canal 2, ya que es el que proporciona los datos más relevantes, y algún otro canal para tener de complemento, por ejemplo el canal 1 (ch1).

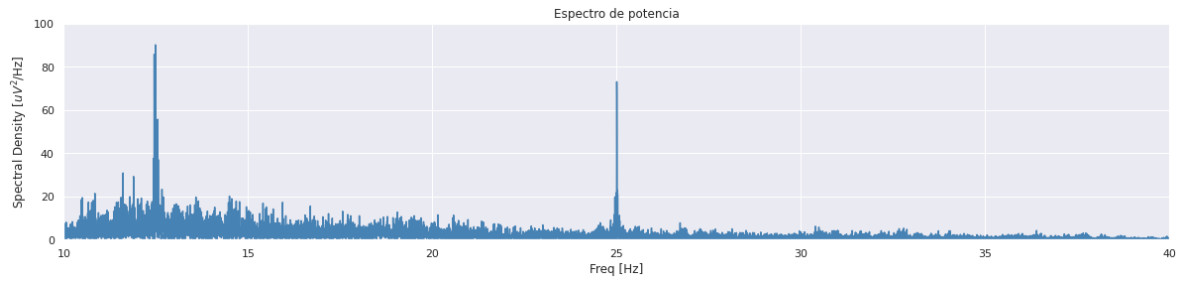
c) Conclusiones:

- A partir de lo observado previamente se puede concluir que no sería necesario incluir todos los canales, debido a que no todos captan la señal de una forma correcta, siendo el canal 2 (ch2) el que más censa la señal correctamente.

• D) Nivel Multi-Paciente:

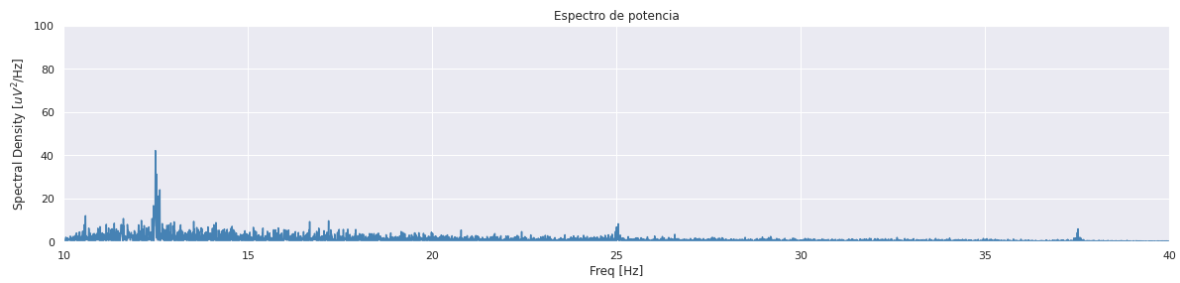
Número de registros por pacientes:

- AA= 63080
- JA= 33631
- HA= 21029
- MA= 12627
- Espectro de potencia estímulo 1 paciente AA:



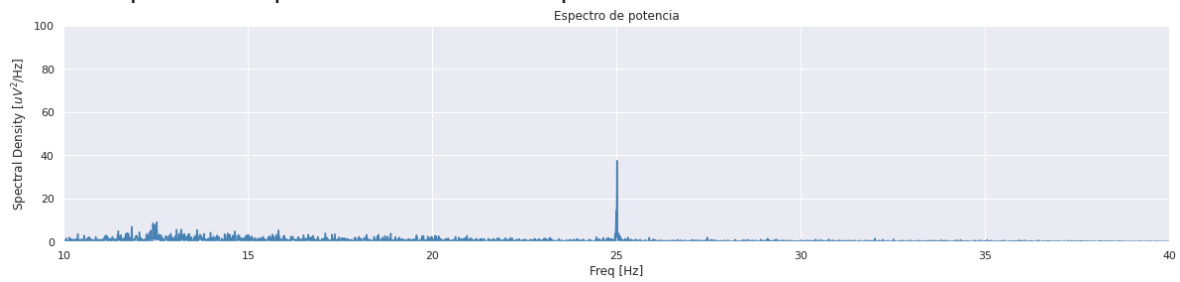
Frecuencia máxima: 12.49

- Espectro de potencia estímulo 1 paciente JA:



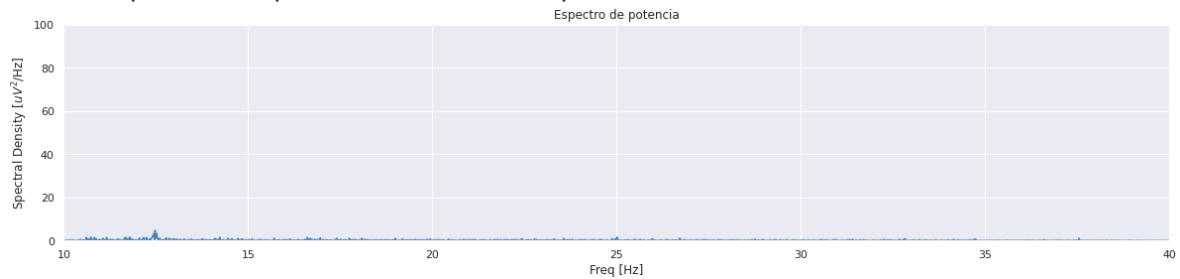
Frecuencia máxima: 12.49

- Espectro de potencia estímulo 1 paciente HA:



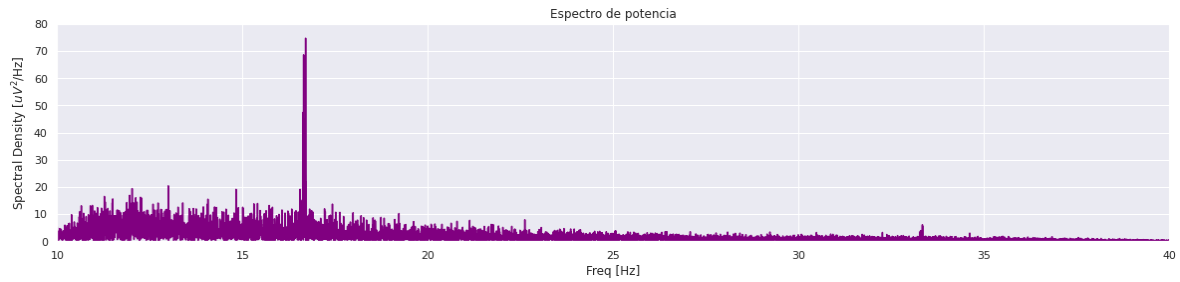
Frecuencia máxima: 25.01

- Espectro de potencia estímulo 1 paciente MA:



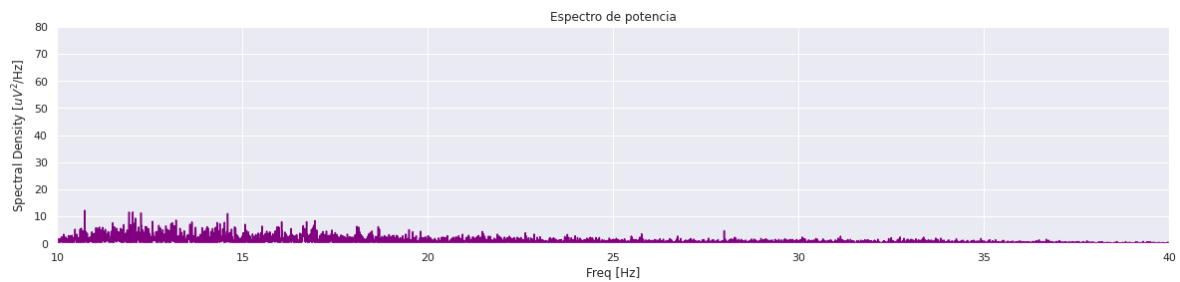
Frecuencia máxima: 12.48

- Espectro de potencia estímulo 2 paciente AA:



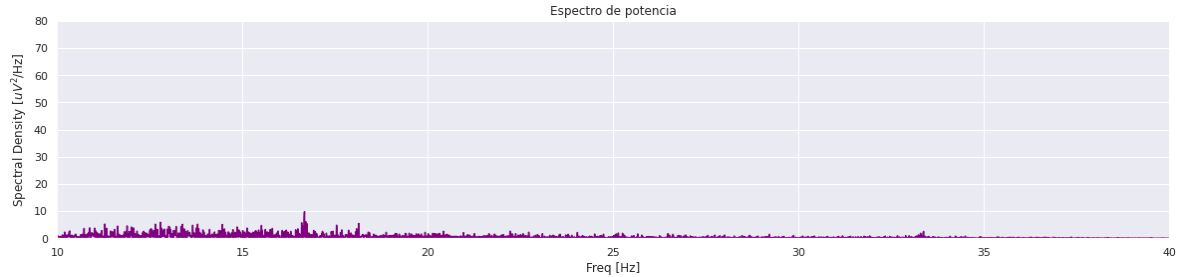
Frecuencia máxima: 16.70

- Espectro de potencia estímulo 2 paciente JA:



Frecuencia máxima: 12.04

- Espectro de potencia estímulo 2 paciente HA:



Frecuencia máxima: 16.67

- Espectro de potencia estímulo 2 paciente MA:

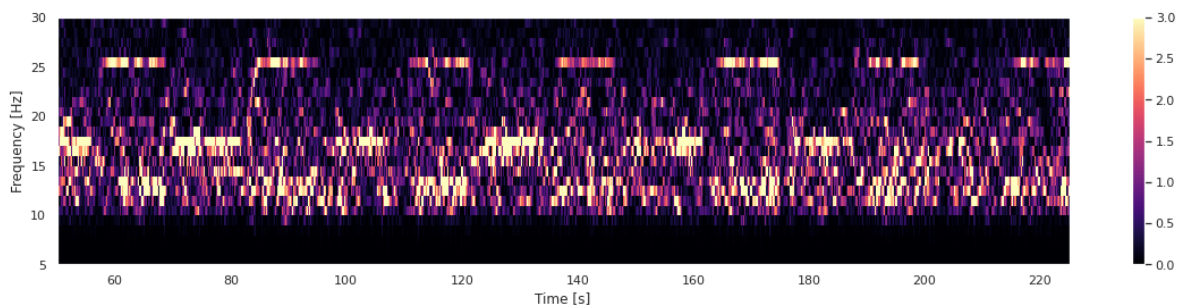


Frecuencia máxima: 16.69

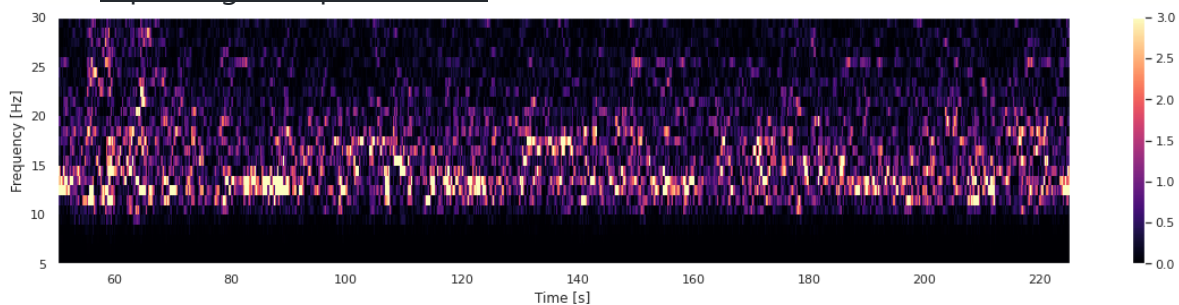
Como se puede observar en los espectros de potencia del canal 2 (ch2), de los distintos pacientes, diferenciados por estímulos, se puede observar que en todos

los casos se encuentran los picos de frecuencias de interés, exceptuando en un caso en particular correspondiente al paciente JA, estímulo 2, en donde la señal máxima corresponde a 12.05 Hz (lo que puede deberse un inconveniente con el electrodo, en donde la señal no fue captada correctamente); lo que también se puede observar es que mientras mayor cantidad de registros posea, mayor es la potencia de los picos de frecuencia de interés, y mayor es la diferenciación con respecto al resto, habiendo una clara diferencia entre los pacientes.

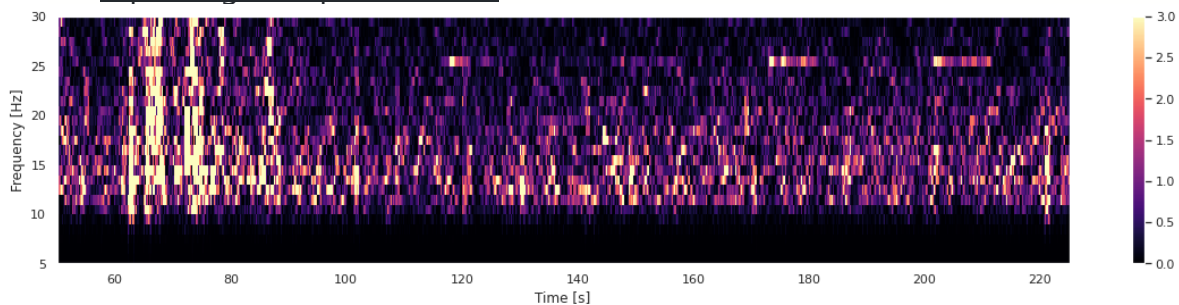
- Espectrograma paciente AA:



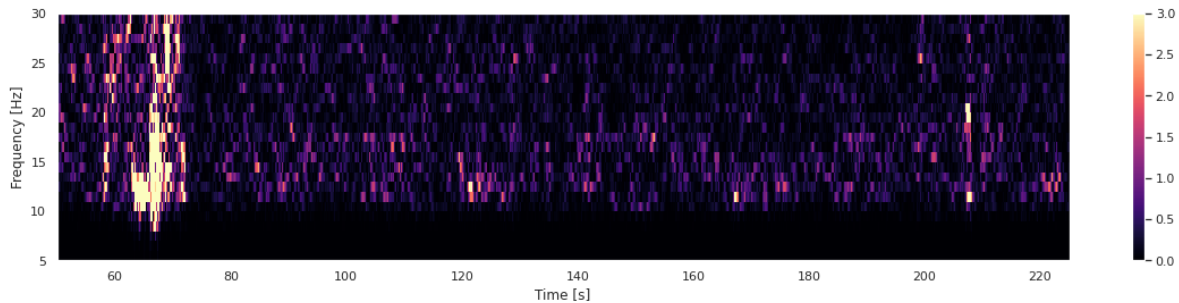
- Espectrograma paciente JA:



- Espectrograma paciente HA:



- Espectrograma paciente MA:



Con los espectrogramas se puede visualizar que si bien los espectrogramas muestran que existe una relación entre el estado y la frecuencia, esta diferenciación de la señal es dependiente de la cantidad de registros que se posean, ya que a medida que mayor cantidad de registros, mayor es la diferenciación de la señal con respecto al estímulo recibido, por lo que el número de registros por pacientes es un número a considerar.

- **E) Comparando con el Trabajo Práctico anterior:**

Cambia la complejidad del análisis de acuerdo a si se trabaja en un dominio o en otro, debido a que para la evaluación del dominio de frecuencia se requiere de conocimiento de dominio de los datos, ya que para su análisis se utilizan herramientas específicas que requieren de un conocimiento previo para ejecutar el análisis, pero esta es la forma correcta de analizar los datos obtenidos, ya que es el tipo de análisis que proporciona la información relevante al objeto de estudio. Por otra parte, el estudio del dominio del tiempo es más complejo en el sentido de que no se obtienen conclusiones que permita una diferenciación entre los estados, y además presenta muchas particularidades de cada sujeto, por lo que no se pueden obtener resultados generalizables, a diferencia de lo que ocurre en el estudio del dominio de la frecuencia, donde se llega en todos los casos a las mismas conclusiones.

Para implementar un modelo de clasificación supervisada se requiere realizar el análisis por medio del dominio de frecuencia ya que este tipo de análisis brinda la información necesaria para poder clasificar un estímulo en las distintas categorías, y esto no lo brinda el análisis temporal, ya que el estímulo que esté recibiendo el paciente se relaciona con las frecuencias de las señales que se reciben, y no con las variaciones de voltaje a lo largo del tiempo, ya que en el trabajo práctico anterior se determinó que no existe una correlación entre los valores de amplitud de los registros y el estado. El análisis en el dominio del tiempo no brinda información que permita una diferenciación evidente y clara de los estados, pero el análisis en el dominio de frecuencia sí.