

G. Fundamentos del procesamiento espectral implementado

INTRODUCCIÓN

Durante el procesamiento de las mediciones espectrales del barrido TVWS (470–698 MHz), se implementaron dos técnicas estadísticas fundamentales:

Coeficiente de correlación de Pearson: Para seleccionar automáticamente el bloque SDR más representativo de cada barrido. Y Umbral adaptativo ($\mu + k \cdot \sigma$): Para detectar automáticamente qué canales están ocupados. Este anexo explica cómo funcionan ambas técnicas y por qué se eligieron los parámetros utilizados.

G.0.1 Coeficiente de Pearson

Al realizar la adquisición espectral con GNU Radio, se generaron vectores que contienen los datos espectrales capturados. Estos se leyeron a través del algoritmo como se muestra a continuación, para más profundidad volver a consultar los (Anexo A) y (Anexo C).

El problema fue que no eran iguales, algunos tenían transitorios, ruido al inicio de la adquisición y al final solo se quería obtener el vector más estable y representativo. Para esto, se comparó cada medición hecha con el SDR con la del analizador de espectro y se seleccionó la más similar.

¿Qué es el coeficiente de Pearson?

El coeficiente de correlación de Pearson (r) mide qué tan similar es la forma de dos curvas, sin importar sus valores absolutos:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

donde:

- **n**: número total de muestras o puntos en cada señal.
- **x_i, y_i**: valores de magnitud (en dBm) de las dos señales en el punto i del eje de frecuencias.
- **\bar{x}** : media aritmética de todos los valores x_i , calculada como $\bar{x} = (1/n) \sum x_i$.
- **\bar{y}** : media aritmética de todos los valores y_i , calculada como $\bar{y} = (1/n) \sum y_i$.
- **r**: coeficiente de correlación resultante, con $-1 \leq r \leq 1$.

Interpretación de los valores de r:

- **$r \approx +1$** : las señales presentaron formas muy similares (picos, valles y pendientes coincidieron en posición y proporción).
- **$r \approx 0$** : no existió relación lineal entre las formas de ambas señales.
- **$r \approx -1$** : las formas resultaron inversas (anticorrelación).

Normalización de las señales

Previo al cálculo de la correlación, ambas señales (analizador y cada vector de GNU Radio) fueron normalizadas mediante la transformación:

$$z_i = \frac{x_i - \bar{x}}{\sigma_x} \quad (2)$$

donde:

- x_i : valor original de magnitud en el punto i .
- \bar{x} : media de la señal original.
- σ_x : desviación estándar de la señal original, calculada como $\sigma_x = \sqrt{[(1/n) \sum (x_i - \bar{x})^2]}$
- z_i : valor normalizado resultante (media = 0, desviación estándar = 1).

La normalización cumplió un papel muy importante ya que eliminó las diferencias de escala y offset, permitiendo que el coeficiente de Pearson evaluara únicamente la similitud geométrica de las curvas, no sus amplitudes absolutas.

El proceso de selección se ejecutó mediante el siguiente algoritmo:

```
mejor_idx = -1
```

```
mejor_corr = -2 # Inicialización fuera del rango [-1, 1]
```

```
for i, bloque in enumerate(datos_txt):
```

```
    bloque_norm = (bloque - np.mean(bloque)) / np.std(bloque)
```

```
    corr, _ = pearsonr(mag_ref_resampled, bloque_norm)
```

```
    if corr > mejor_corr:
```

```
        mejor_corr = corr
```

```
        mejor_idx = i
```

1. Se normalizó el espectro del analizador (referencia) según la **¡Error! No se encuentra el origen de la referencia..**
2. Se normalizó cada vector por medio de la medición SDR con GNU Radio disponible mediante la misma transformación.
3. Se calculó el coeficiente r entre la referencia (Analizador) y GNU Radio usando la **¡Error! No se encuentra el origen de la referencia..**
4. Se seleccionó el bloque que presentó el valor máximo de r como representativo del barrido.

Ventajas del método de correlación

La implementación de este criterio de selección presentó las siguientes ventajas operativas:

- Robustez ante diferencias: No fue afectado por diferencias de ganancia entre instrumentos.
- Comparación basada en forma, no en amplitud: Detectó canales ocupados en las mismas posiciones frecuenciales independientemente del nivel absoluto.
- Automatización completa: Eliminó la necesidad de inspección visual de cada bloque.
- Descarte de artefactos: vectores con transitorios o ruido impulsivo presentaron r bajo y fueron automáticamente rechazados.

G.0.2 Umbral Adaptativo

Una vez seleccionado el vector más representativo de cada barrido, se requirió establecer un criterio objetivo para determinar qué segmentos del espectro correspondían a canales ocupados y cuáles únicamente a ruido de fondo.

Un umbral fijo (por ejemplo, clasificar como ocupado todo lo que superara -80 dBm) presentó limitaciones importantes:

- En zonas con ruido elevado, generó falsas alarmas que es ruido clasificado erróneamente como señal.
- En zonas con ruido bajo, produjo detecciones perdidas que son señales débiles reales no detectadas.

Se implementó entonces un umbral adaptativo que se ajustó automáticamente al nivel de ruido observado en cada medición específica.

Interpretación matemática del umbral

El umbral de detección λ se calculó mediante la Ecuación 1.

$$\text{umbral } \lambda = \mu_{\text{ruido}} + k\sigma_{\text{ruido}} \quad (5)$$

donde:

- λ : umbral adaptativo de detección (en dBm).
- μ_{ruido} : media aritmética del piso de ruido, calculada como el promedio de las magnitudes más bajas observadas.
- σ_{ruido} : desviación estándar del ruido, que cuantificó la dispersión de las muestras de ruido alrededor de su media.
- $k = 2.5$: factor de confianza estadística.

La estimación de μ_{ruido} y σ_{ruido} se realizó utilizando únicamente el 20% de las muestras de menor magnitud, como se puede observar en esta porción del algoritmo, y también teniendo en cuenta la Tabla 1:

```
n_ruido = int(len(magnitud_suav) * porcentaje_ruido / 100) # 20%
```

```
indices_ruido = np.argsort(magnitud_suav)[:n_ruido]
```

```
piso_ruido = np.mean(magnitud_suav[indices_ruido])
```

```
sigma_ruido = np.std(magnitud_suav[indices_ruido])
```

1. Se ordenaron todas las muestras del espectro de menor a mayor magnitud.
2. Se seleccionaron únicamente las muestras correspondientes al 20% inferior.
3. Se calculó la media (μ_{ruido}) y desviación estándar (σ_{ruido}) exclusivamente sobre este subconjunto.

En la banda TVWS, la mayor parte de las frecuencias correspondieron a canales libres (solo ruido), mientras que una fracción menor presentó canales activos (TV analógica, TV digital, LTE). Al seleccionar el 20% más bajo, se garantizó que la estimación del piso de ruido no estuviera contaminada por las señales reales presentes en el espectro.

Percentil	Características	Limitaciones
10%	Muy robusto ante presencia de señales; ignora prácticamente todas las emisiones activas	Puede subestimar σ si el ruido presenta alta variabilidad espacial

Percentil	Características	Limitaciones
20%	Balance óptimo: suficientes muestras para estabilidad estadística sin incluir señales	— (seleccionado)
30%	Mayor número de muestras, mejor estimación de σ en entornos uniformes	Riesgo de incluir señales débiles, elevando artificialmente μ y σ
50% (mediana)	Representativo si el espectro está mayormente desocupado	Falla cuando >50% del espectro está ocupado; umbral resultante es excesivamente alto

Tabla 1. Comparación de diferentes percentiles para estimación del piso de ruido

Selección del factor de confianza $k = 2.5$

El factor k controló el compromiso entre sensibilidad (capacidad de detectar señales débiles) y especificidad (probabilidad de no clasificar ruido como señal). Bajo el supuesto de que el ruido siguió una distribución gaussiana [2] cada valor de k se asoció a una probabilidad de falsa alarma (Pfa), definida como la probabilidad de que una muestra de ruido puro supere el umbral por fluctuación aleatoria (Gorcin & Arslan, 2010).

k	Percentil del ruido bajo el umbral	Pfa (falsa alarma)	Nivel de confianza	Aplicabilidad
1.0	84.13%	15.87%	84.13%	Excesivamente sensible; genera muchas falsas alarmas
2.0	97.72%	2.28%	97.72%	Moderadamente sensible; apropiado para entornos con señales fuertes
2.5	98.76%	1.24%	98.76%	Balance óptimo entre sensibilidad y especificidad
3.0	99.87%	0.135%	99.87%	Conservador; puede perder canales con potencias moderadas
4.0	99.997%	0.0032%	99.997%	Muy conservador; solo detecta señales muy fuertes

Tabla 2. Relación entre factor k y probabilidad de falsa alarma bajo modelo gaussiano

Interpretación para $k = 2.5$:

Con este valor, solo el 1.24% de las muestras de ruido puro superaron el umbral. En otras palabras, existió un 98.76% de confianza de que una muestra por encima del umbral correspondió a una señal real y no a una fluctuación aleatoria del ruido.

Justificación de la selección:

El valor $k = 2.5$ fue seleccionado porque:

1. Suficiente sensibilidad: Permitió detectar canales con potencias moderadas típicas del entorno de medición.
2. Baja tasa de falsas alarmas: Solo 1-2% del ruido fue clasificado erróneamente como ocupación.
3. Consistencia con estándares: Los límites de detección con instrumentación científica utilizan típicamente factores entre 3σ y 3.3σ [1]; el uso de 2.5σ representó una configuración ligeramente más sensible, apropiada para la detección de canales débiles en espectros de radiofrecuencia.

Fundamento teórico: distribución gaussiana del ruido

El método del umbral adaptativo se fundamentó en una distribución normal (gaussiana).

La función de densidad de probabilidad del ruido fue:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3)$$

donde:

x : valor de magnitud observado (variable aleatoria, en dBm).

μ : media de la distribución (piso de ruido promedio).

σ : desviación estándar (dispersión del ruido).

$p(x)$: probabilidad de observar el valor x .

Aproximadamente el 68.27% de las observaciones cayeron dentro del intervalo $[\mu - \sigma, \mu + \sigma]$

Aproximadamente el 95.45% cayeron dentro de $[\mu - 2\sigma, \mu + 2\sigma]$

Aproximadamente el 99.73% cayeron dentro de $[\mu - 3\sigma, \mu + 3\sigma]$

Para $k = 2.5$, el 98.76% de las muestras de ruido quedaron por debajo del umbral $\lambda = \mu + 2.5\sigma$ [2].

Si una muestra superó el valor $\mu + 2.5\sigma$, la probabilidad de que correspondiera únicamente a ruido fue inferior al 1.24%, permitiendo clasificarla con alta confianza como parte de un canal ocupado.

El umbral de detección se estableció en: $-89.45 + (2.5 \times 2.31) = -83.68$ dBm

Se identificaron dos canales con magnitudes significativamente superiores a este umbral, correspondientes a canales UHF.

Ventajas del método adaptativo

La implementación del umbral adaptativo presentó las siguientes ventajas frente a métodos de umbral fijo:

- Adaptación automática: El umbral se recalculó para cada medición según las condiciones reales del ruido
- Independencia de la banda de frecuencia: No requirió recalibración manual al cambiar de frecuencia central
- Robustez ante cambios de ganancia: Si se modificó la configuración del receptor, el método continuó funcionando correctamente
- Balance optimizado: El factor $k = 2.5$ minimizó tanto las falsas alarmas como las detecciones perdidas.

[1] AcademiaLab. (s.f.). Límite de detección. AcademiaLab. <https://academia-lab.com/enciclopedia/limite-de-deteccion/>

[2] Observatoire de Paris – PSL. (s.f.). El ruido gaussiano. Aprender. <https://media4.obspm.fr/public/VAU/instrumentacion/observer/analizar/ruido-gaussiano/APPRENDRE.html>