

BCI_TP2

August 2, 2022

Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones Edición 2022

0.1 Mentoría: Data Science aplicado a BCI

0.2 Grupo 2

Integrantes: Gastón Briozzo, Pablo Ventura

Profesor de Práctico: Juan Manuel Lopez

1 A) Nivel Segmento/Estado: Seleccione los datos correspondientes a un paciente y un canal, y para él defina un conjunto de datos para cada estado presente en el dataset. Para cada uno de ellos estudie los siguientes elementos y luego compárelos.

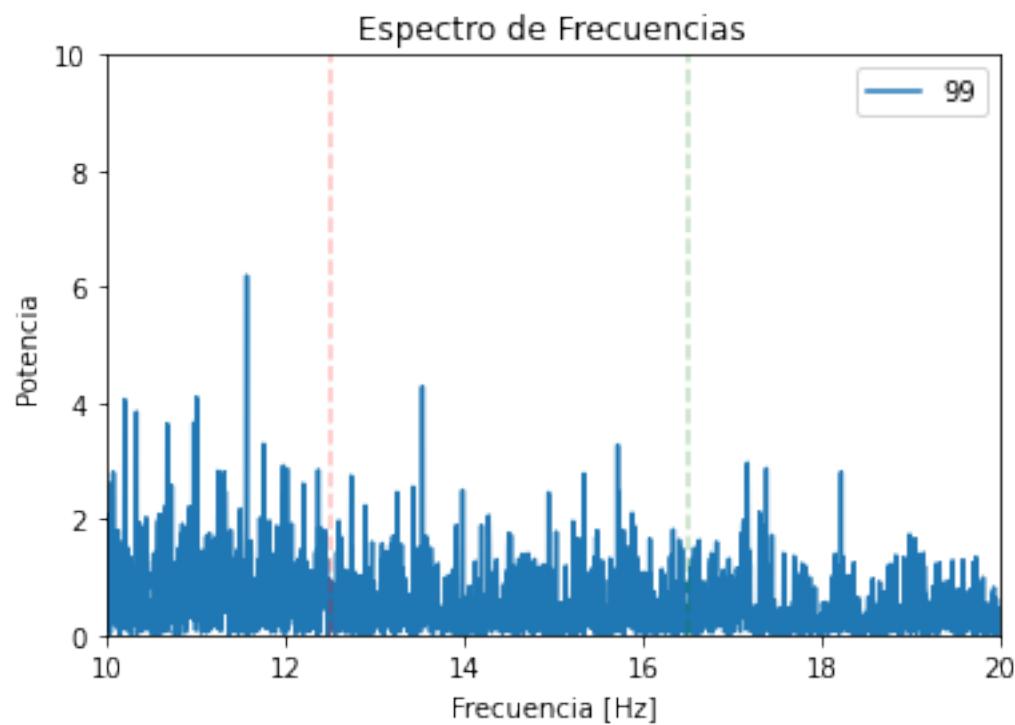
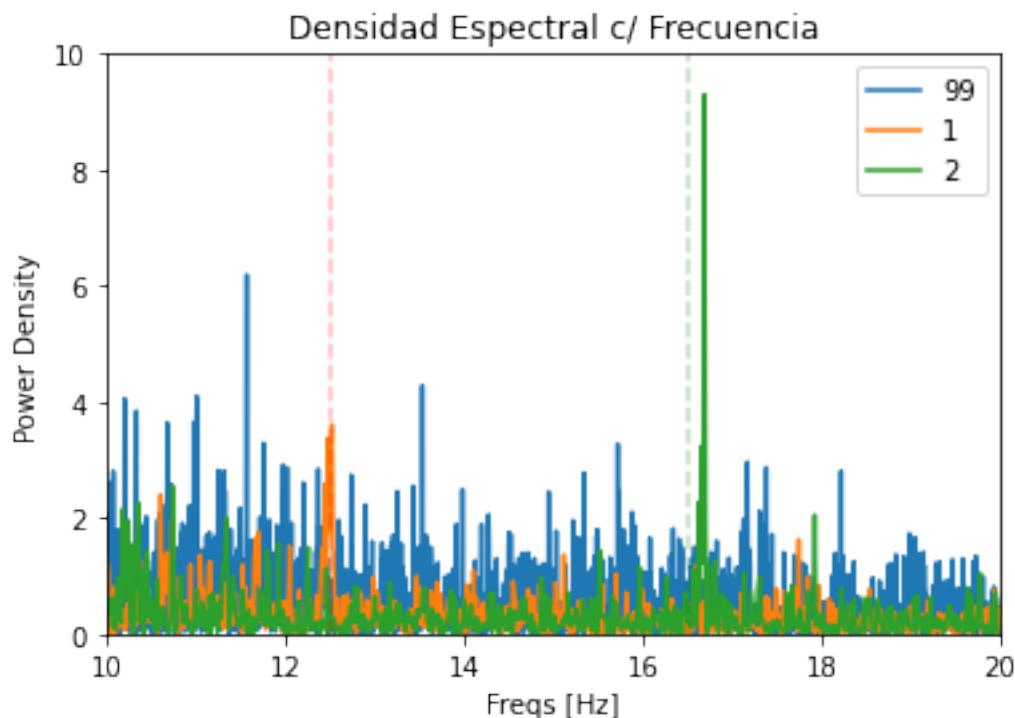
1.1 A) a) A partir de la frecuencia de muestreo previamente determinada y usando el teorema de Nyquist ($f_{NQ} = f_0/2$). ¿Qué frecuencia máxima pudo ser registrada en los datos disponibles?

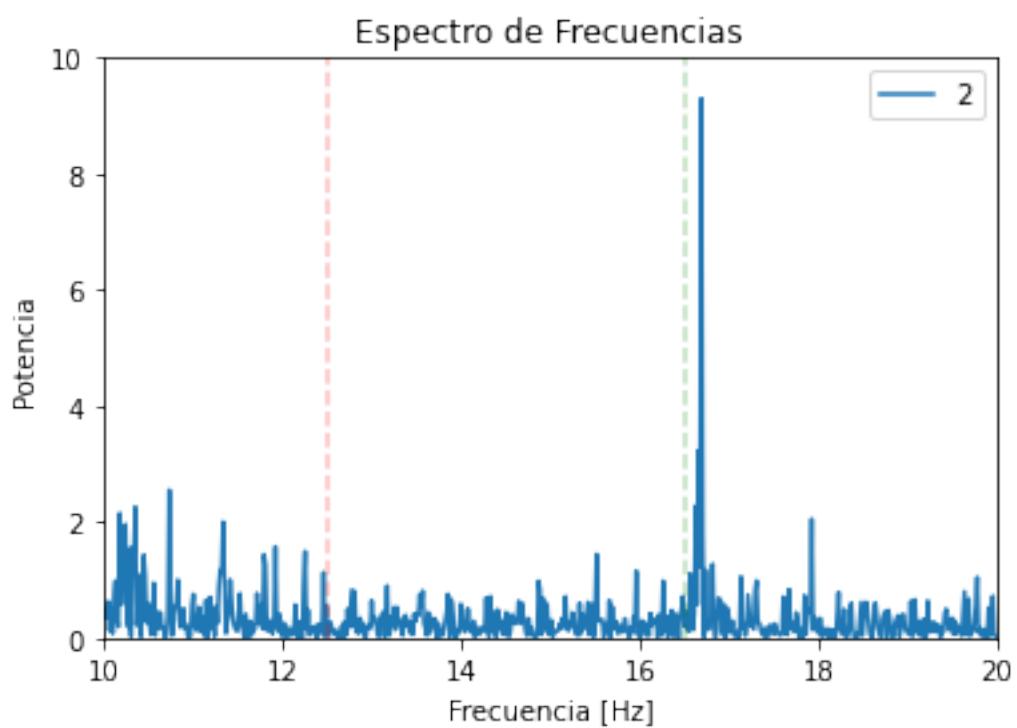
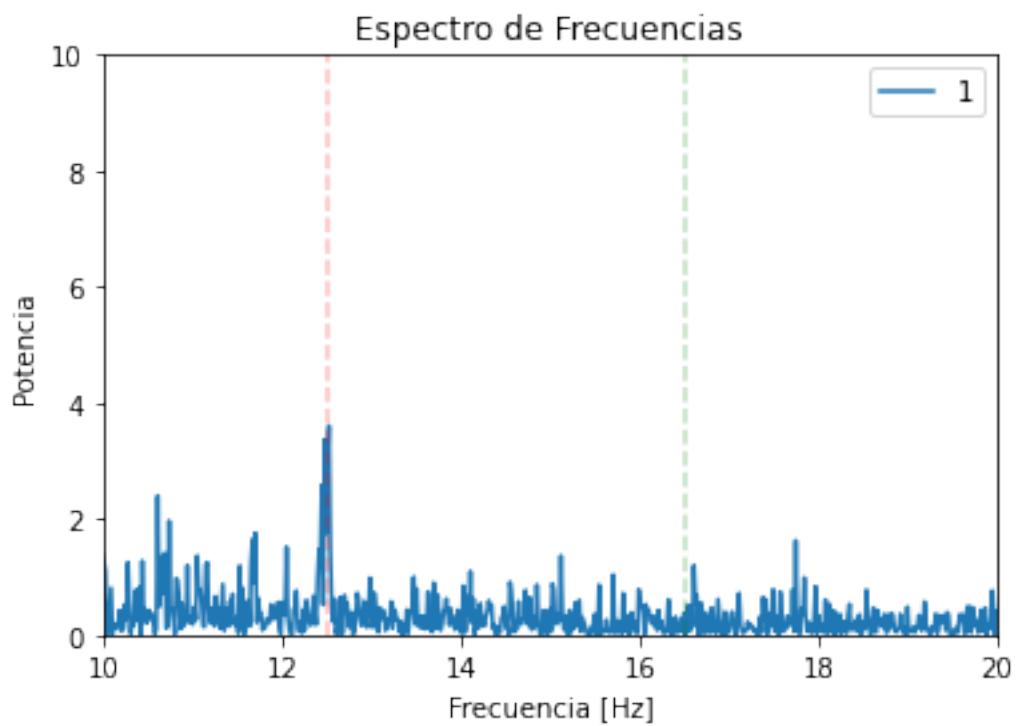
La frecuencia de muestreo experimental es $f_0 = 200\text{Hz}$. Esto significa que si los instrumentos registran una señal periódica cuya frecuencia es de 200Hz , medirán un valor constante, por lo que esta no será interpretada correctamente.

El Teorema de Nyquist nos dice que la mayor frecuencia de entrada interpretable por un instrumento con frecuencia de medición f_0 está dada por $f_{NQ} = f_0/2$.

De esta forma, la mayor frecuencia registrable en este experimento es de $f_{NQ} = 200/2\text{Hz} = 100\text{Hz}$.

- 1.2 A) b) Compute y grafique la densidad espectral de potencia para cada segmento. Elija las escalas apropiadas para su visualización (qué magnitudes están graficando? corresponde el uso de decibeles?).





- 1.3 A) c)** Determine la/las frecuencias con mayor presencia en cada caso. ¿Qué puede concluir con esta nueva información? ¿Encuentra diferencias en este aspecto entre los tramos 99 y los tramos 1-2? ¿Hay frecuencias no asociadas con el fenómeno a detectar?

En cada caso, las frecuencias con mayor presencia se corresponden con las frecuencias de estimulación a la que se encuentra sometido el paciente. Esto es, 12.5 y 16.5 Hz para los estados 1 y 2 respectivamente. Por su parte, para el estado 99 no vemos ninguna frecuencia predominante, mostrando un espectro que se corresponde con ruido.

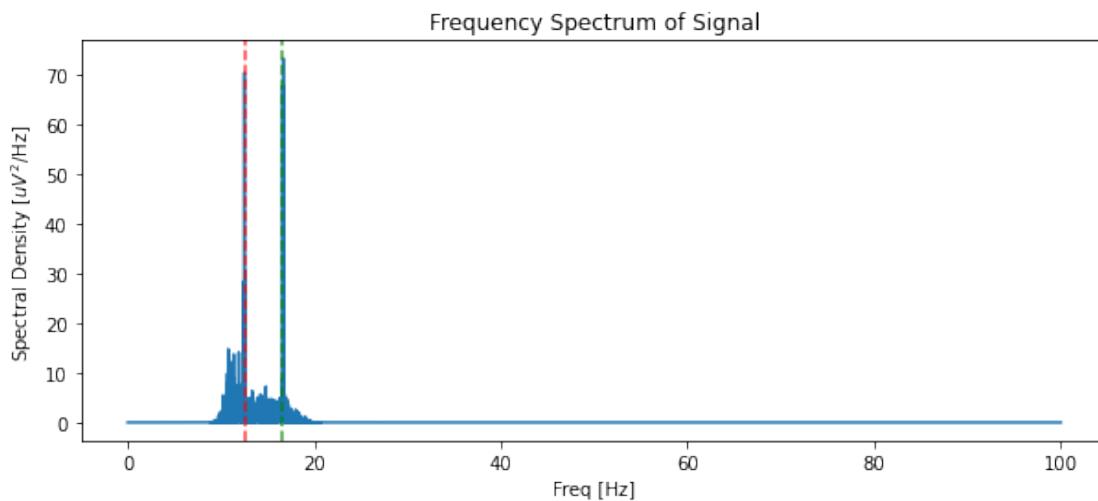
Las contribuciones de las demás frecuencias son claramente distinguibles del pico central correspondiente al estado de estimulación.

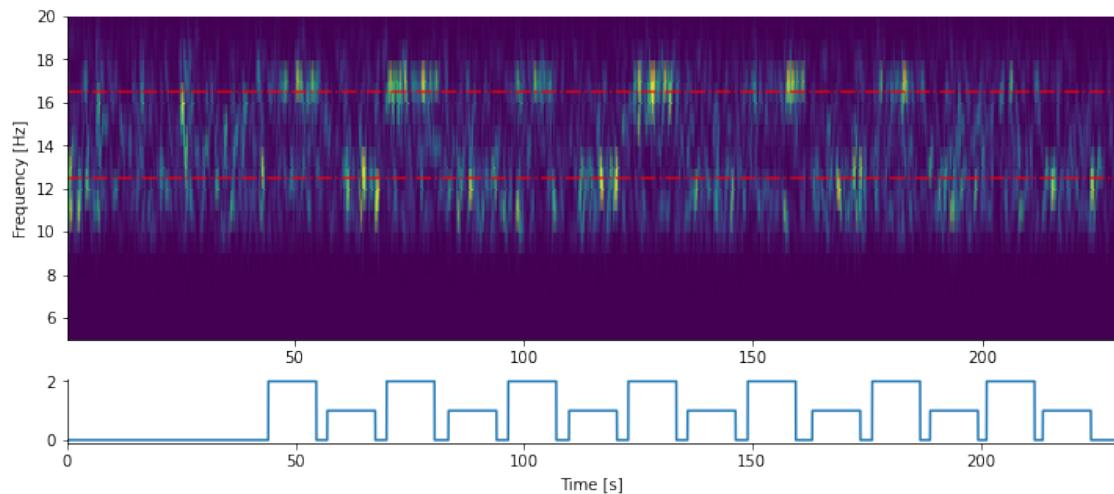
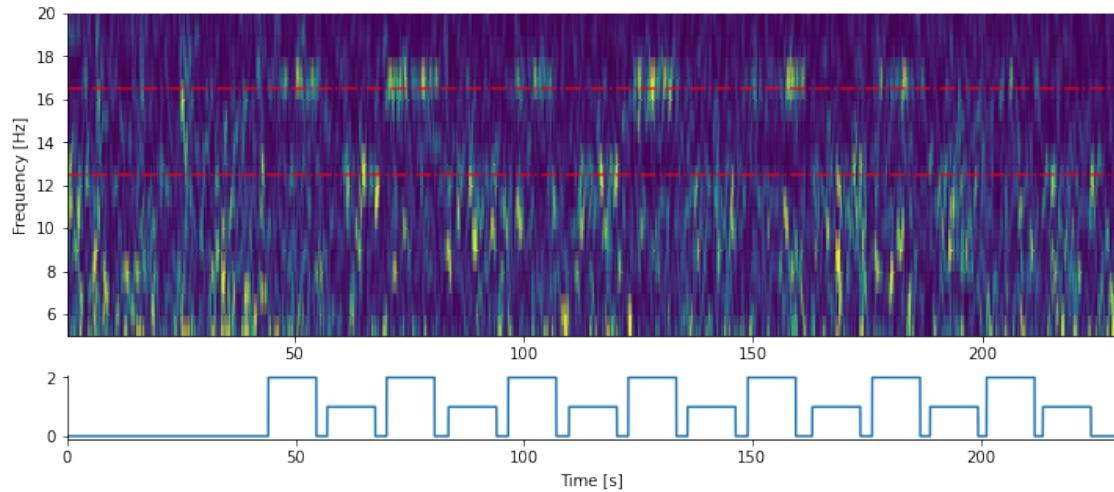
- 1.4 A) d)** ¿Es útil el concepto de outliers o conviene usar otros criterios? ¿Se pueden descartar señales?

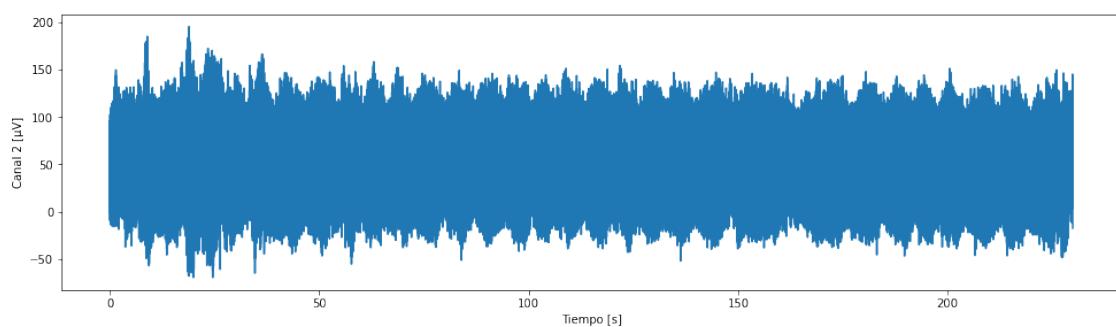
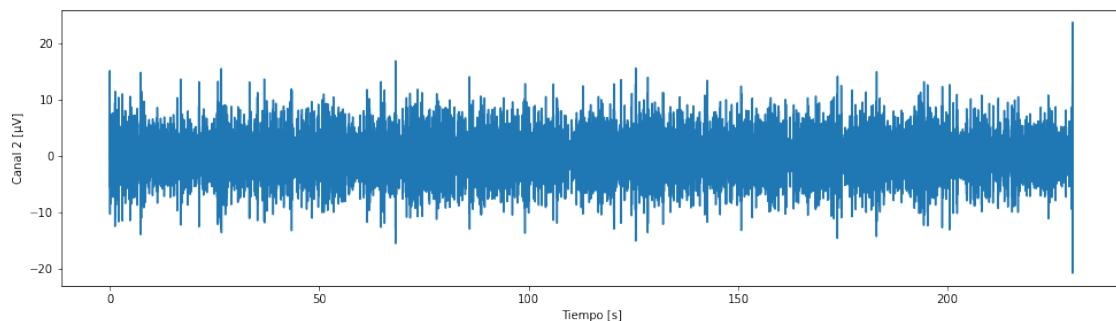
Al considerar la transformada de fourier, la información relevante se corresponde a la posición central del pico, es decir, la frecuencia principal. Por este motivo, el concepto de outlier deja de ser útil.

Podemos descartar el estado 99, correspondiente al paciente sin ser estimulado. Esta señal no presenta un pico central, sino una distribución aleatoria de ruido.

- 1.5 A) e)** Comparar en el dominio del tiempo las señales con y sin filtrado de frecuencias indeseadas. ¿Qué diferencias encuentra?



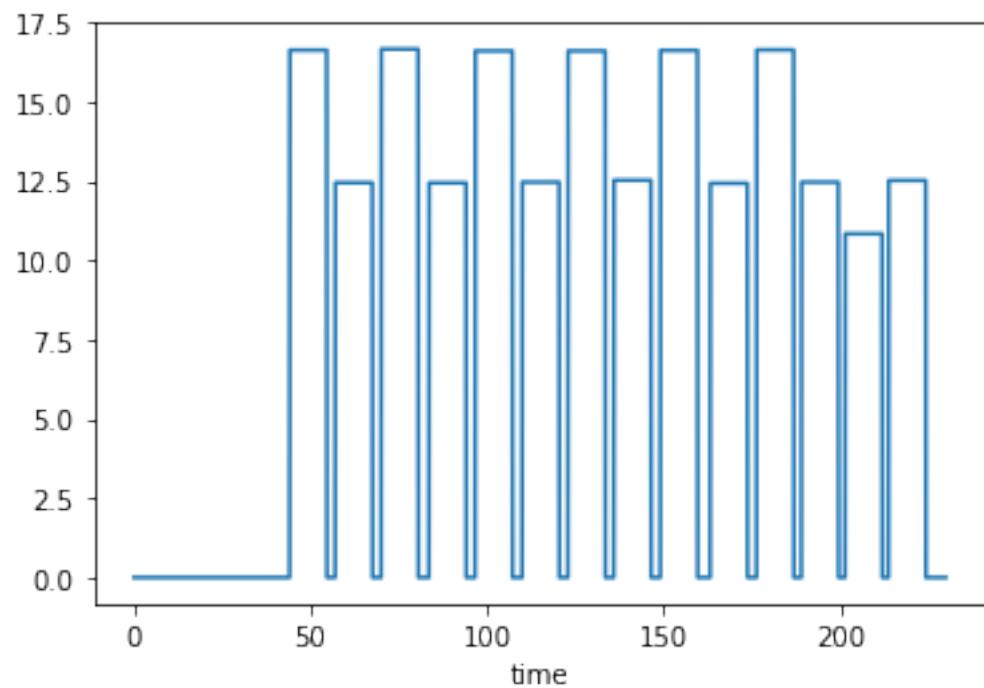
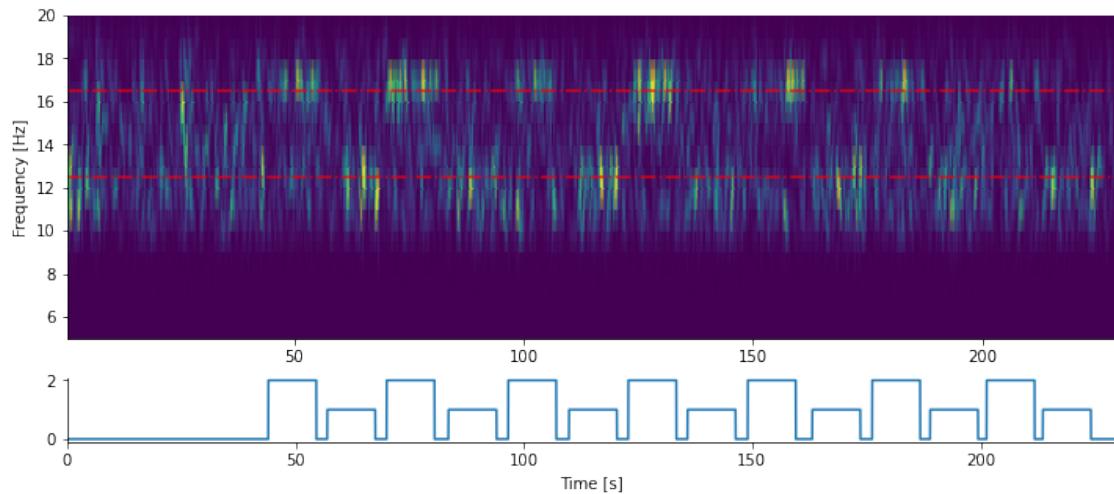




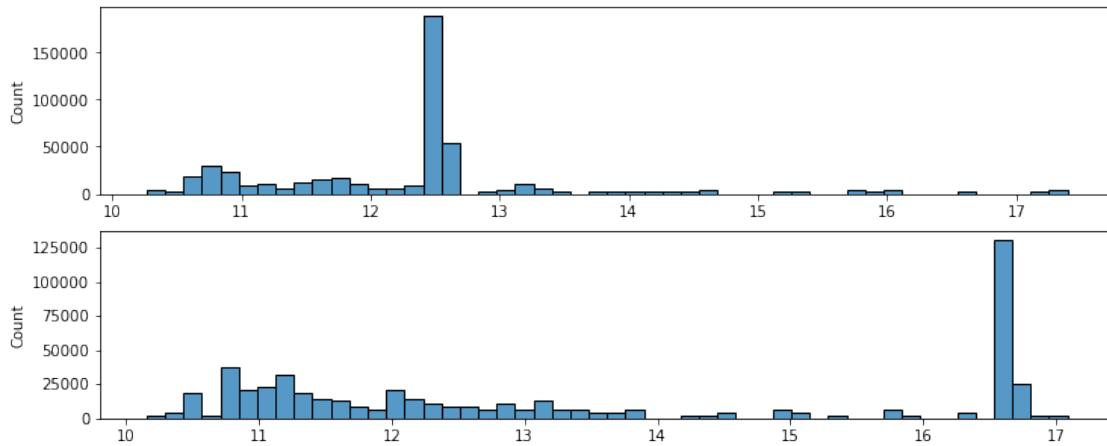
La diferencia fundamental es que en la señal filtrada no se ve valor central ni deriva.

2 B) Nivel Paciente - un canal: Seleccione los datos correspondientes a un paciente y un canal de adquisición y para ese caso estudie los siguientes elementos:

- 2.1 B) a) Para cada intervalo de adquisición de cada estado, determine la frecuencia de mayor presencia relevante al problema estudiado. Ej: Un canal tiene 10 intervalos 2, para cada uno de ellos la frecuencia máxima oscila alrededor de 12.5.



2.2 B) b) Realice un análisis estadístico de dichas frecuencias a lo largo de todo el conjunto de intervalos para cada clase. Determine su distribución y su resumen estadístico según considere apropiado. Ej: Siguiendo el ejemplo anterior, cómo es la distribución de esas frecuencias alrededor de 12.5 en el intervalo 2 y de la misma manera para los otros.



Vemos que, para cada estado, el histograma presenta un pico en la frecuencia correspondiente, mostrando un conteo sumamente superior al resto. En particular, para el estado 2, existe una gran cantidad de ruido en las frecuencias bajas, que aleja tanto a la media como a la mediana del valor esperado.

2.3 B) c) Determine si existe una diferencia estadísticamente significativa, entre los valores centrales de frecuencias para los estados existentes.

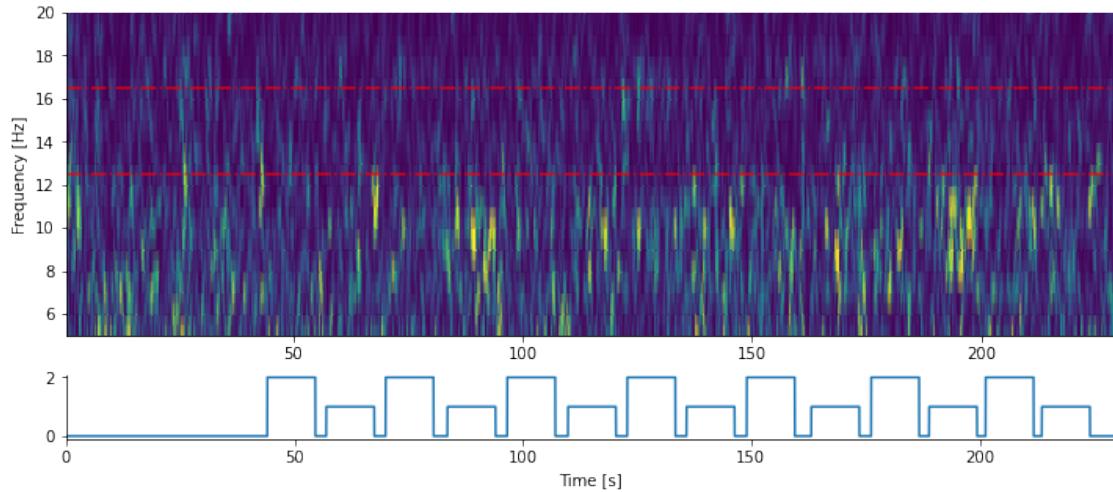
La diferencia entre las distribuciones es suficiente como para afirmar que son diferentes.

2.4 B) d) ¿Son variables independientes el estado registrado de la señal y la frecuencia mayoritaria presente en el registro? Use herramientas cuantitativas y cualitativas para justificar su respuesta.

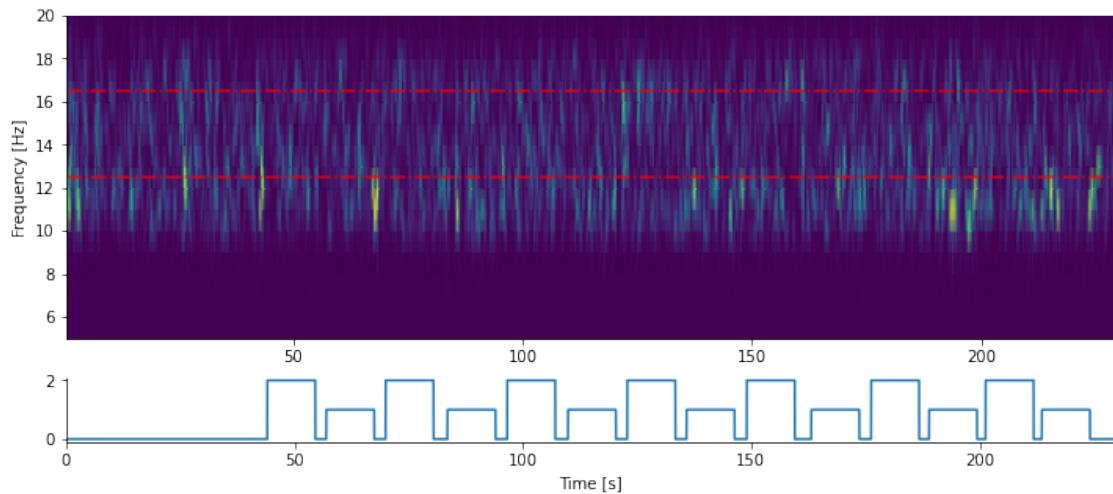
El estado registrado de la señal y la frecuencia mayoritaria son variables claramente dependientes, como se puede ver en la correlación.

2.5 B) e) Compute y grafique el espectrograma del registro completo. Probar con el registro filtrado y sin filtrar y comentar las diferencias. Ver link informativo, si quedan dudas conceptuales no duden en preguntar.

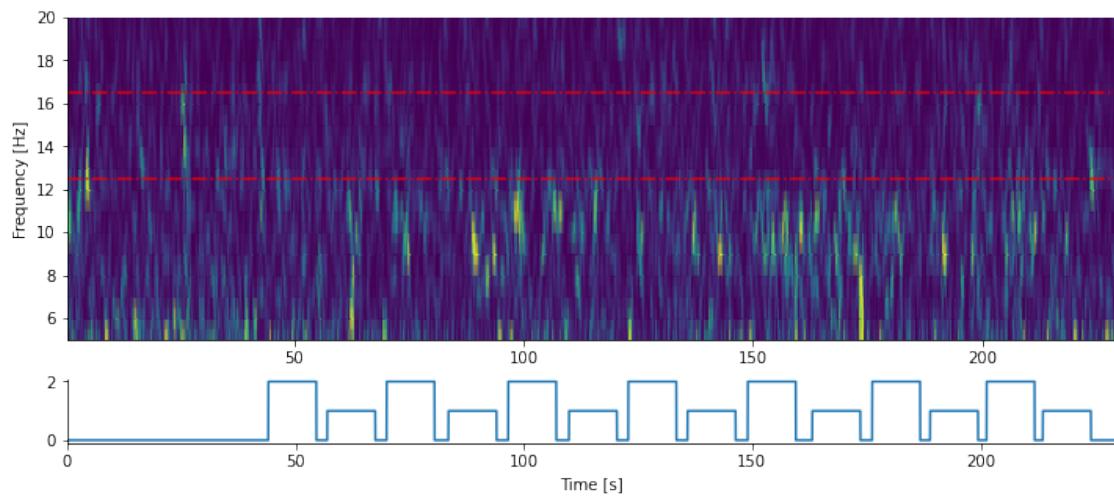
AA0:ch0 sin filtrar



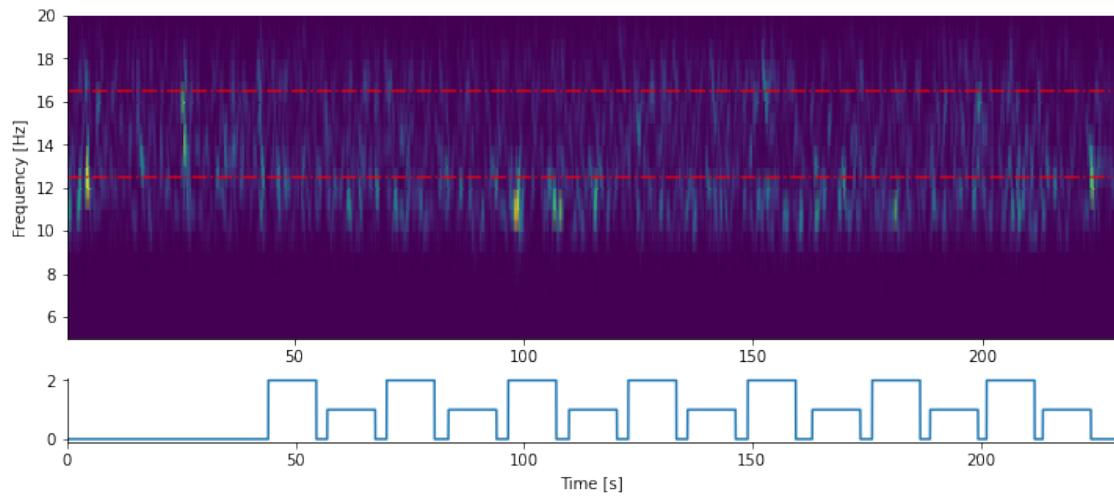
AA0:ch0 filtrado



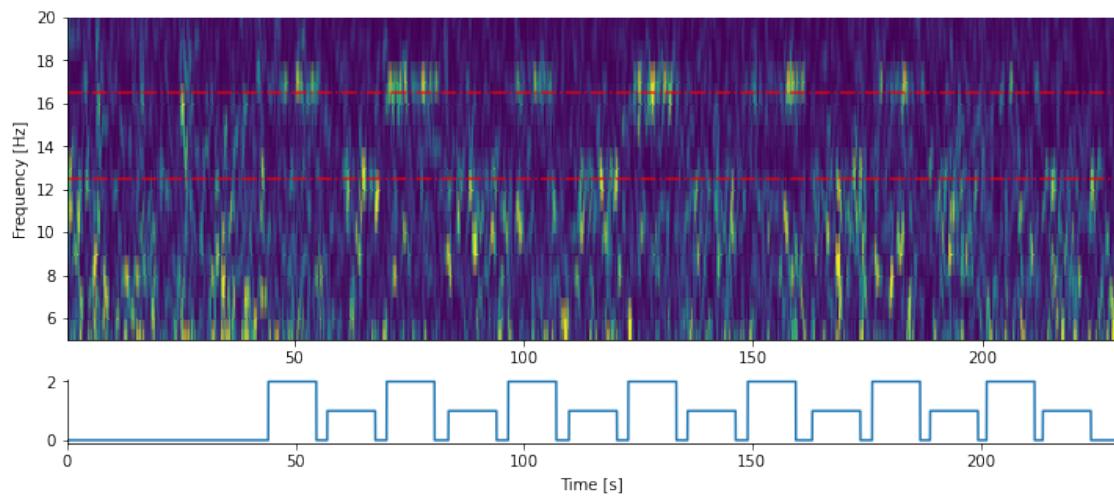
AA0:ch1 sin filtrar



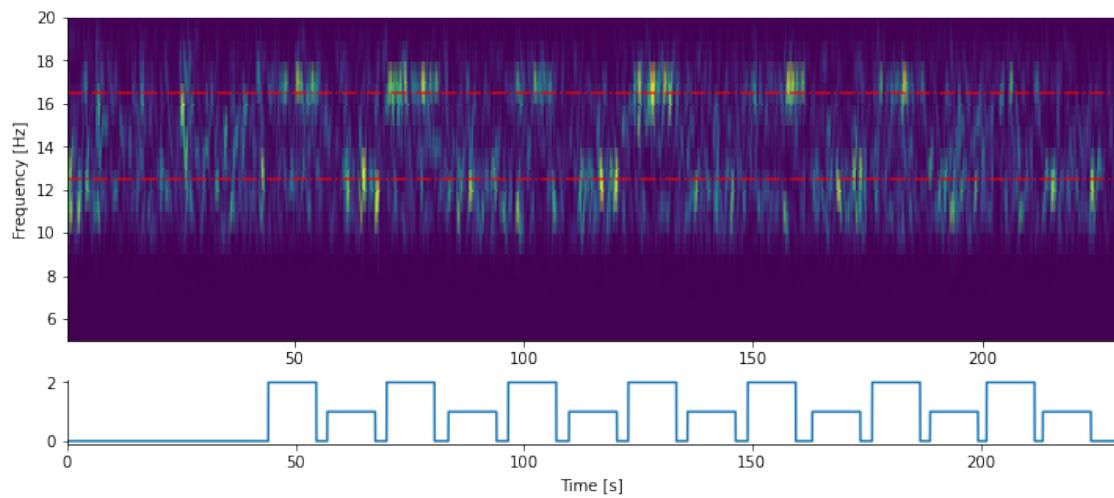
AA0:ch1 filtrado



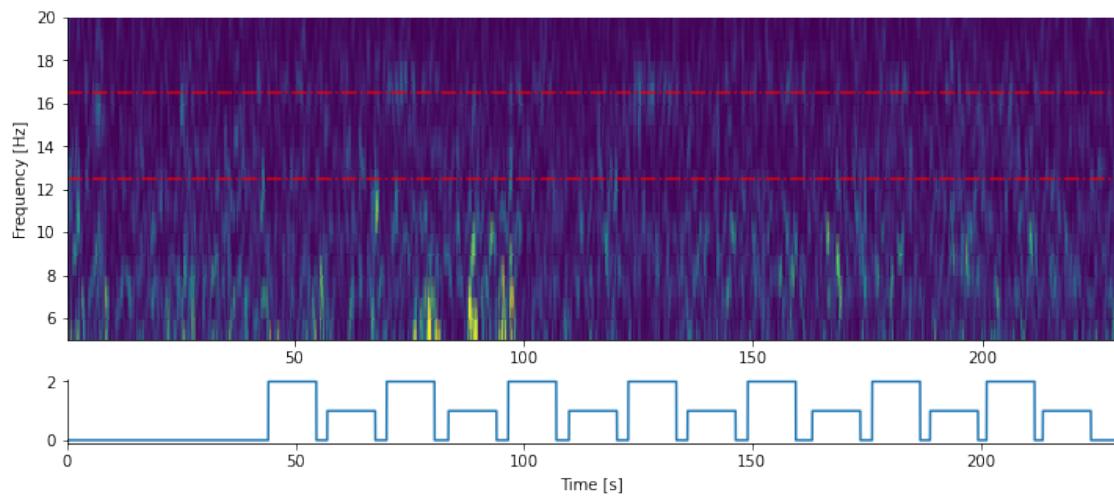
AA0:ch2 sin filtrar



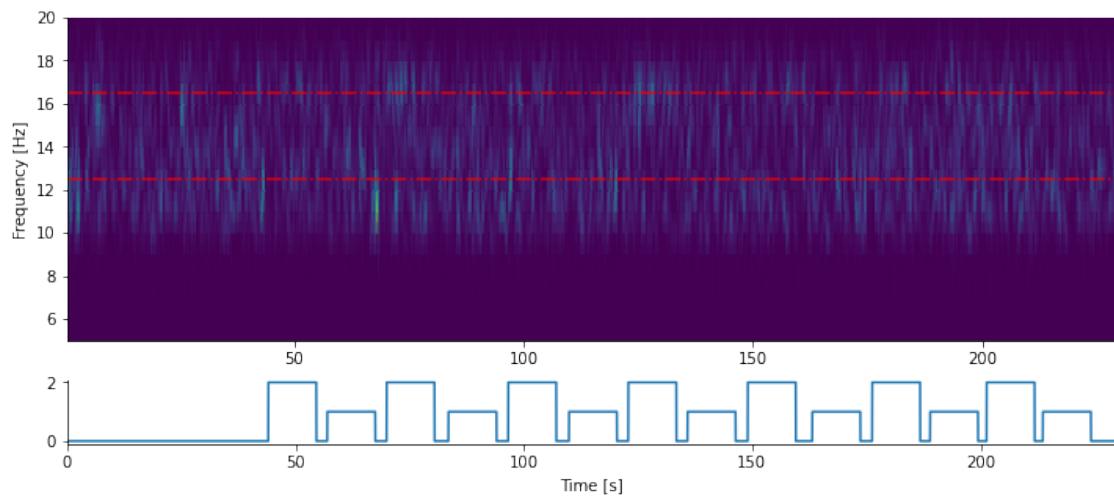
AA0:ch2 filtrado



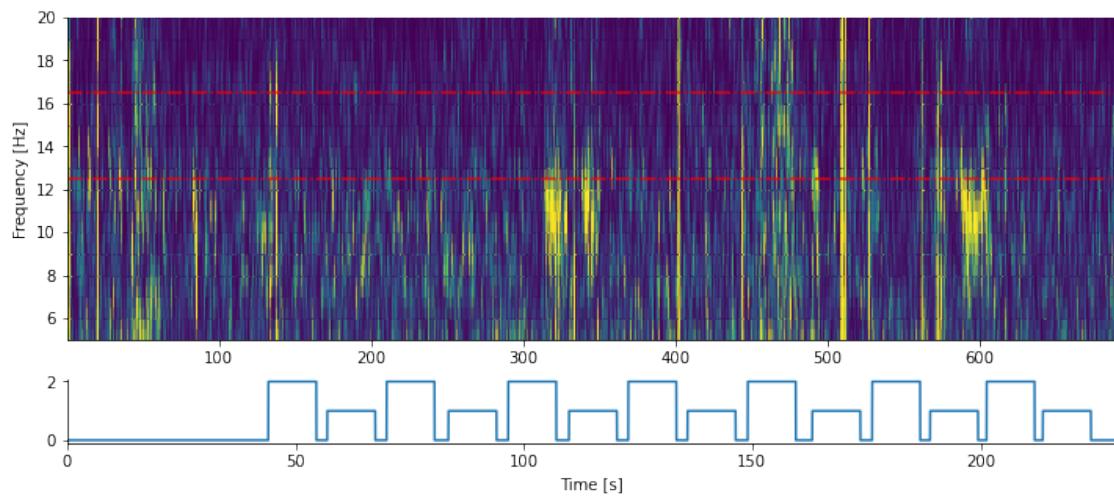
AA0:ch3 sin filtrar



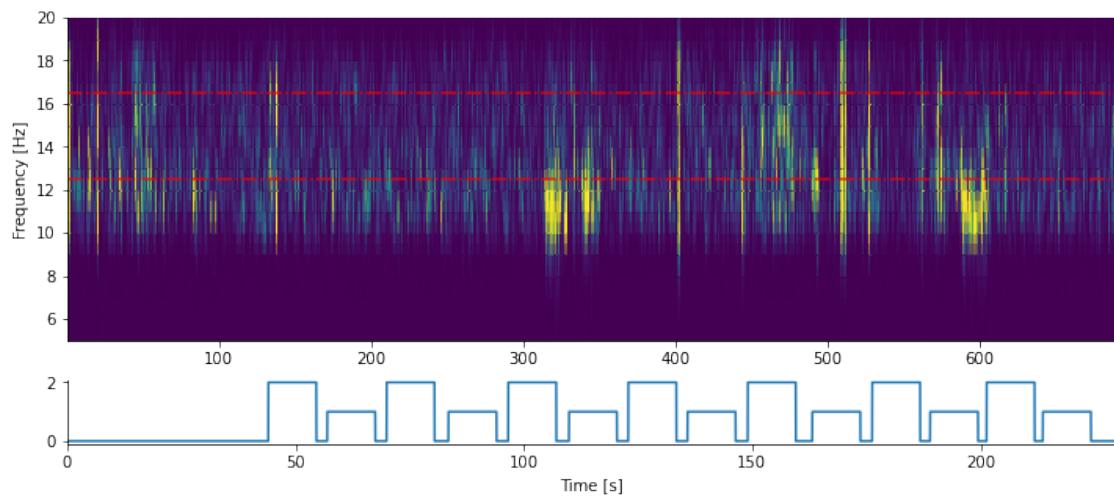
AA0:ch3 filtrado



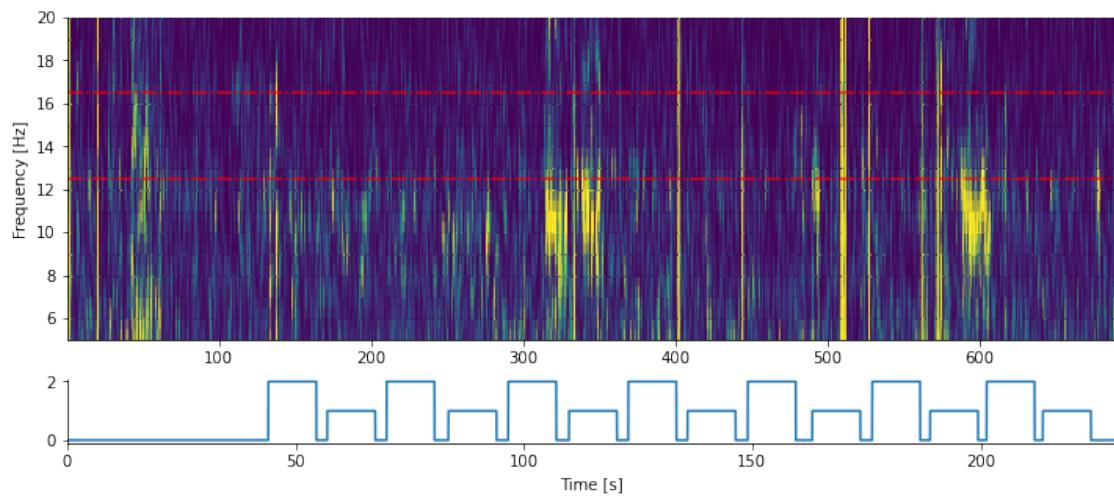
AA1:ch0 sin filtrar



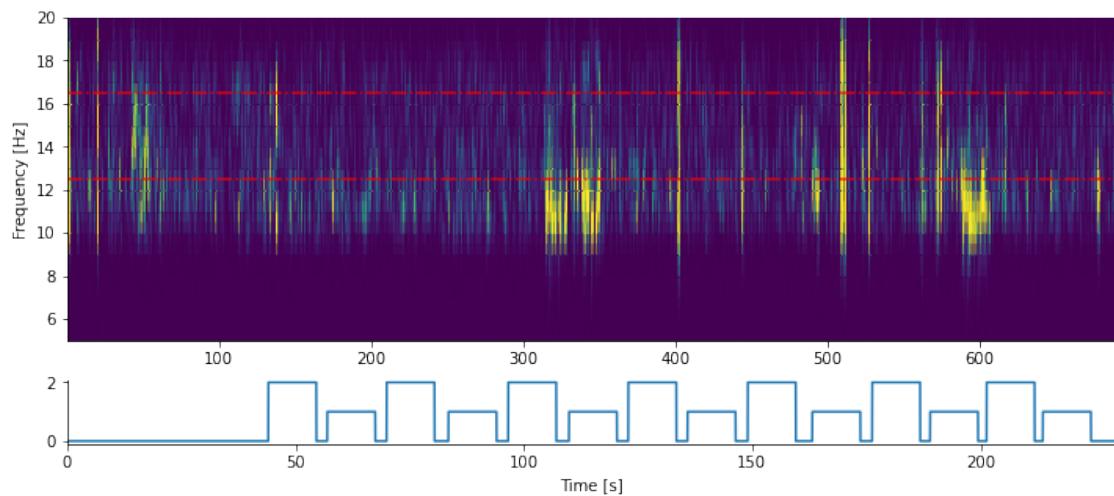
AA1:ch0 filtrado



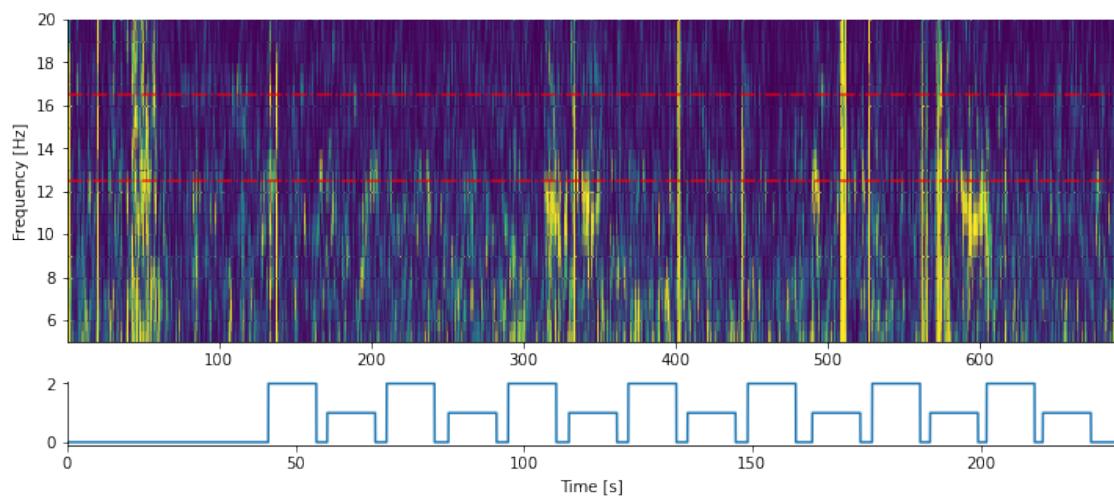
AA1:ch1 sin filtrar



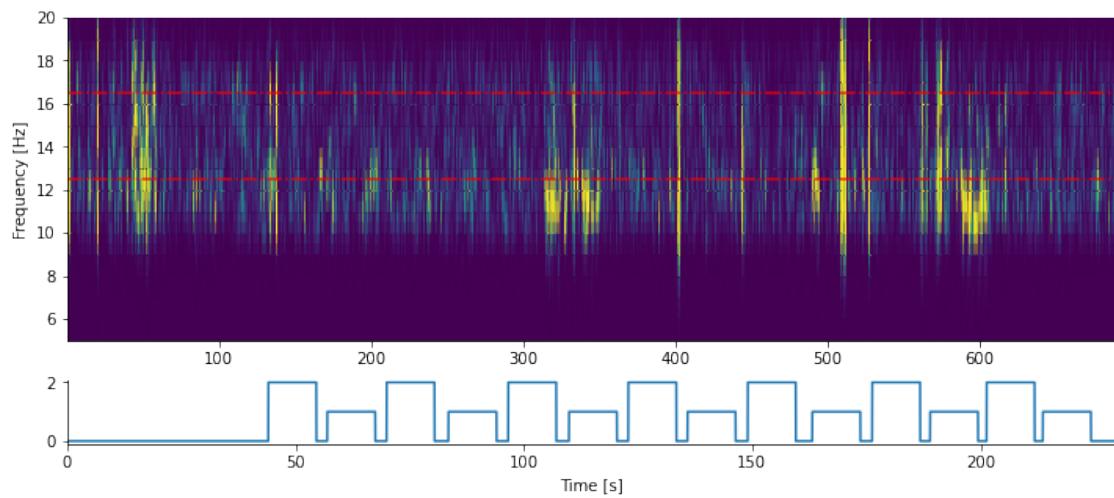
AA1:ch1 filtrado



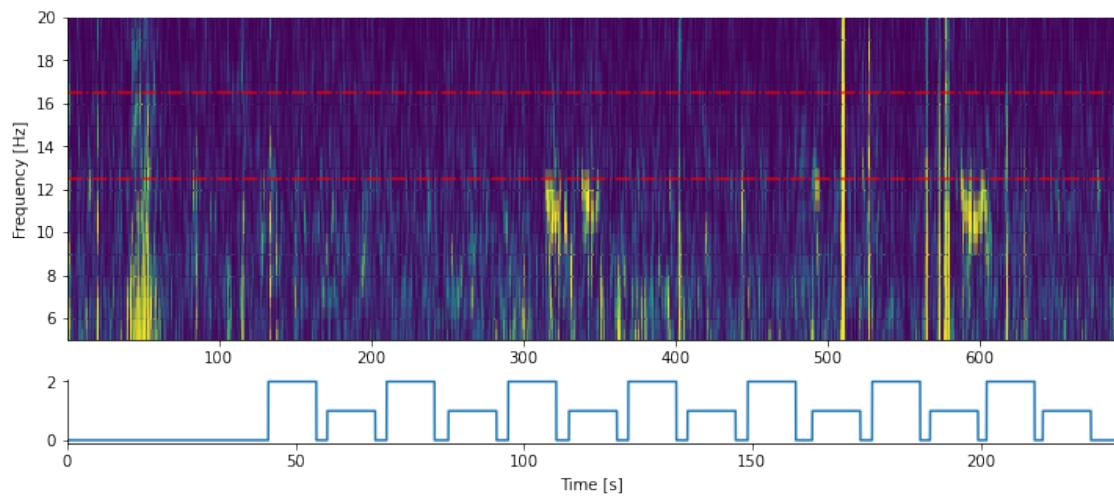
AA1:ch2 sin filtrar



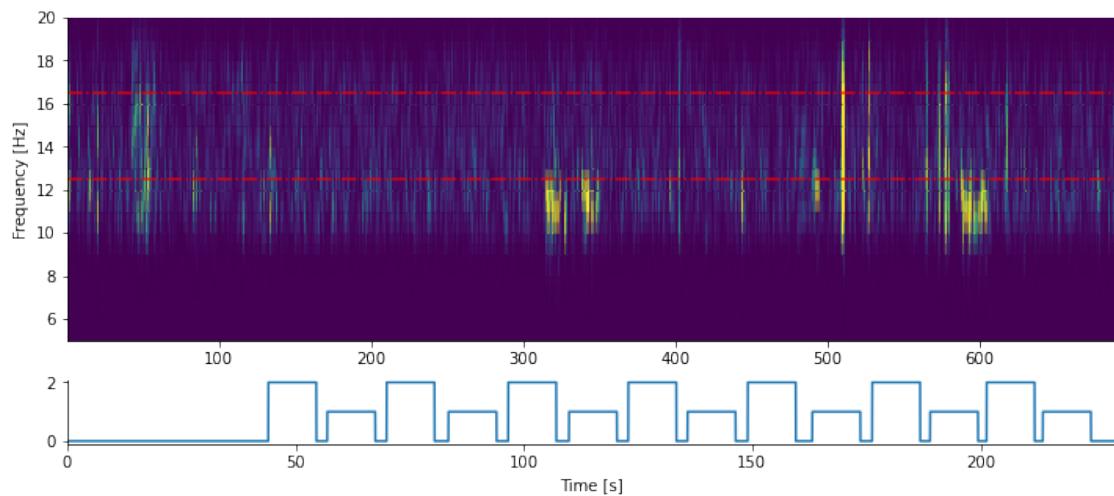
AA1:ch2 filtrado



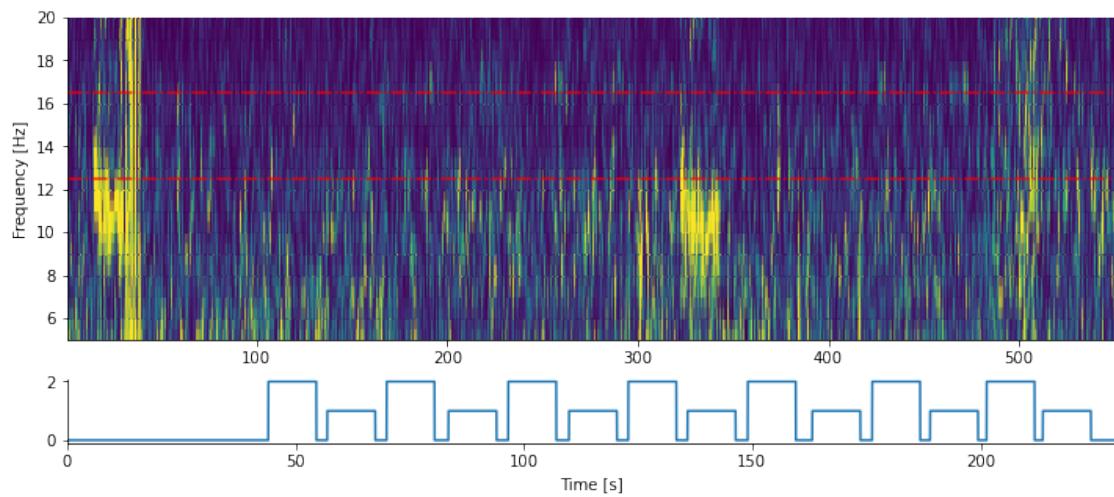
AA1:ch3 sin filtrar



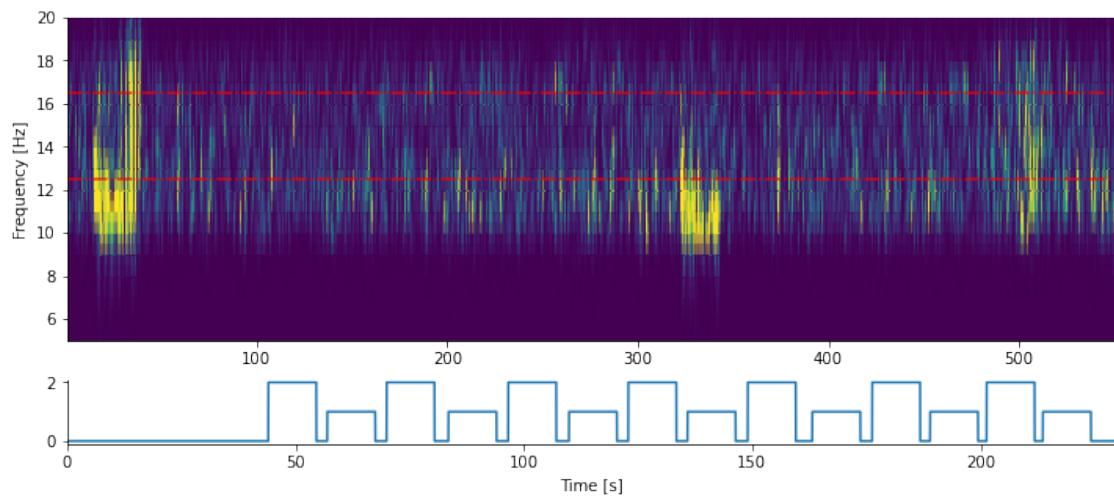
AA1:ch3 filtrado



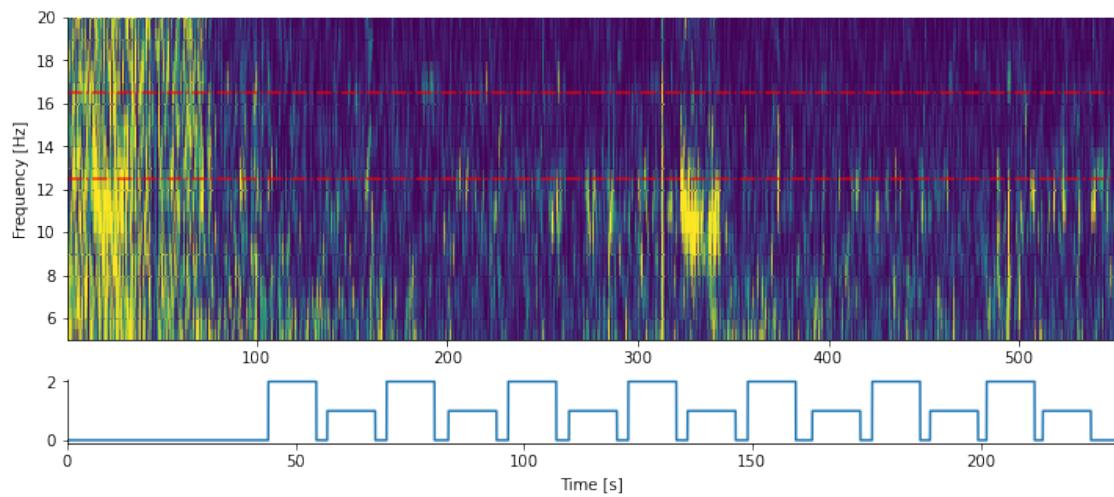
AA2:ch0 sin filtrar



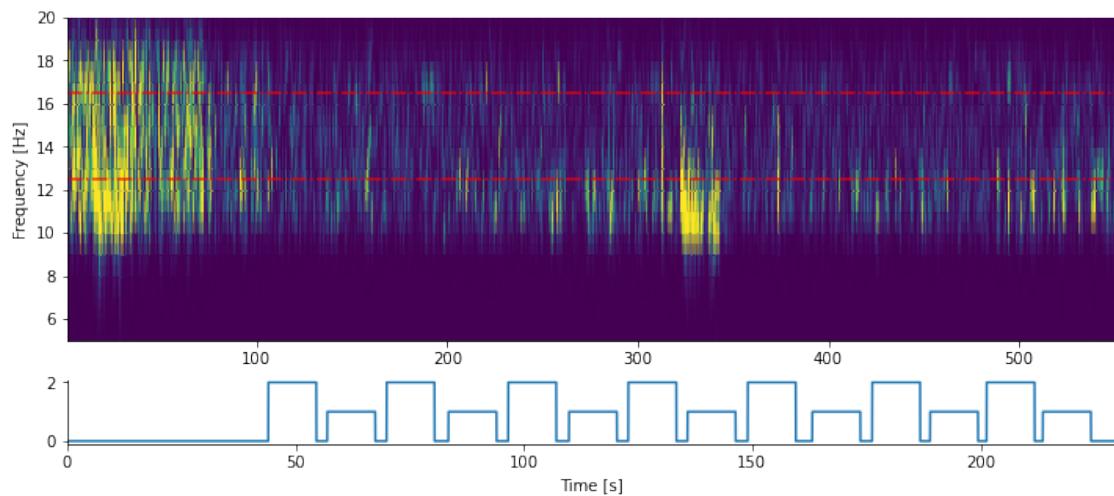
AA2:ch0 filtrado



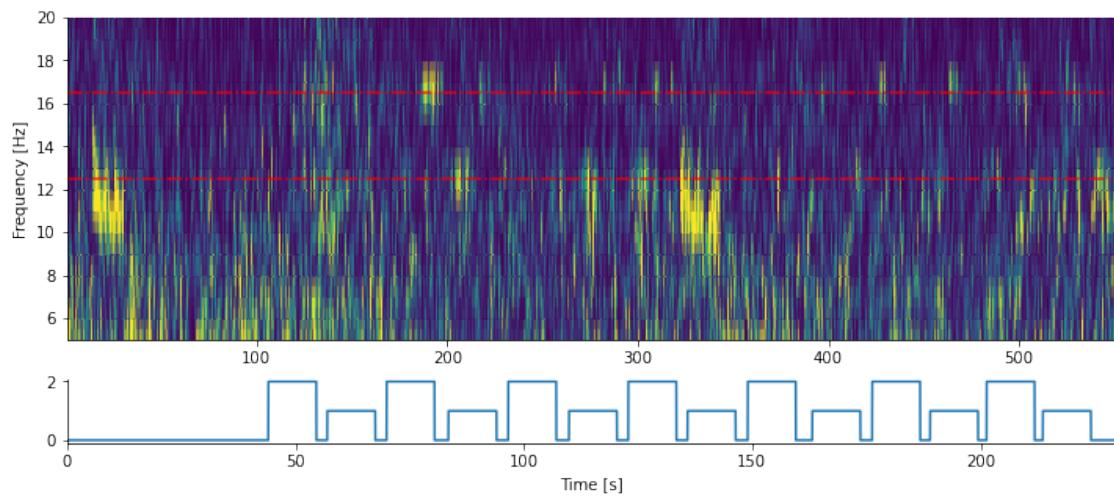
AA2:ch1 sin filtrar



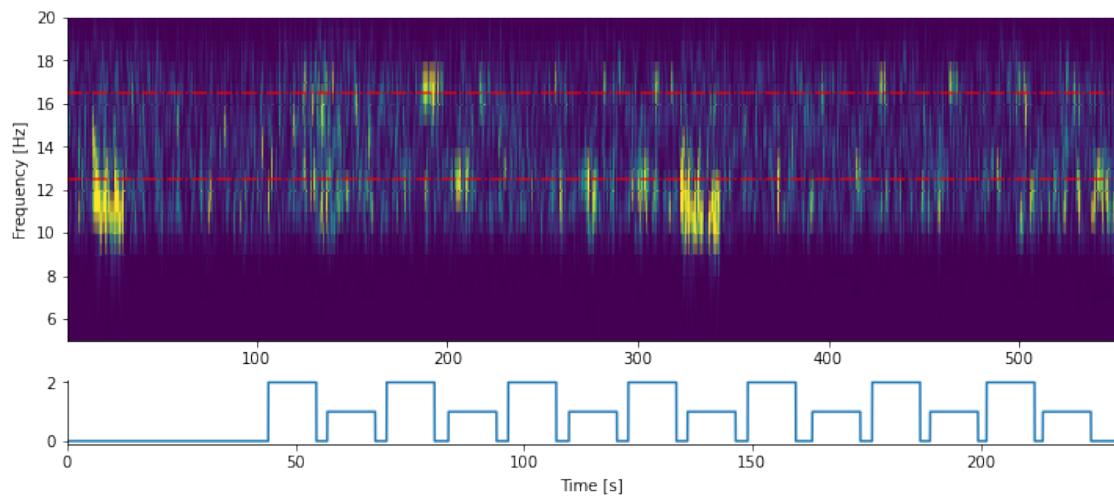
AA2:ch1 filtrado



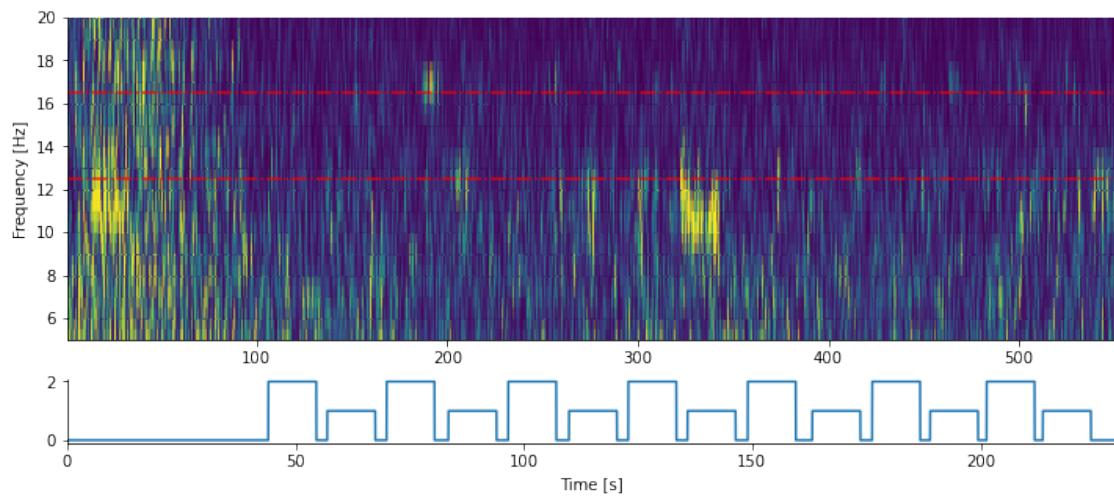
AA2:ch2 sin filtrar



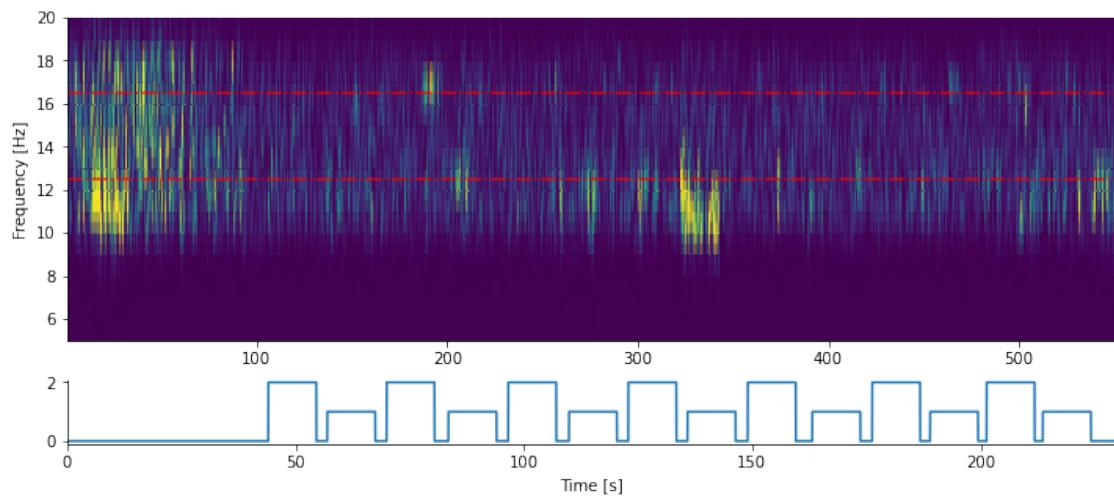
AA2:ch2 filtrado



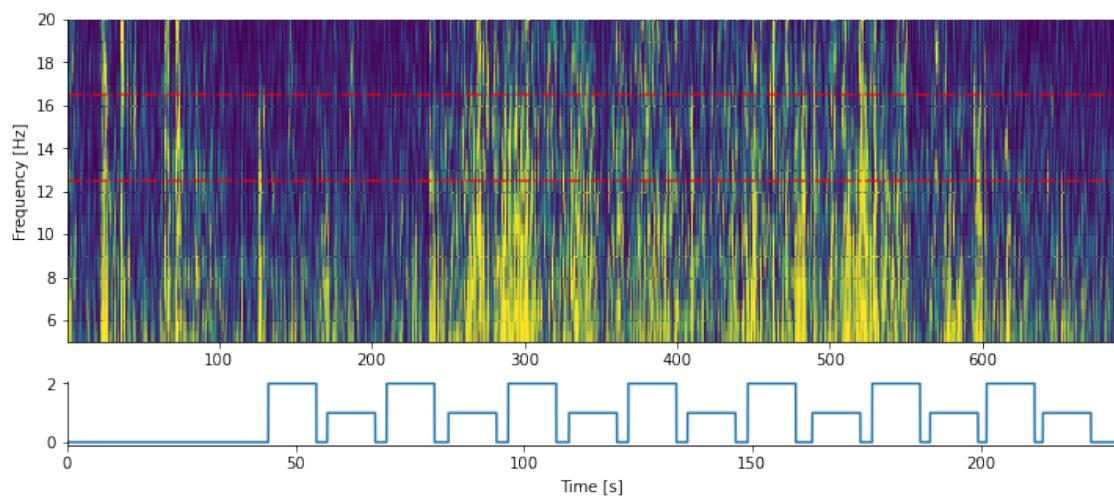
AA2:ch3 sin filtrar



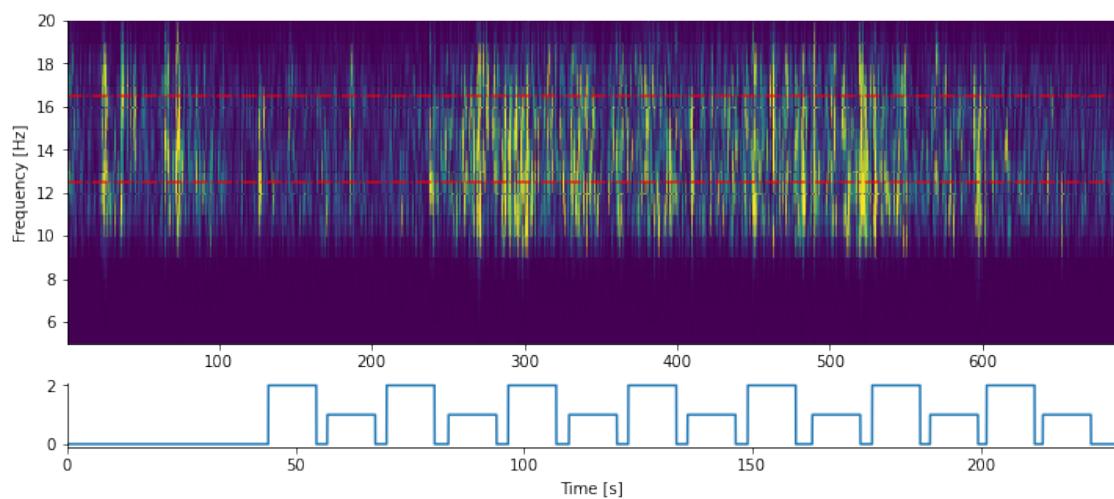
AA2:ch3 filtrado



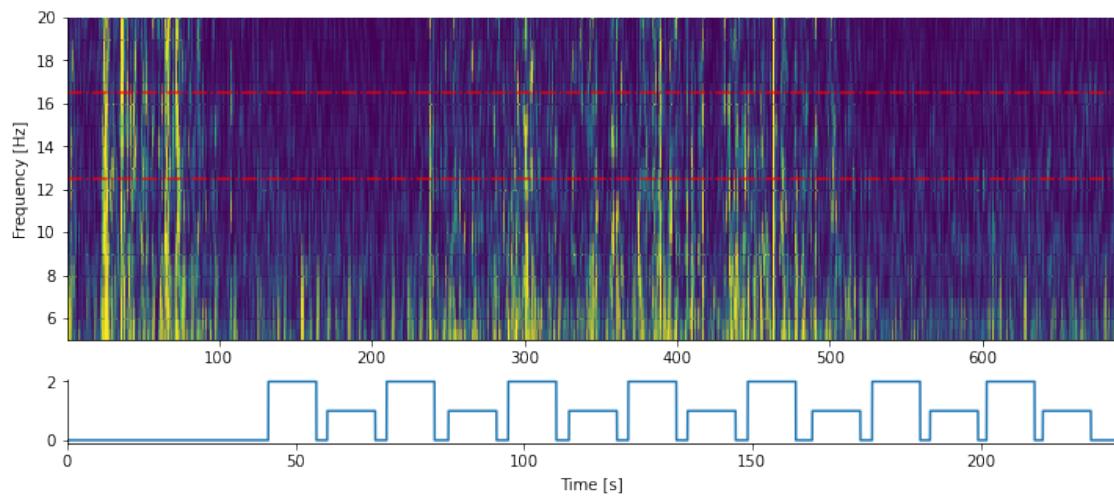
HA1:ch0 sin filtrar



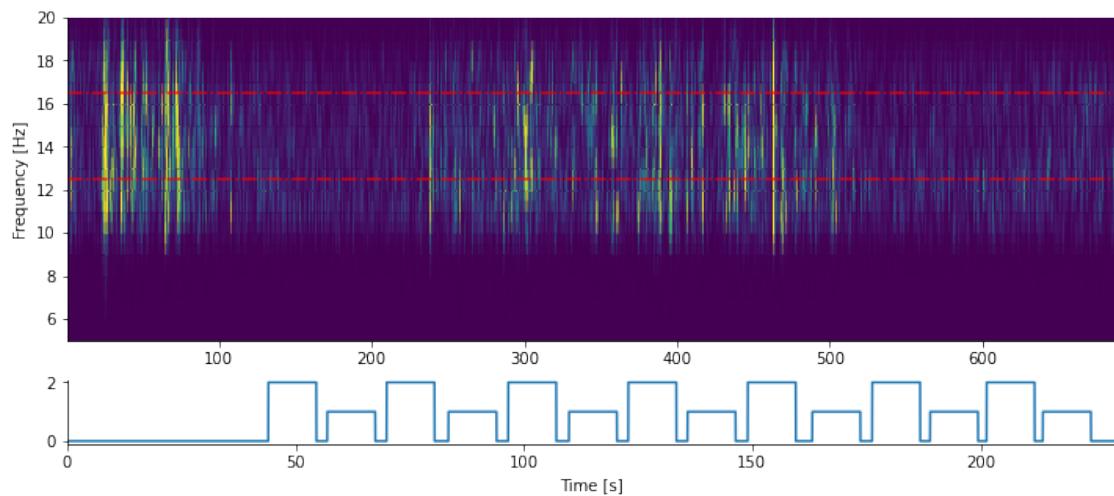
HA1:ch0 filtrado



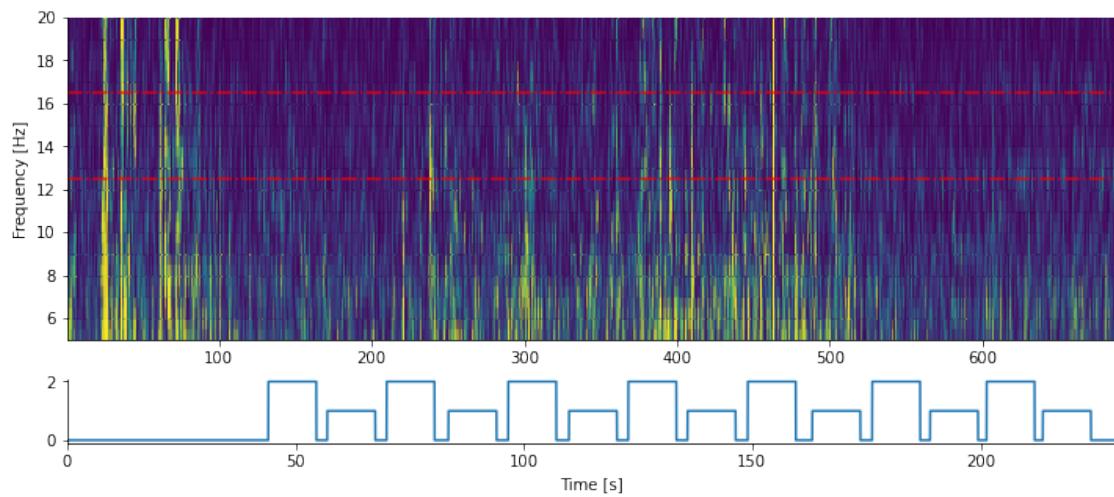
HA1:ch1 sin filtrar



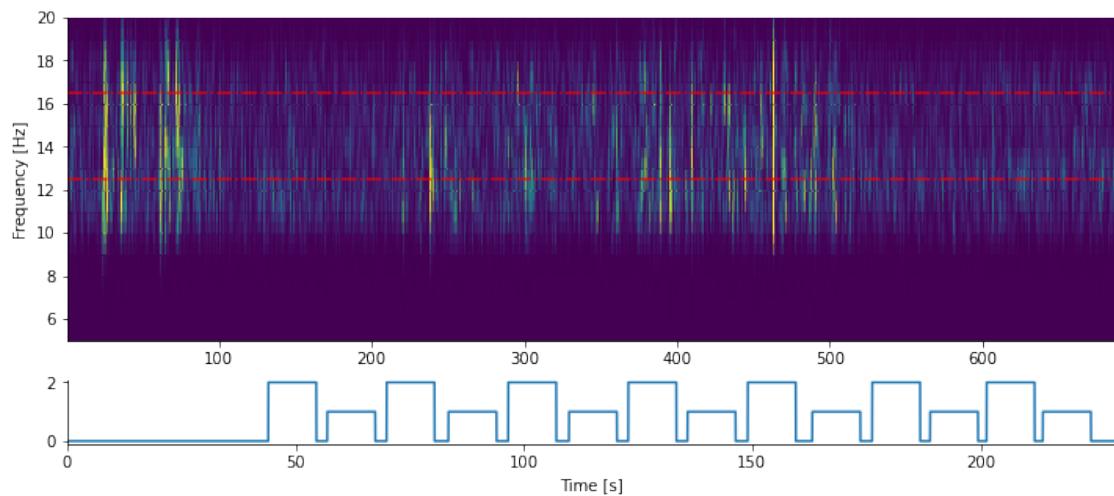
HA1:ch1 filtrado



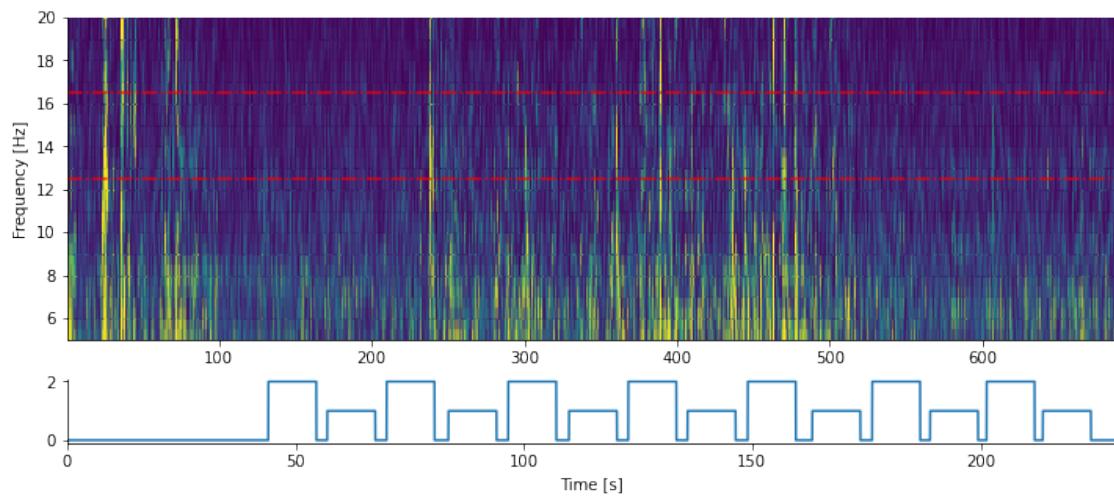
HA1:ch2 sin filtrar



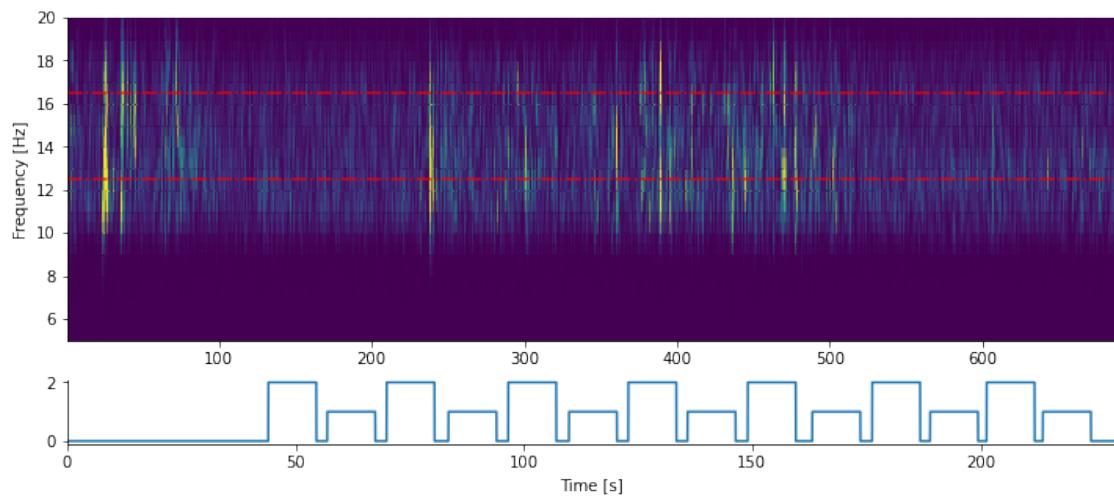
HA1:ch2 filtrado



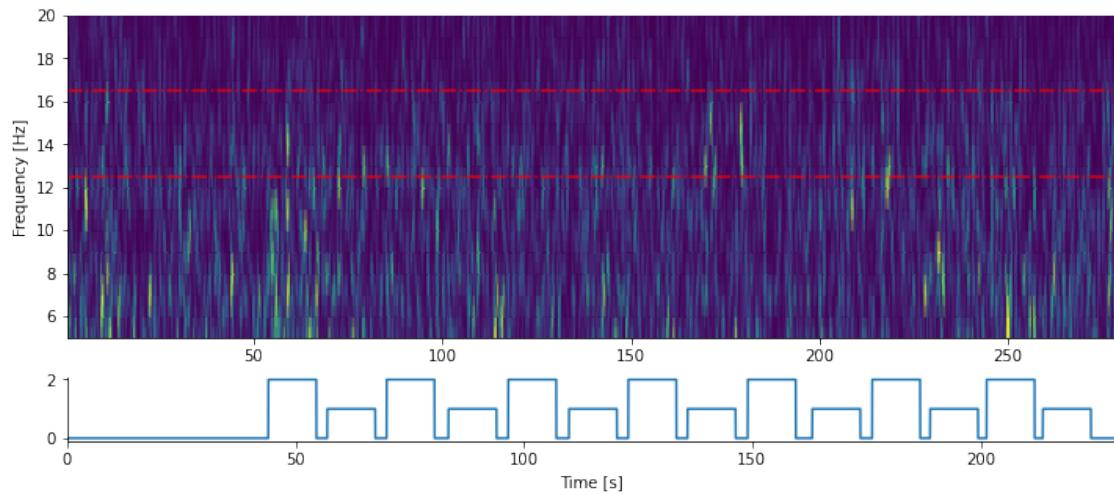
HA1:ch3 sin filtrar



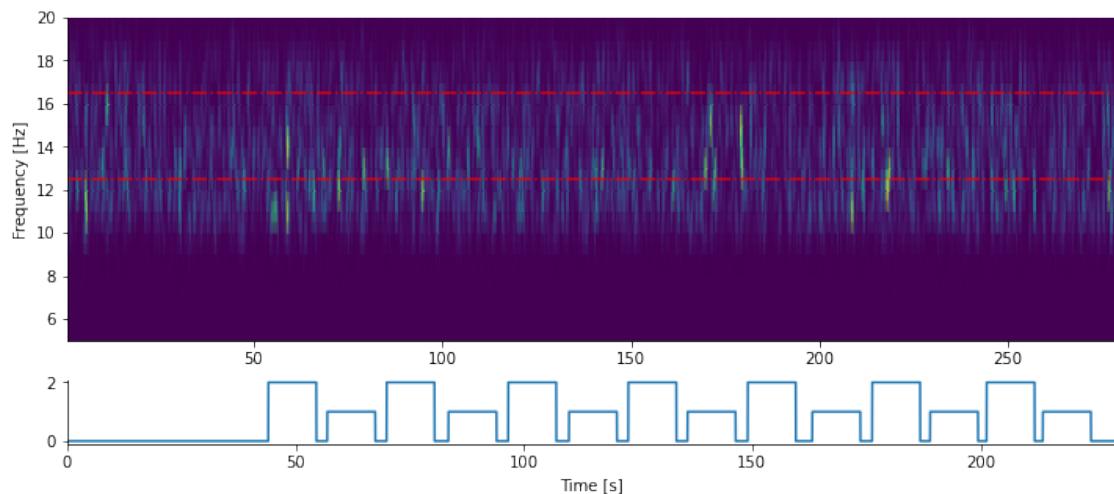
HA1:ch3 filtrado



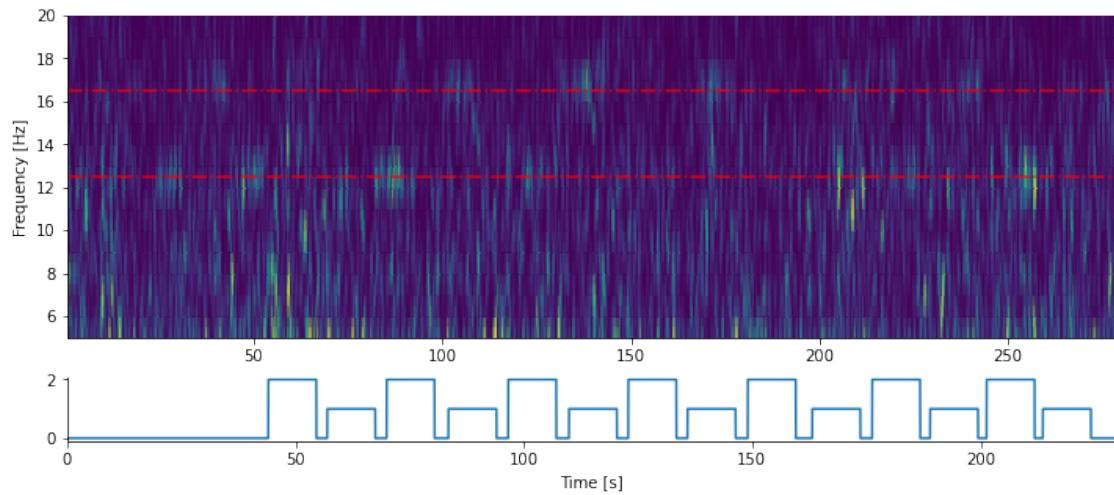
J1:ch0 sin filtrar



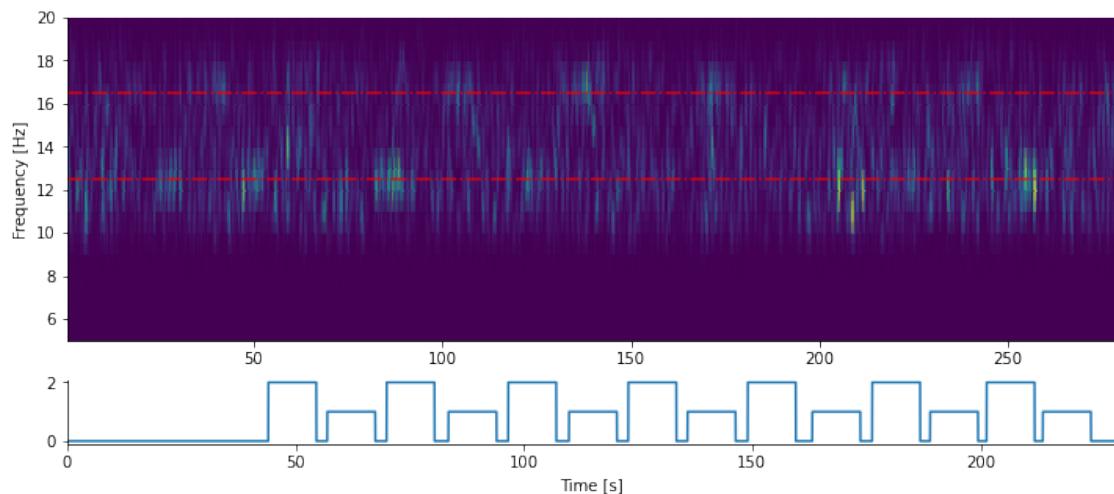
J1:ch0 filtrado



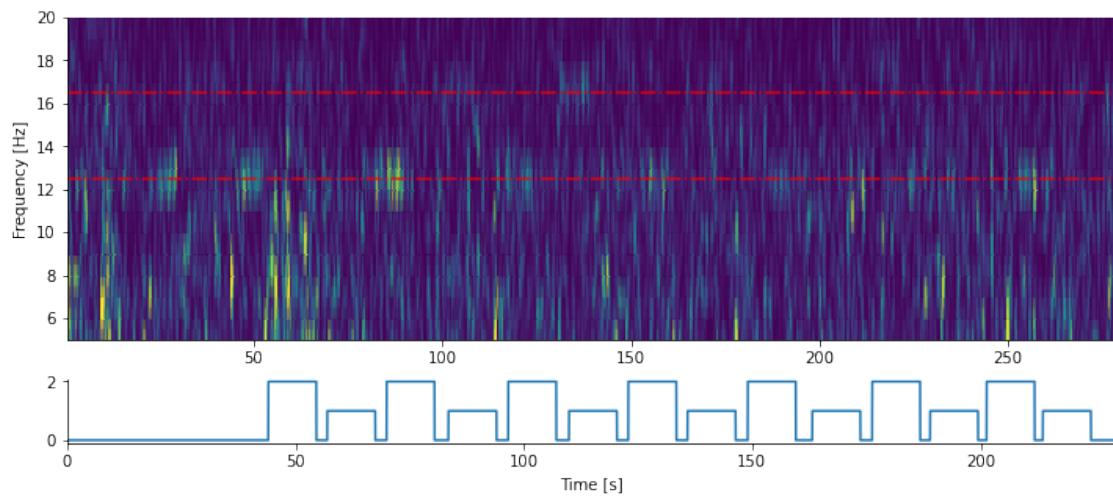
J1:ch1 sin filtrar



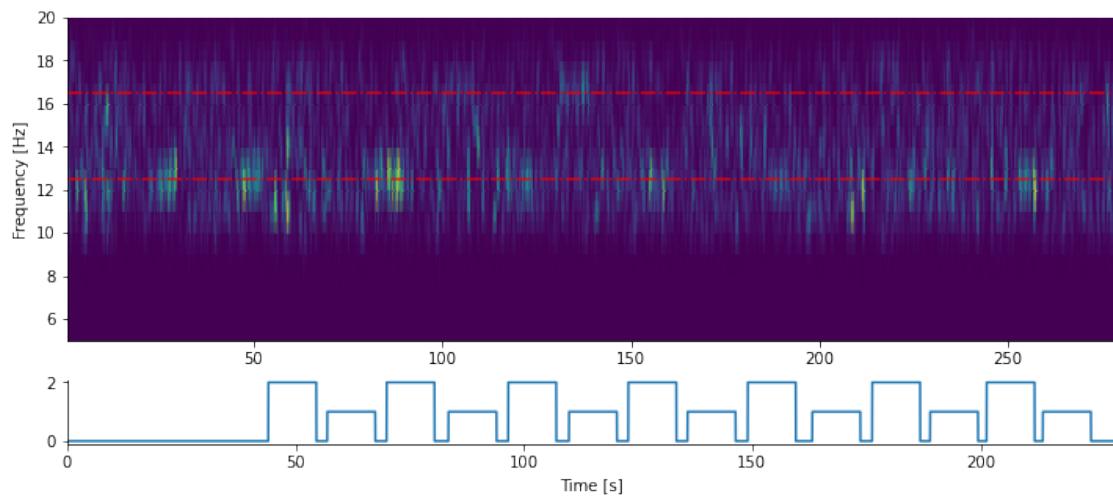
J1:ch1 filtrado



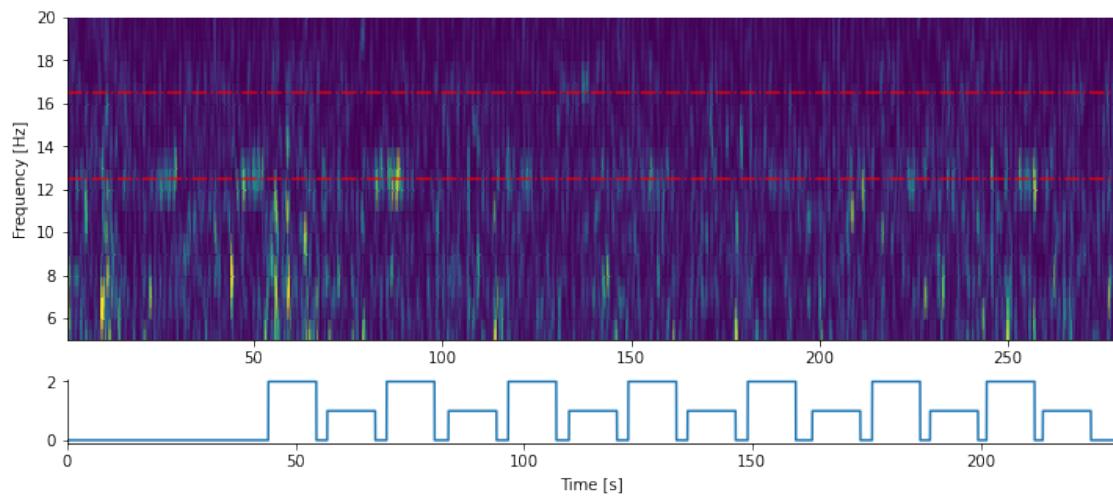
J11:ch2 sin filtrar



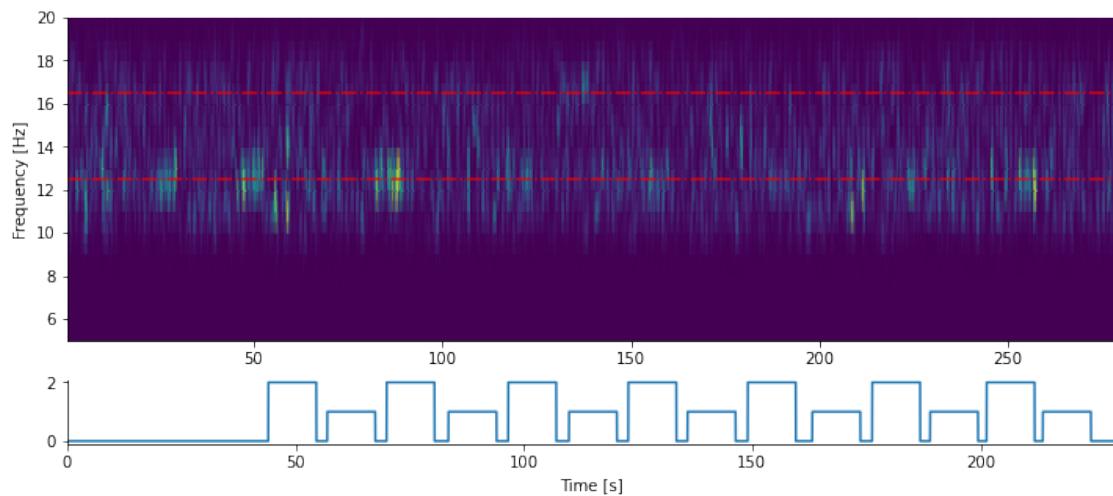
J11:ch2 filtrado



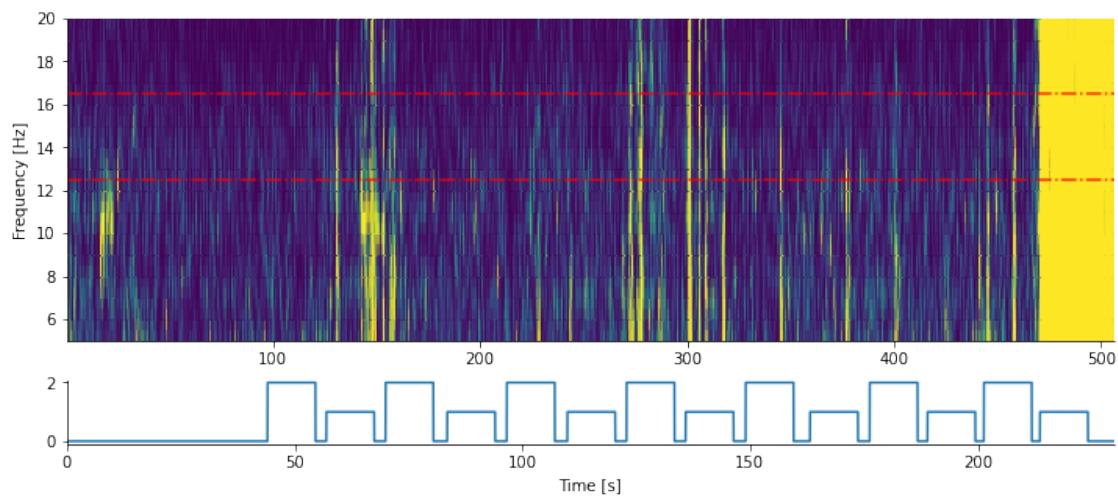
J11:ch3 sin filtrar



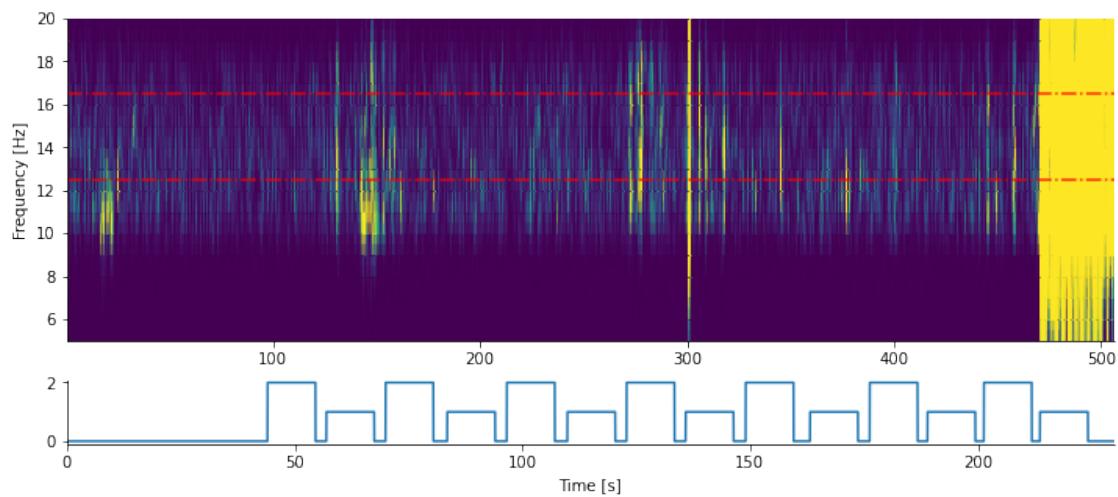
J11:ch3 filtrado



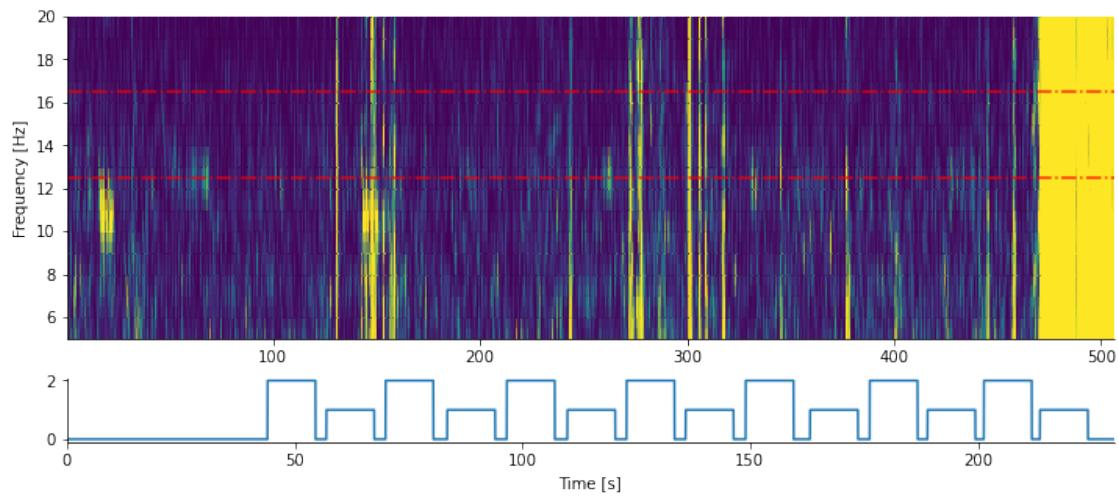
JA2:ch0 sin filtrar



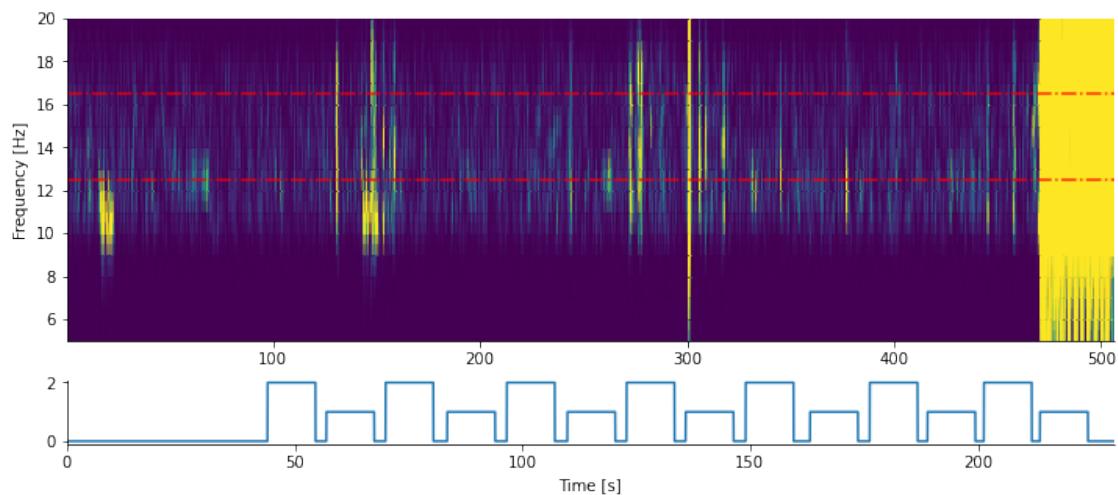
JA2:ch0 filtrado



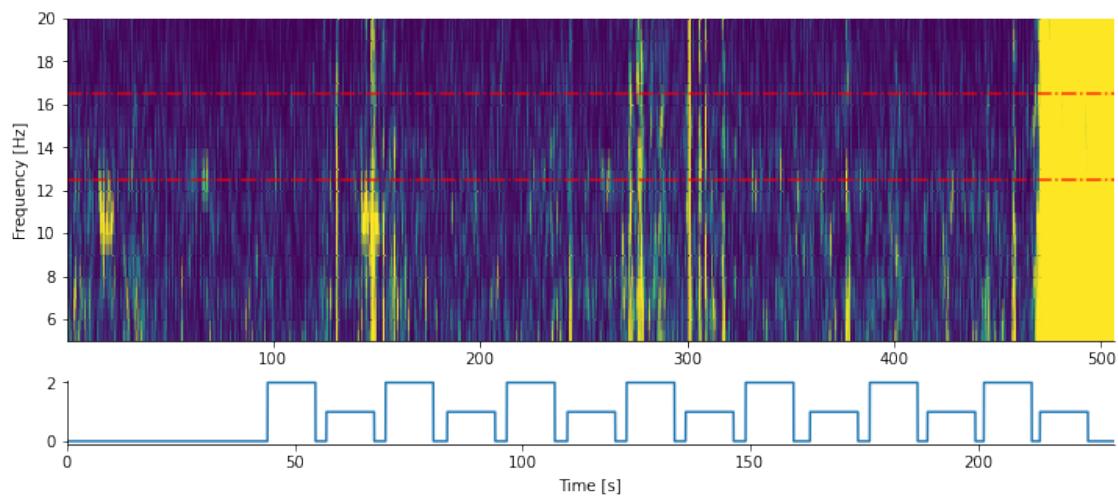
JA2:ch1 sin filtrar



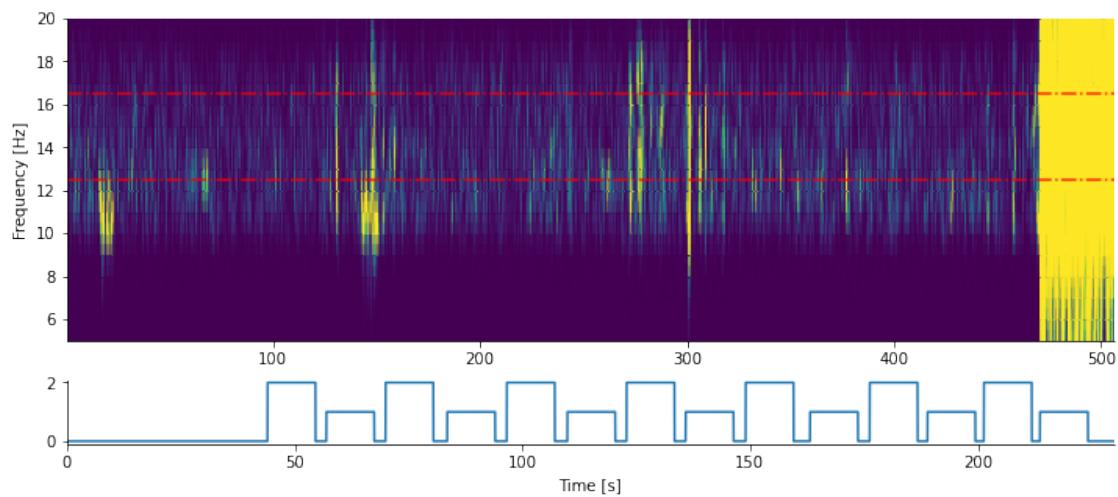
JA2:ch1 filtrado



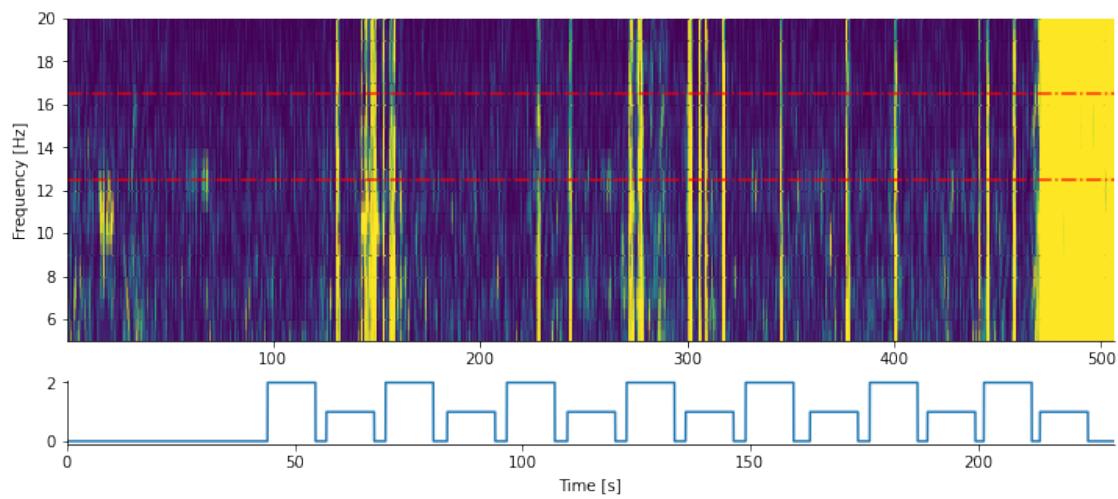
JA2:ch2 sin filtrar



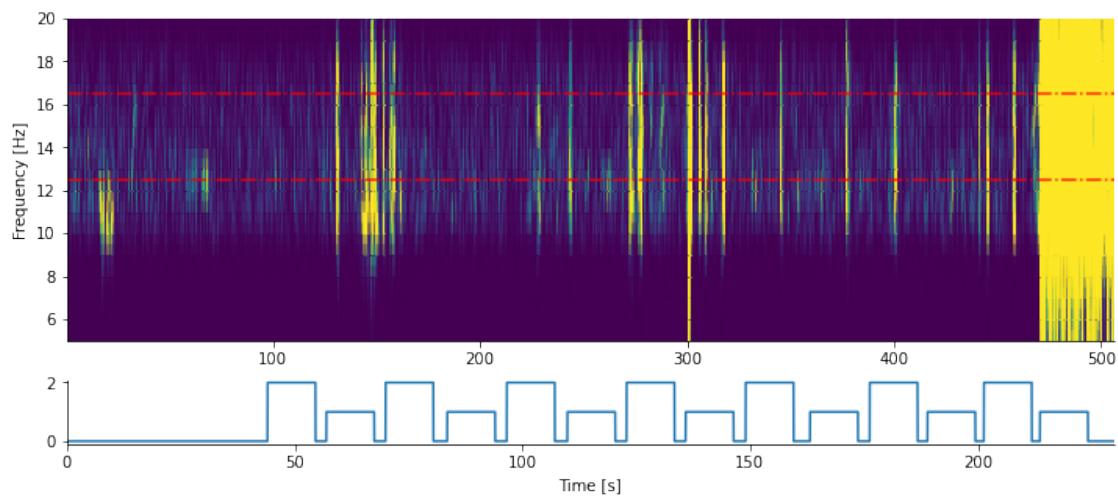
JA2:ch2 filtrado



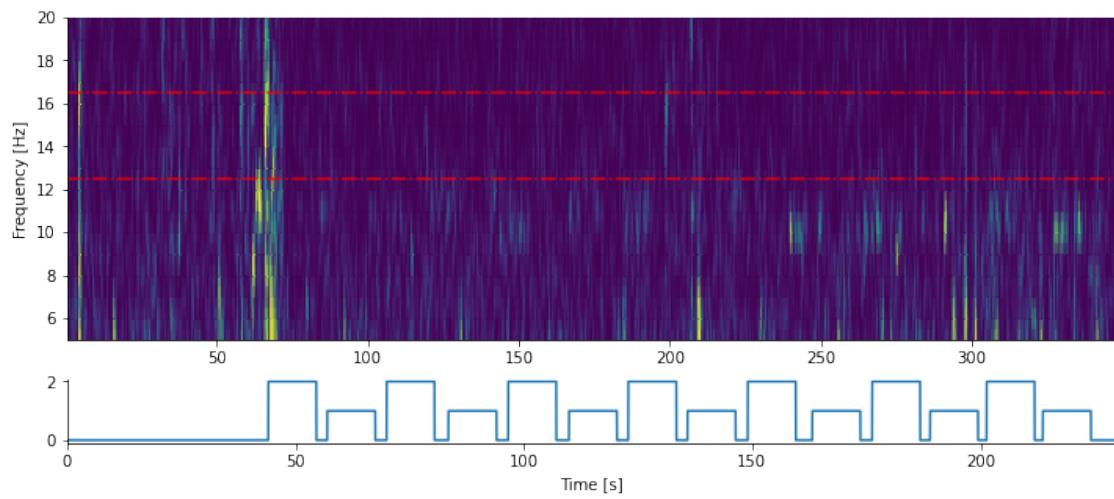
JA2:ch3 sin filtrar



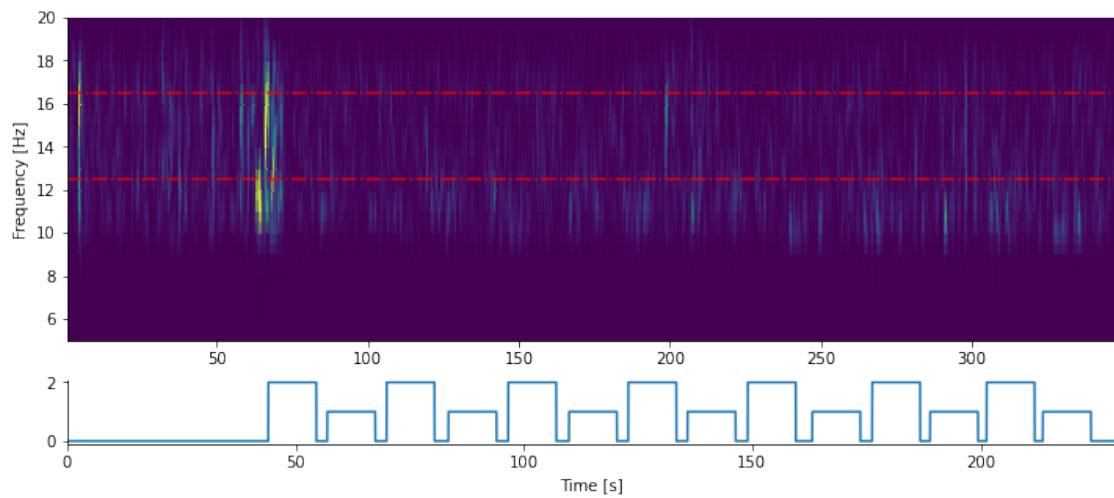
JA2:ch3 filtrado



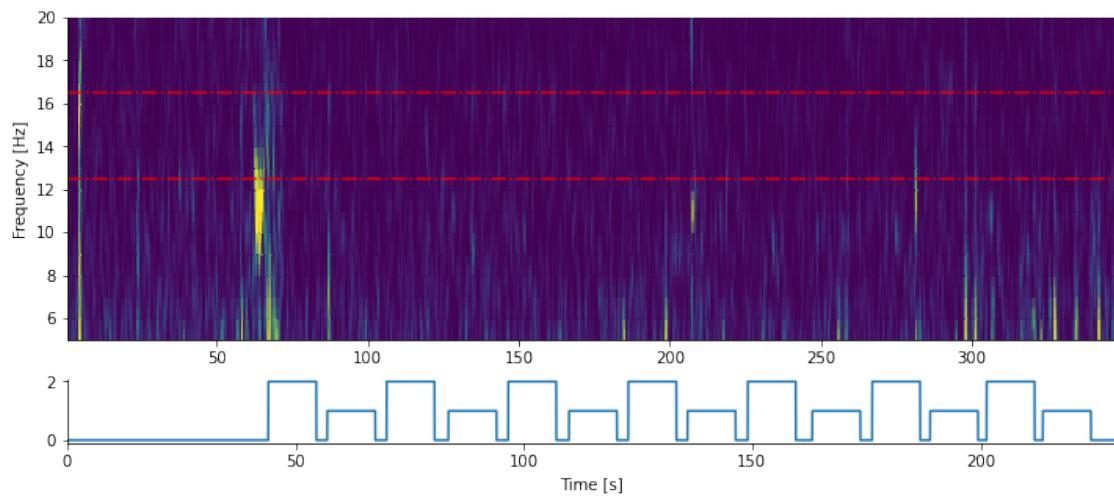
MA1:ch0 sin filtrar



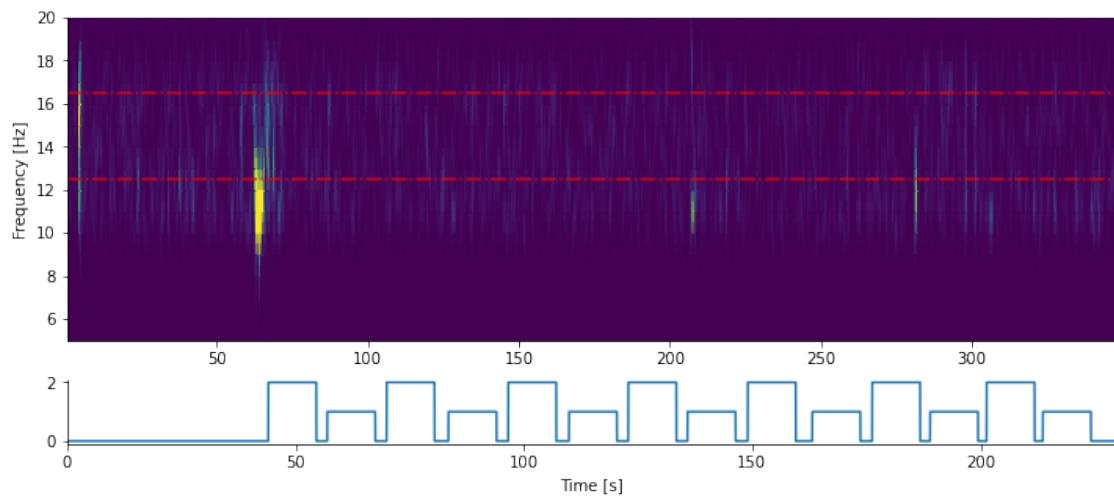
MA1:ch0 filtrado



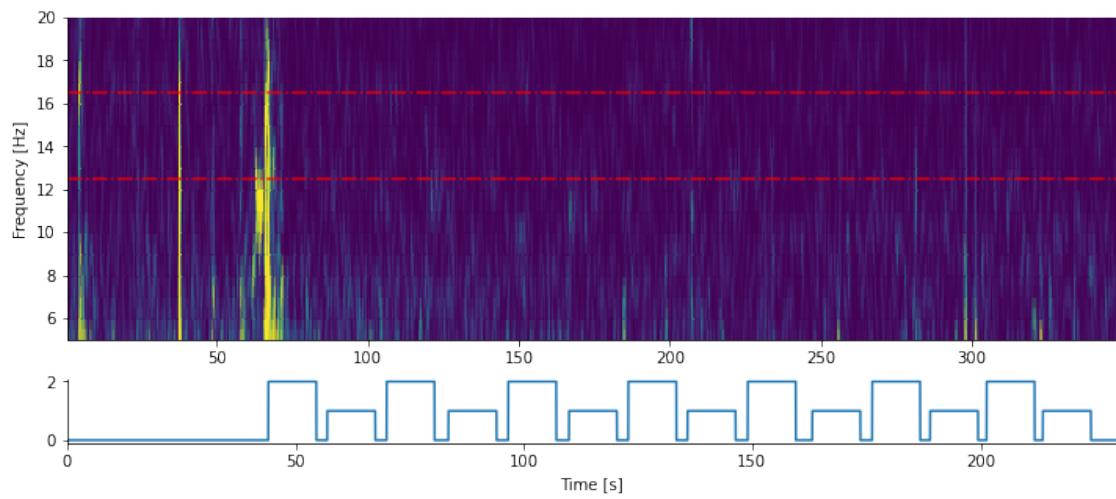
MA1:ch1 sin filtrar



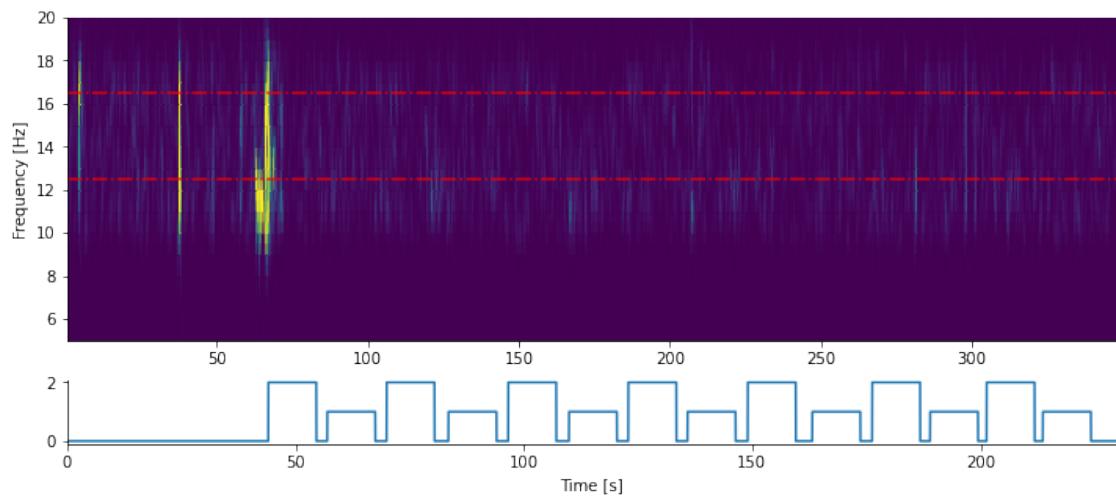
MA1:ch1 filtrado

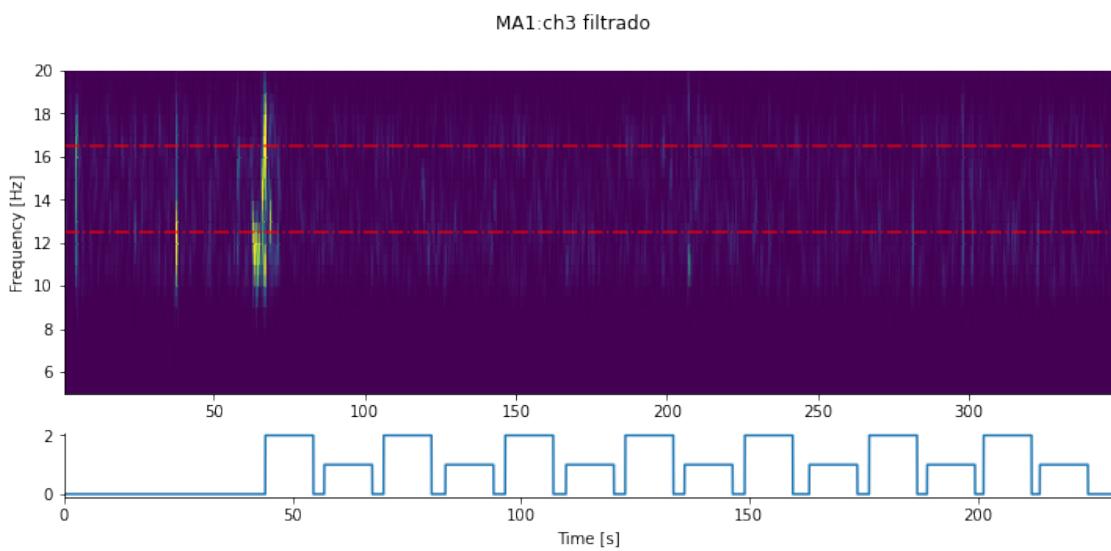
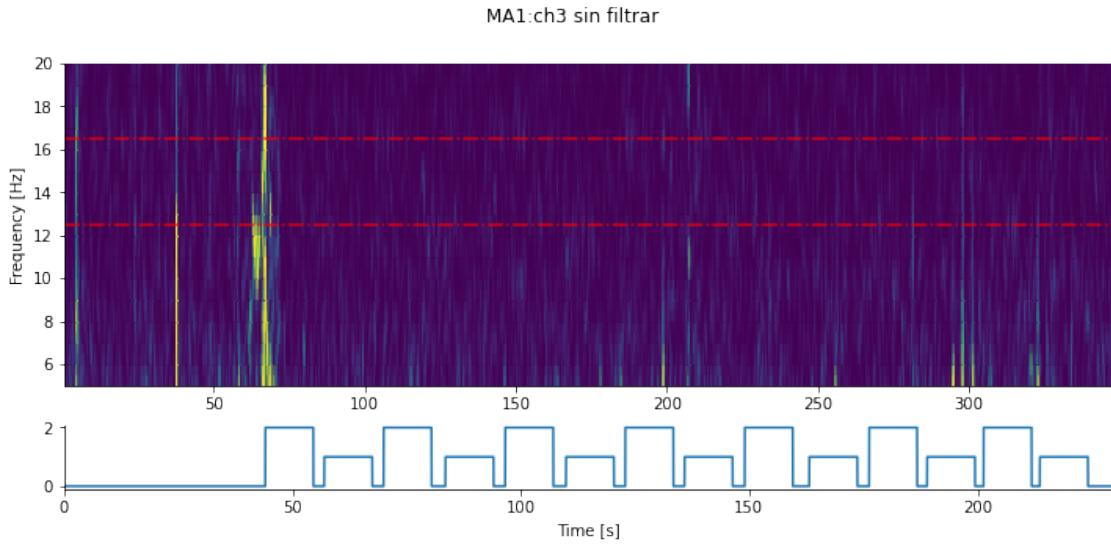


MA1:ch2 sin filtrar



MA1:ch2 filtrado





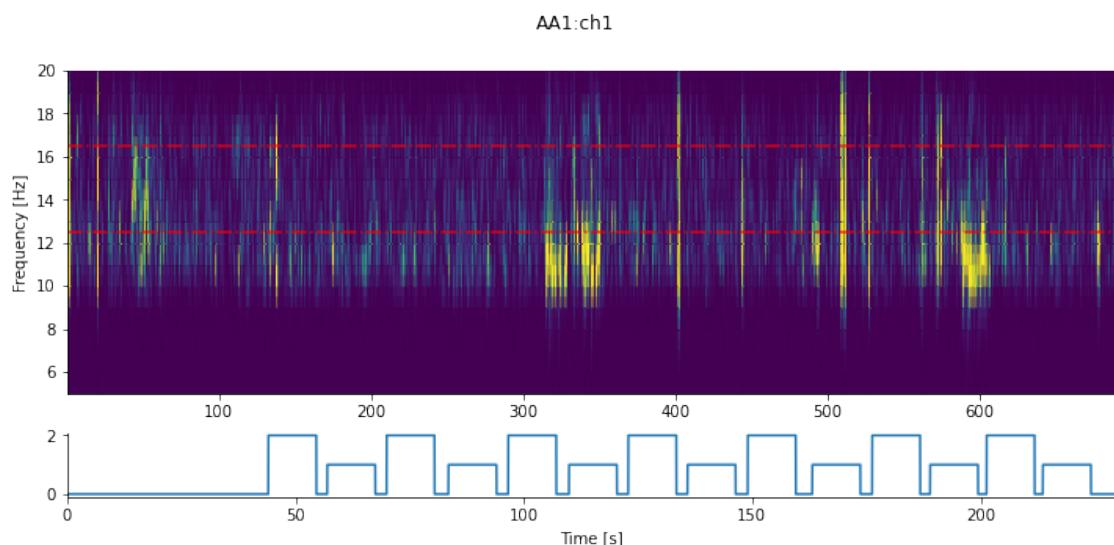
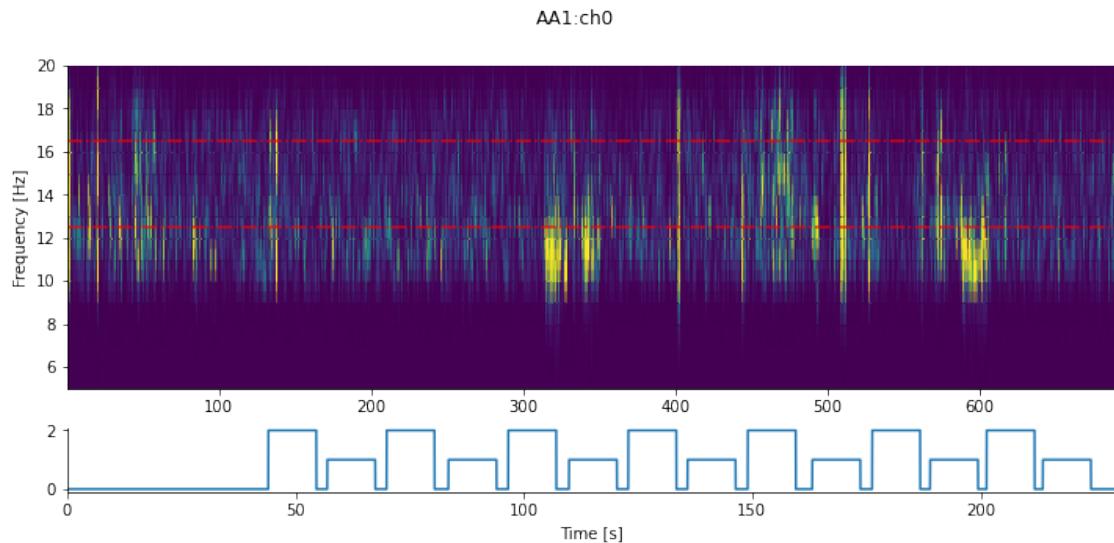
2.6 B) f) Resuma las principales conclusiones de este nivel de análisis.

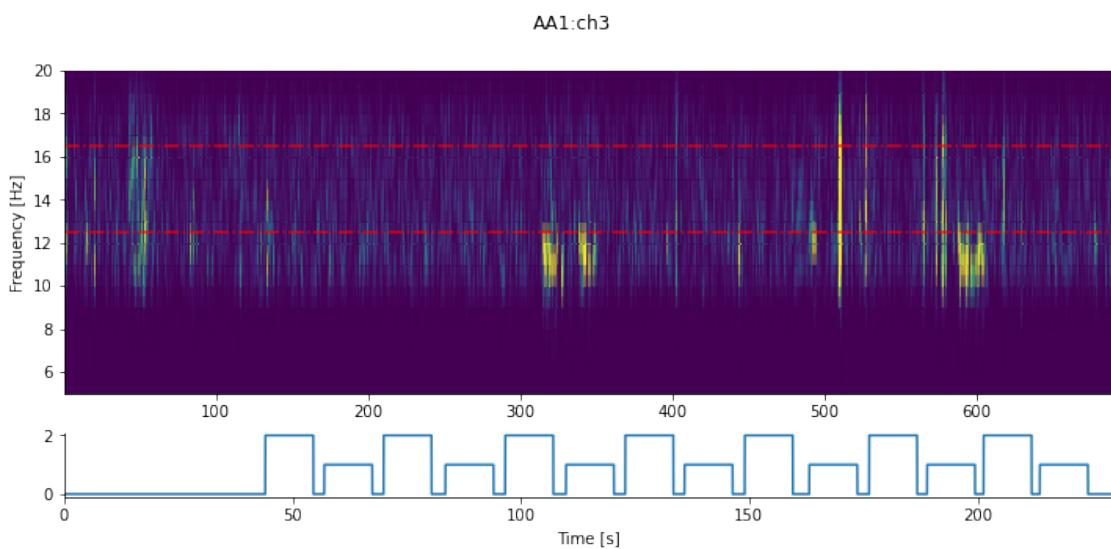
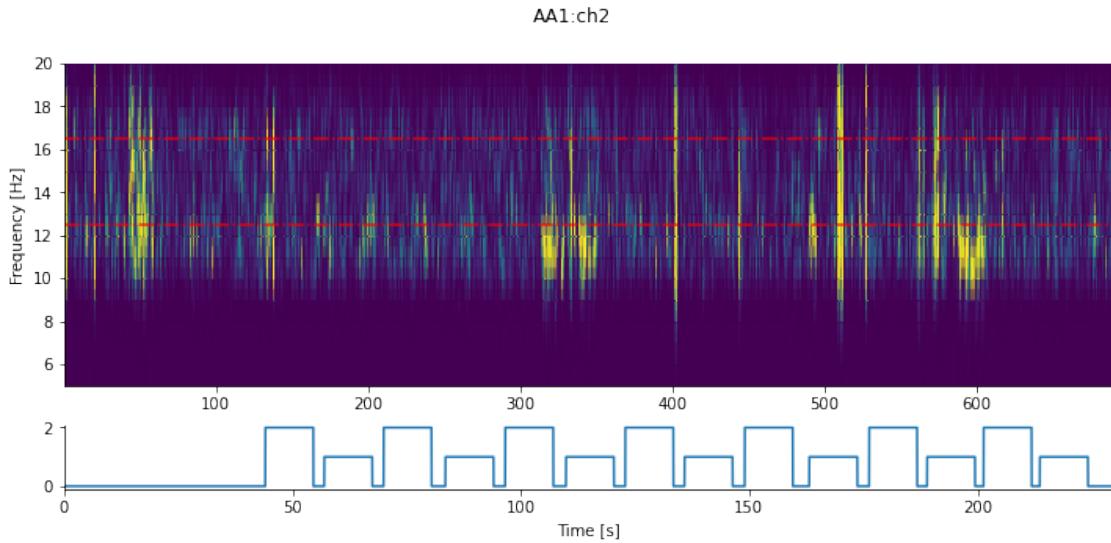
A partir del estudio de las transformadas de Fourier de las señales originales observamos que:

- La frecuencia de mayor presencia coincide con la frecuencia de estimulación a la que se expone al paciente. De esta forma se facilita el estudio, eliminando todos los posibles errores y fuentes de ruido en la señal original.
- Debido a las contribuciones de la red eléctrica y del background del ambiente que contaminan notablemente el espectro resultante, es conveniente realizar un filtrado por frecuencias para quedarnos con el intervalo de frecuencias relevante.

3 C) Nivel Paciente - multi canal: Seleccione los datos correspondientes a un paciente y para ese caso estudie los siguientes elementos:

3.1 C) a) Compute el espectrograma para cada canal, y compárelos. ¿A simple vista, existe alguna correlación?





Se nota correlacion a simple vista.

3.2 C) b) A partir de la respuesta anterior. ¿Considera relevante trabajar con todos los canales disponibles o podría quedarse con un subconjunto?

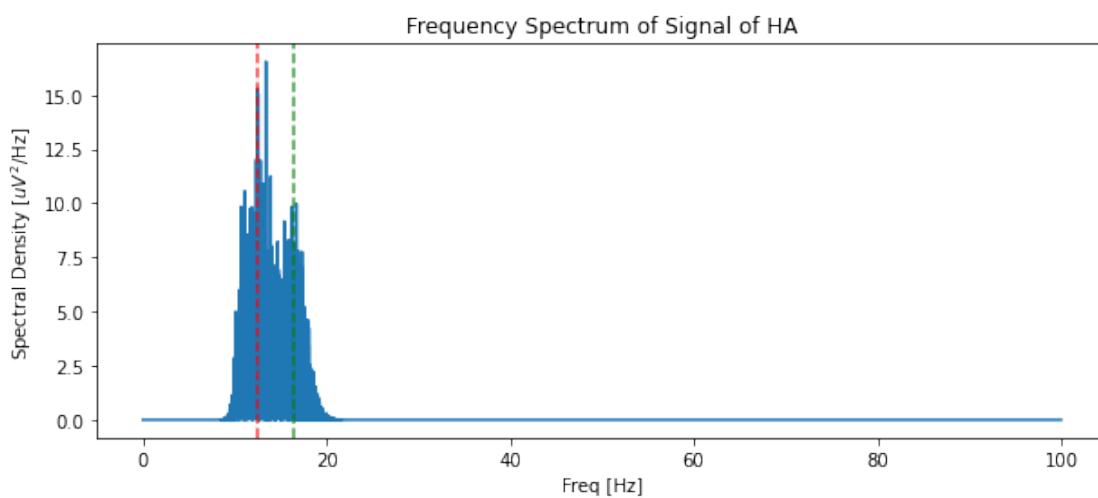
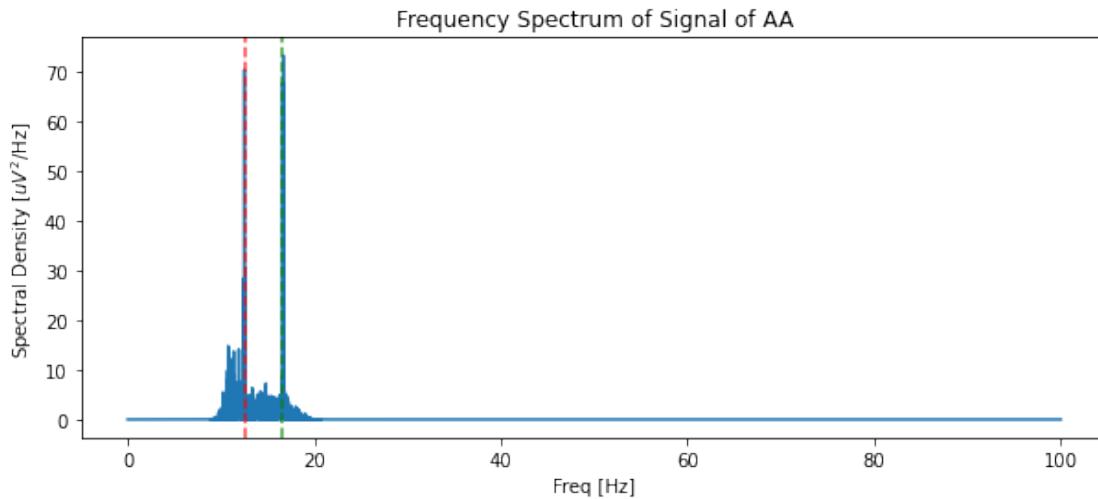
A menos de que existiera algun error en el dataset, la eleccion del canal es irrelevante. Por lo que podemos realizar todo el analisis con solo uno de ellos.

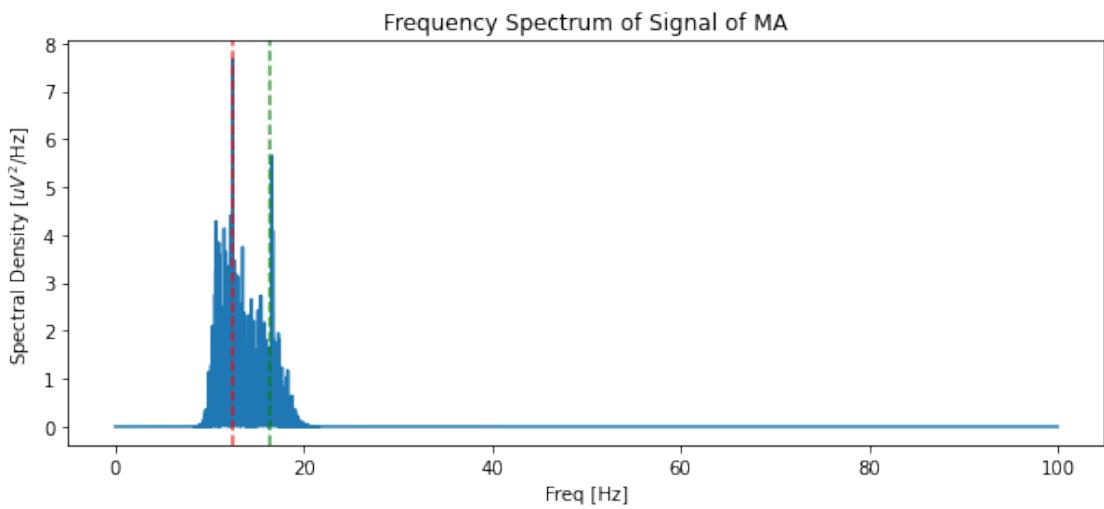
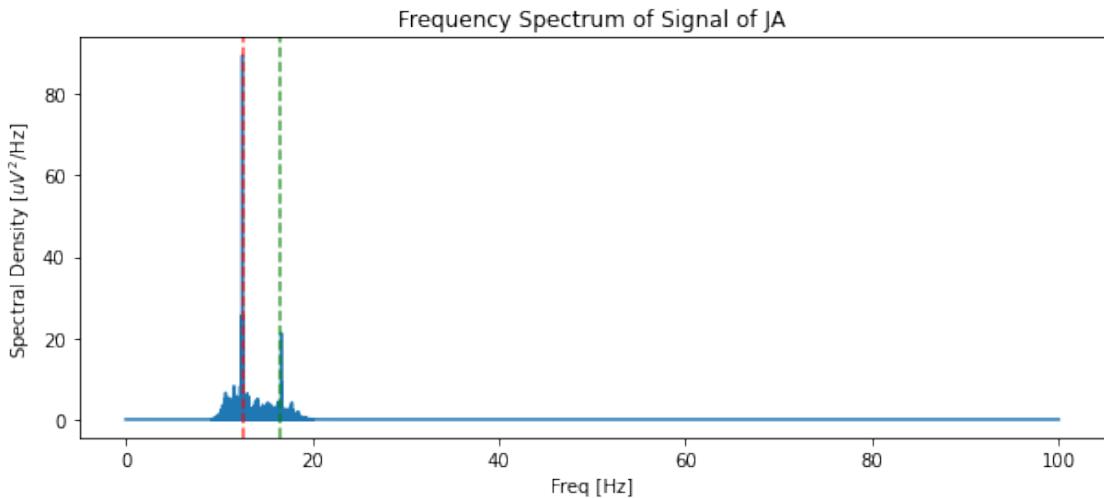
3.3 C) c) Resuma las principales conclusiones de este nivel de análisis.

Como los canales estan correlacionados, podemos realizar todo el analisis con solo uno de ellos.

4 D) Nivel Multi-Paciente.

- 4.1 D) a) A partir de las conclusiones extraídas de los niveles de análisis anteriores. Decida cuáles son los aspectos más importantes a analizar de los registros de un paciente y compárelos entre pacientes. ¿Encuentra diferencias significativas entre pacientes?





El aspecto mas importante para analizar de un paciente es la frecuencia de mayor presencia al hacer la transformada de Fourier sobre la señal original. Una vez analizado esto, existen diferencias entre los pacientes, pero no son relevantes al momento de realizar el análisis.

5 E) Comparando con el Trabajo Práctico anterior.

5.1 E) a) ¿Cambia la complejidad del análisis según el dominio en el que se estudien los datos?

El análisis de los datos a partir de la transformada de Fourier se simplifica considerablemente dado que solo necesitamos encontrar la frecuencia de mayor presencia para saber el estímulo al que está siendo expuesto el paciente.

5.2 E) b) Pensando en un problema de clasificación supervisada, ¿resulta más cómodo trabajar los datos en uno de los dos dominios?

Al momento de realizar una clasificación supervisada, resulta mucho más natural trabajar en el dominio de frecuencia, dado que esta es la naturaleza de las señales neuronales. En este dominio, la información a manejar es mínima y los problemas se resuelven con mayor facilidad.