



ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIÓN

Titulación:

INGENIERO DE TELECOMUNICACIÓN

Título del proyecto:

“MODELADO NO ESTACIONARIO DE LA ENERGÍA
ARMÓNICA Y TRANSITORIA EN SEÑALES
MUSICALES”

Federico Ruiz Esquiroz

Tutor: Miroslav Zivanovic

Pamplona, 16 de Septiembre de 2010

Índice

<u>1. Introducción</u>	<u>2</u>
1.1. Clasificación de picos	6
1.2. Modelado no estacionario	7
<u>2. Antecedentes, Objetivos, Motivación y Aplicaciones</u>	<u>12</u>
<u>3. Descripción de las tramas</u>	<u>14</u>
3.1. Trama 1	15
3.2. Trama 2	16
3.3. Trama 3	17
3.4. Trama 4	18
3.5. Trama 5	20
<u>4. Estudio del Overfit en el Modelado no estacionario</u>	<u>22</u>
<u>4.1. Estudio del Overfit en el Modelado Polinomial</u>	<u>22</u>
4.1.1. Modelado Polinomial Trama 1	23
4.1.2. Modelado Polinomial Trama 2	35
4.1.3. Modelado Polinomial Trama 3	46
4.1.4. Modelado Polinomial Trama 4	48
4.1.5. Modelado Polinomial Trama 5	58
<u>4.2. Preprocesado de la trama mediante Clasificación de picos espectrales y posterior modelado polinomial</u>	<u>65</u>
4.2.1. Preprocesado y modelado polinomial Trama 5	65
4.2.2. Preprocesado y modelado polinomial Trama 4	69
4.2.3. Preprocesado y modelado polinomial Trama 2	72
4.2.4. Preprocesado anulando los picos ruidosos y modelado polinomial Trama 2	74
4.2.5. Conclusiones al preprocesado mediante Clasificación de picos y posterior modelado	75
<u>4.3. Estudio del Overfit en el modelado lineal por tramos</u>	<u>76</u>
4.3.1. Modelado lineal por tramos Trama 3	77
4.3.2. Modelado lineal por tramos Trama 5	84
<u>4.4. Conclusiones al estudio del Overfit en el modelado no estacionario</u>	<u>86</u>

<u>5. Estudio y búsqueda de un baremo del grado de modulación de una trama</u>	<u>87</u>
<u>5.1. Método 1: Tiempo medio y Duración temporal</u>	<u>88</u>
5.1.1. Descripción	88
5.1.2. Grado de modulación según los valores del Tiempo medio y la Duración temporal	90
5.1.3. Estimación de unos umbrales de decisión	91
5.1.4. Conclusiones. Umbrales de decisión estimados	100
<u>5.2. Método 2: Descriptor NBD usado en Clasificación de picos</u>	<u>101</u>
5.2.1. Descripción	101
5.2.2. Estudio del grado de modulación según los valores NBD	102
5.2.3. Conclusiones al estudio del grado de modulación según los valores NBD	119
<u>5.3. Método 3: Sparsity de la DWT y la MDCT</u>	<u>120</u>
5.3.1. Descripción	120
5.3.2. Prueba del método	122
5.3.3. Conclusiones al método basado en el Sparsity	129
<u>5.4. Conclusiones a la búsqueda de un método de baremo</u>	<u>129</u>
<u>6. Búsqueda de un método para estimar el impacto del ruido en una trama</u>	<u>130</u>
6.1. Prueba del método y estimación de un Umbral	131
<u>7. Implementación del modelado optimizado y estudio comparativo</u>	<u>132</u>
7.1. Estudio Comparativo	133
<u>8. Conclusiones Finales</u>	<u>141</u>
<u>9. Líneas futuras</u>	<u>142</u>
<u>10. Bibliografía</u>	<u>143</u>

1. Introducción

Modelo de la señal musical:

El sonido producido por instrumentos musicales o por cualquier tipo de fuente física armónica se puede modelar como la suma entre una serie de sinusoides y un ruido residual. La parte sinusoidal o determinista corresponde normalmente a los principales modos de vibración del sistema, mientras que la parte residual recoge la energía producida por el mecanismo de excitación que no es transformada por el sistema en vibraciones estacionarias, además de otros tipos de componentes energéticas de naturaleza no sinusoidal. Por ejemplo, en el sonido de los instrumentos de viento, la señal determinista es resultado de las oscilaciones producidas en la cavidad, y la parte residual es una señal ruidosa generada por la corriente turbulenta de aire introducida por el músico a través de una estrecha apertura. En instrumentos de arco y cuerda, las sinusoides estables son resultado de los modos principales de vibración de las cuerdas, mientras la parte residual es producida por el deslizamiento del arco contra las cuerdas, además de otros comportamientos no lineales del sistema resonador arco-cuerda. La misma separación puede ser aplicada a sonidos vocales, instrumentos de percusión e incluso sonidos no musicales encontrados en la naturaleza [1].

En el presente proyecto se llamará señal determinista a la suma de componentes cuasi sinusoidales, o sinusoides cuya amplitud y frecuencia varía lentamente a través del tiempo. Cada senoide modela una estrecha banda del sonido original y se describe mediante una función de amplitud y una función de frecuencia.

A la componente estocástica también se le suele llamar ruido convolutivo, por ser un ruido producido por el propio sistema de excitación, aunque puede incluirse en dicha componente el ruido aditivo de fondo existente en grabaciones de baja calidad. Este ruido queda totalmente caracterizado por su densidad espectral de potencia, que proporciona la potencia esperada para cada banda frecuencial. Cuando una señal es considerada estocástica, no es necesario conservar información sobre su fase instantánea o su magnitud exacta para cada FFT individual.

Por lo tanto el sonido de entrada $s(t)$ se puede modelar como:

$$s(t) = \sum_{r=1}^R A_r(t) \cdot \cos[\theta_r(t)] + e(t) \quad (\text{Ecuación 1.1})$$

donde $A_r(t)$ y $\theta_r(t)$ son la amplitud y fase instantáneas de la r -ésima senoide, respectivamente, y $e(t)$ es la componente de ruido en el tiempo t .

El modelo asume que **las sinusoides, también llamadas parciales, son componentes estables del sonido y poseen una amplitud y frecuencia que varían lentamente en comparación con la propia frecuencia de la señal.** La fase instantánea se puede tomar por tanto como la integral de la frecuencia instantánea, verificando:

$$\theta_r(t) = \int_0^t w_r(\tau) d\tau \quad (\text{Ecuación 1.2})$$

donde $w(t)$ es la frecuencia en radianes y r es el número de senoide.

Asumiendo que $e(t)$ es una señal estocástica, se puede describir como el filtrado de un ruido blanco:

$$e(t) = \int_0^t h(t, \tau) u(\tau) d\tau \quad (\text{Ecuación 1.3})$$

donde $u(t)$ es un ruido blanco y $h(t, \tau)$ es la respuesta impulsional de un filtro variante en el tiempo en el instante t . Es decir, el residuo es modelado por la convolución de ruido blanco con un filtro variante en el tiempo que da forma al espectro del mismo.

El modelo encuentra dificultades en el caso de “parciales ruidosos”, situados a medio camino entre la parte determinista y la estocástica de la señal, encontrados en determinados tipos de sonido. La separación entre ambos no es, pues, clara, y es necesario dotar al proceso de cierta flexibilidad.

Enfoque utilizado:

Existen tres tipos básicos de enfoques ó modelos con los que estudiar el sonido producido por cualquier sistema físico:

- **Modelos instrumentales:** intentan proponer un modelo físico de la misma fuente que produce el sonido, ya sea un instrumento como un violín, un clarinete, o el propio tracto vocal.
- **Modelos espectrales:** intentan parametrizar el sonido tal y como es percibido por el oído, descartando cualquier información que para el oído no sea importante.
- **Modelos abstractos:** tratan de proporcionar parámetros musicalmente útiles mediante la obtención de una fórmula abstracta, sin que esta tenga una interpretación directa en términos físicos o perceptuales.

El presente proyecto esta basado en el segundo tipo, **modelos espectrales.**

La principal ventaja de trabajar con el modelo espectral es la existencia de diferentes procedimientos de análisis capaces de extraer de sonidos reales los parámetros de síntesis, teniendo por tanto la capacidad de sintetizar dichos sonidos, modificarlos y analizarlos en profundidad. **Estas técnicas están basadas en el análisis de Fourier, mediante la FFT (Fast Fourier Transform) y la STFT (Short Time Fourier Transform).**

Modulación en la señal musical:

El sonido musical no es una señal estacionaria en el tiempo, sino que presenta variaciones instantáneas en la frecuencia y en la potencia asociadas a cada componente frecuencial del mismo, que pueden ser más o menos pronunciadas, ligadas al grado de modulación en frecuencia y en amplitud respectivamente. La fase, aunque también registra variaciones, tiene un papel menos relevante, dado que el oído humano se puede considerar en una primera aproximación insensible a la fase.

La información contenida en un sonido es descrita como la modulación conjunta de una portadora en frecuencia (FM) y en amplitud (AM), esto es, ambas modulaciones se encuentran acopladas y es muy difícil separarlas, aunque existen desarrollos teóricos que pueden informarnos sobre la contribución al ancho de banda total de cada una de las modulaciones [5].

El análisis ha de realizarse en pequeños fragmentos o tramas de la señal real, donde esta se pueda considerar cuasi-estacionaria, y donde se pueden llevar a cabo mediciones de validez local. La señal de entrada es dividida en tramas de longitud fija. La longitud de una trama se elige según el criterio del usuario. Una trama larga presenta una gran variación de los diferentes parámetros entre el extremo inicial y final. Una trama corta es más regular en sus parámetros, pero es necesario operar sobre un mayor número de ellas para recorrer la señal en su totalidad.

En el presente proyecto se trabajará con **tramas de 500 muestras y en algunos casos de 1000**, para velocidades de muestreo en torno a 44100 muestras/seg.

Tipos de trama y energía:

Al separar la secuencia de entrada en tramas de tamaño fijo, pese a ser cortas (500 muestras) se dará el caso en el que dentro de la propia trama exista tal grado de variación de las componentes sinusoidales en cuanto a amplitud y frecuencia que no pueda considerarse cuasi estacionaria y no podrán realizarse unas medidas válidas de los parámetros de las sinusoides o parciales que puedan describir posteriormente la trama.

Por tanto a la hora de separar la secuencia en tramas nos encontraremos ante dos tipos posibles, según el grado de modulación AM-FM:

- **Trama tonal:** será una trama donde la potencia y la frecuencia de las sinusoides varíe lentamente, el grado de modulación conjunto AM-FM es bajo, pueden considerarse las sinusoides cuasi estacionarias y podrán realizarse medidas válidas. Las tramas tonales **conforman la energía Armónica.**
- **Trama transitoria:** será una trama donde la potencia y/o la frecuencia de las sinusoides varíe rápidamente, en gran magnitud, el grado de modulación AM-FM es alto, en algunos casos una de ambas modulaciones puede ser más pronunciada y visible que la otra, no podrán realizarse medidas válidas sobre los parámetros sinusoidales y no será fácil extraer la parte determinista del ruido. Las tramas transitorias **conforman la energía Transitoria.**

Será necesario establecer un umbral a partir del cual una trama se considere transitoria o se considere tonal.

La energía armónica está asociada a las partes vocales de un sonido, también llamadas zonas “ring”. **La energía transitoria se asocia a las partes consonantes**, correspondientes a chasquidos y transiciones bruscas en el tono o en la amplitud correspondientes a cambios de nota o comienzo/fin de una voz o instrumento, aunque puede considerarse como transitoria una trama correspondiente a un sonido vocal que varíe rápidamente en amplitud y/o frecuencia (como cuando se produce un trino).

Tanto la energía tonal o armónica como la energía transitoria forman parte de la componente determinista de la señal, por tanto no hay que confundir energía transitoria con ruido. **La señal de entrada podrá separarse por tanto en 3 componentes: parte tonal asociada a la energía armónica, parte transitoria asociada a la energía transitoria y parte estocástica asociada al ruido:**

$$x = x_{trans} + x_{tonal} + r \quad (\text{Ecuación 1.4})$$

El ruido estará presente en ambos tipos de tramas, por tanto también se podrán dividir las tramas como ruidosas o poco ruidosas. La parte transitoria estará formada por el conjunto de tramas consideradas transitorias y la parte tonal por el conjunto de tramas consideradas tonales.

Métodos de análisis:

Para estudiar las características de un sonido en profundidad es necesario separar eficazmente la parte determinista (tanto armónica como transitoria) de la estocástica.

Para ello actualmente existen planteados dos métodos o estrategias:

- **Clasificación de picos espectrales:** se trabaja en el dominio frecuencial y para cada trama se separa la parte determinista extrayendo del espectro real los picos que son considerados parciales, el resto se considera el residuo.
- **Modelado no estacionario:** se trabaja en el dominio temporal, se modela para cada trama la parte determinista con un conjunto de polinomios que ponderan una base ortonormal de senos y cosenos. El residuo se obtiene restando la trama modelada de la real.

Describiremos más profundamente ambos métodos, sus ventajas, sus inconvenientes y el grado de desarrollo actual.

1.1 CLASIFICACION DE PICOS

Descripción:

La clasificación de picos trabaja en el **dominio frecuencial**, para ello hace uso de la **STFT** consistente en separar la señal de entrada en tramas, aplicarles una ventana adecuada¹ y computar la FFT para cada trama.

Para cada espectro en magnitud de cada trama obtenido mediante la FFT, se aplican **criterios basados en descriptores derivados de las propiedades tiempo-frecuencia de cada pico frecuencial para decidir si se trata de un parcial o de un pico ruidoso [3,4]**. Los picos considerados como parciales formarán la parte determinista y el resto se considerará el residuo o la parte estocástica.

Ventajas:

Es un método muy desarrollado y estudiado, con buenos resultados para sonidos monofónicos (una sola fuente acústica).

Inconvenientes:

Para **sonidos o tramas muy ruidosas** no es capaz de discernir entre picos deterministas (parciales) y picos ruidosos ya que los primeros quedan enmascarados.

Para **sonidos polifónicos** (varias fuentes acústicas) no funciona correctamente en las zonas donde se producen solapamientos entre parciales armónicos de los diferentes instrumentos. Los solapamientos son muy frecuentes en señales musicales por la propia armonicidad entre instrumentos y voces de las composiciones musicales. Dichos solapamientos adquieren formas erráticas que son difíciles de discernir de picos ruidosos.

Para **tramas con mucha modulación AM-FM consideradas transitorias**, pese a haberse desarrollado descriptores adecuados para sinusoides no estacionarias [3], la detección de picos no funciona adecuadamente. El espectro correspondiente a este tipo de tramas presenta parciales muy ensanchados y deformados debido al efecto de la modulación y con formas diferentes para cada parcial. En el caso de una fuerte modulación FM los parciales más altos son más afectados por la modulación ensanchándose más que los parciales bajos. Esto hace difícil aplicar los algoritmos de detección de picos y discernir entre parciales y picos ruidosos.

¹El inventariado de las tramas es necesario para asegurar una **buena resolución frecuencial**, pues el proceso ha de ser capaz de identificar los picos espectrales correspondientes a sinusoides de la componente determinista y no confundirlos con lóbulos laterales por el **efecto Leakage**. Conocida la frecuencia fundamental podría aprovecharse esta información para la elección del tamaño de la ventana, evitando el efecto Leakage. En el actual proyecto y en las aplicaciones reales no se tiene conocimiento de la frecuencia fundamental, y a la hora de calcular la FFT de cada trama se utilizará un tamaño de ventana fijo igual al tamaño de la trama (500 ó 1000 muestras) y de tipo Hanning. Para mejorar así mismo la resolución espectral, el número de muestras de la FFT deberá ser suficientemente alto en relación al tamaño de la trama y será potencia de 2 para asegurar una mayor eficiencia computacional. En el actual proyecto cuando se calcule la FFT de cada trama se hará con un número de bins (muestras del espectro) de **2048 puntos**.

1.2 MODELADO NO ESTACIONARIO

Descripción:

El modelado no estacionario trabaja en el dominio temporal. Para cada trama se modela su parte determinista con un conjunto de polinomios que ponderan una base ortonormal de senos y cosenos, similar a la empleada en el desarrollo de Fourier. Si la base ortonormal estuviera ponderada por constantes, se hablaría de modelado estacionario, esto es, para cada trama se supondrían constantes los parámetros que describen las sinusoides.

Con el modelado no estacionario, al trabajar con polinomios, **podemos describir las variaciones en amplitud y frecuencia que se producen dentro de una trama** para las distintas sinusoides y por tanto **podemos modelar las variaciones AM-FM** que se producen dentro de la propia trama, dentro de unos límites de linealidad en la variación de dichas modulaciones que vendrán asociados a los límites de tonalidad o transitoriedad de la trama.

Cuando una trama presente una modulación AM-FM dentro de los límites de linealidad se dirá que se tiene una trama **cuasi estacionaria**.

Tipos de modelado no estacionario:

En este punto hay que distinguir entre los dos tipos de modelado no estacionario que se han desarrollado actualmente, los cuales cubren los dos tipos de energía que nos podemos encontrar, tanto la armónica (tramas tonales) como la transitoria (tramas transitorias):

- **Modelado no estacionario polinomial:** es el ya mencionado, la **base ortonormal es ponderada por un conjunto de polinomios** que describen la modulación AM-FM dentro de la propia trama pero dentro de unos límites de linealidad. Se utiliza para **modelar tramas tonales o cuasi estacionarias. Necesario definir los límites de estacionariedad.**
- **Modelado no estacionario lineal por tramos:** en este caso la **base ortonormal es ponderada por funciones triangulares**, con ello se consigue describir mediante líneas rectas las variaciones rápidas de las modulaciones AM-FM producidas dentro de la propia trama, que no se llegan a describir con funciones polinómicas. Se utiliza para **modelar tramas no estacionarias o transitorias**, con rápidas variaciones en amplitud y frecuencia de las componentes sinusoidales.

Ambos tipos de modelado no estacionario pueden ser combinados para implementar una herramienta que modele los dos tipos de energía descritos y la totalidad de las tramas que nos podamos encontrar (modelado mixto).

Descripción matemática del modelado:

Modelado no estacionario polinomial:

Se asume que la frecuencia dentro de la trama varía linealmente en torno a un valor medio F_0 . La estimación del valor de F_0 para cada trama es un paso crítico en presencia de ruido y modulaciones al límite, es eficazmente solucionado con el método ASA (Automated Spectral Analysis method) [2]. Una vez estimado el valor medio de la frecuencia fundamental de la trama F_0 , el resto de parámetros armónicos son estimados mediante una función de coste lineal basada en mínimos cuadrados (least-squares o LS).

El sonido discretizado producido por una fuente armónica viene descrito por el siguiente modelo:

$$s(n) = \sum_{i=1}^I A_i(n) \cdot \cos[2\pi i F_0(n)n + \varphi_i] + r(n) \quad (\text{Ecuación 1.2.1})$$

para $n=0,1,\dots,N-1$ donde N es el número de muestras de la trama o ventana dada.

Basado en el modelo descrito inicialmente, la parte determinista es descrita como una superposición de I armónicos cuyas variaciones en amplitud y frecuencia vienen dadas por los parámetros variantes en el tiempo (no estacionarios) $A_i(n)$ y $F_0(n)$.

La parte estocástica o residuo viene dada por $r(n)$.

Aplicando identidades trigonométricas podemos descomponer la parte determinista del modelo del siguiente modo:

$$\cos(\alpha + \beta) = \cos\alpha \cdot \cos\beta - \text{sen}\alpha \cdot \text{sen}\beta$$

$$s(n) = \sum_{i=1}^I a_i(n) \cdot \sin[2\pi i F_0(n)n] + b_i(n) \cos[2\pi i F_0(n)n] \quad (\text{Ecuación 1.2.2})$$

donde:

$$a_i(n) = -A_i(n) \sin \varphi_i \quad \text{y} \quad b_i(n) = A_i(n) \cos \varphi_i$$

Asumiendo que las variaciones en amplitud y frecuencia respecto al valor medio se producen linealmente a lo largo de la trama (ventana corta), podemos descomponer las funciones $a_i(n)$, $b_i(n)$ y $F_0(n)$ dependientes del tiempo (n) como:

$$a_i(n) = a_0^{(i)} + a_1^{(i)} \cdot n$$

$$b_i(n) = b_0^{(i)} + b_1^{(i)} \cdot n$$

$$F_0(n) = F_0 + \beta \cdot n$$

Sustituyendo en (1.1.2) y aplicando identidades trigonométricas obtenemos una **combinación lineal** de productos de amplitudes por senos y cosenos con argumentos $2\pi F_0 n$ y $2\pi\beta n^2$.

Para señales cuasi estacionarias e incluso para señales con fuerte modulación AM-FM como en el caso de un vibrato (caso extremo) podemos asumir que:

$$\sin 2\pi\beta n^2 \approx 2\pi\beta n^2, \quad \cos 2\pi\beta n^2 \approx 1 \quad \text{y lo mismo para los términos con } 2\pi F_0 n$$

Simplificando obtenemos la siguiente expresión para el modelo:

$$s(n) = \sum_{i=1}^I p_s^{(i)}(n) \cdot \sin[2\pi F_0 n] + p_c^{(i)}(n) \cos[2\pi F_0 n] \quad (\text{Ecuación 1.2.3})$$

donde:

$$p_s^{(i)}(n) = a_0^{(i)} + a_1^{(i)} \cdot n - 2\pi b_0^{(i)} \beta \cdot n^2 - 2\pi b_1^{(i)} \beta \cdot n^3$$

$$p_c^{(i)}(n) = b_0^{(i)} + b_1^{(i)} \cdot n - 2\pi a_0^{(i)} \beta \cdot n^2 - 2\pi a_1^{(i)} \beta \cdot n^3$$

son polinomios que describen las variaciones en amplitud y en frecuencia. El grado de dichos polinomios suele estar entre 3 y 6 (en este caso son de grado 3) y determinarán el grado de libertad con el que los polinomios se ajusten a las variaciones AM-FM en cada trama.

Con la estimación de F_0 obtenida con el método ASA, solo necesitamos estimar los coeficientes de los polinomios para tener la trama totalmente descrita. La estimación de los coeficientes se realiza mediante mínimos cuadrados. Podemos expresar los polinomios de forma más compacta y general como:

$$p_s^{(i)}(n) = \sum_{k=0}^P p_{sk}^{(i)} n^k$$

$$p_c^{(i)}(n) = \sum_{k=0}^P p_{ck}^{(i)} n^k$$

donde **P** es el **grado de los polinomios**. Computacionalmente siempre será **más eficiente tener un grado de polinomio lo más bajo posible** dentro de que se modele bien la parte determinista.

Definimos el vector p con los coeficientes de ambos polinomios como:

$$p = (p_{s0}^{(0)} \dots p_{sP}^{(0)} \dots p_{s0}^{(I)} \dots p_{sP}^{(I)} \dots p_{c0}^{(0)} \dots p_{cP}^{(0)} \dots p_{c0}^{(I)} \dots p_{cP}^{(I)})^T$$

formado por $2 \cdot I \cdot P$ coeficientes en total.

Si tenemos en el vector s la señal musical original (trama) formularemos la siguiente igualdad matricial:

$$\boxed{s = \Psi \cdot p + r} \quad (\text{Ecuación 1.2.4})$$

donde el vector r es el formado por el ruido de la trama y la matriz Ψ es la formada por la base ortonormal, de tamaño $N \times 2 \cdot I \cdot P$ que viene dada por:

$$\Psi = \begin{pmatrix} \Psi^T(s(n), c(n)) \\ \cdot \\ \cdot \\ \cdot \\ \Psi^T(s(n-N-1), c(n-N-1)) \end{pmatrix}$$

con $\Psi^T(s(n), c(n)) = (1 \dots s(I \cdot n) \dots c(I \cdot n))$, $s(n) = \sin 2\pi F_0 n$ y $c(n) = \cos 2\pi F_0 n$

Despejando el vector p obtendremos la estimación por mínimos cuadrados de los coeficientes de los polinomios que describen la información determinista contenida en la trama original s :

$$\boxed{\hat{p}^{LS} = (\Psi^T \Psi)^{-1} \cdot \Psi^T \cdot s} \quad (\text{Ecuación 1.2.5})$$

Modelado no estacionario lineal por tramos:

Si tomamos el modelado polinomial con grado de polinomio 1 tendremos rectas o funciones triangulares ponderando la base ortonormal, por tanto el desarrollo matemático es el mismo solo que empleando funciones triangulares en lugar de polinomios.

En el modelado lineal por tramos **se define la variable Breaks como el número de funciones triangulares o segmentos lineales que se emplean dentro de la trama para describir las variaciones en amplitud y frecuencia** de las componentes sinusoidales que conforman la base ortonormal.

Ventajas:

Computacionalmente el modelado es mucho más eficiente que la detección de picos, ya que se trabaja en el dominio temporal con matrices, más adecuado para un entorno de programación como el empleado, Matlab.

El modelado es capaz de separar más energía residual de la parte determinista que la detección de picos en condiciones de SNR_p bajas en las que la detección de picos no es capaz de discernir entre picos ruidosos y picos deterministas [2].

Funciona con tramas consideradas transitorias, aplicando el modelado mixto podemos modelar y separar tanto la energía armónica como la transitoria del ruido, la detección de picos solo es capaz de separar el ruido en las tramas armónicas o tonales.

Funciona con señales musicales polifónicas, siempre que se conozca el número de instrumentos o fuentes armónicas y sus frecuencias fundamentales. Se modela en el dominio temporal la parte determinista de cada instrumento como una serie armónica según cada frecuencia fundamental detectada para cada instrumento. Nos olvidamos de los solapamientos que se puedan producir en el dominio frecuencial, que tantos problemas producen en la detección de picos.

Mayor rango de aplicación que la detección de picos, al poder modelar las partes transitorias y separarlas del ruido las aplicaciones en otros campos en los que se necesite estudiar señales transitorias aumentan.

Inconvenientes:

El principal inconveniente a la hora de aplicar el modelado no estacionario es el de **encontrar los parámetros óptimos de modelado**, es decir, los valores del grado del polinomio **P** en el **modelado polinomial** y el número de tramos **breaks** en el **modelado lineal por tramos**, además de saber en que caso se aplica uno u otro tipo de modelado.

- **Para el modelado polinomial:** Cuando la trama a modelar presenta mucha modulación AM-FM y se modela con un grado de polinomio **P** bajo se produce un efecto conocido como **SUBFIT** o **Subestima** en el que el polinomio no es capaz de seguir adecuadamente las variaciones AM-FM que presenta la trama y por tanto **no se modela bien la información determinista de ésta**.

Cuando la trama a modelar presenta bastante ruido y se modela con un grado de polinomio **P** alto se produce un efecto conocido como **OVERFIT** o **Sobreestima** en el que se **modela además de la parte determinista parte del ruido**, ya que el polinomio presenta un gran grado de libertad y sigue pequeñas variaciones y pulsos producidos por el ruido.

- **Para el modelado lineal por tramos:** Si no se ajusta el valor óptimo de número de tramos lineales o breaks que modelan las rápidas variaciones en la modulación AM-FM se producen efectos similares de subestima o sobreestima que en el caso de modelado polinomial.

2. Antecedentes, Objetivos, Motivación y Aplicaciones

2.1 ANTEDECENTES

Dentro del modelado no estacionario, hasta ahora los algoritmos implementados y estudiados aplican a toda la secuencia musical los mismos valores de los parámetros de modelado, es decir, **se fija para todas las tramas un mismo valor del grado de polinomio P o un mismo valor del número de tramos lineales breaks.**

Esto conlleva a que en muchas tramas se produzcan efectos de **Overfit** y de **Subfit** según las características de modulación AM-FM de estas y por tanto **no se modele bien la parte determinista en algunas de ellas y en otras se modele parte del ruido.**

Además **cuando se aplica uno u otro tipo de modelado** (polinomial o lineal por tramos) **no se aplica uno de ellos a cada trama en concreto**, si no que se aplica a las zonas o secuencias en las que se cree que se concentra la energía tonal o la energía transitoria. **La optimización de los parámetros de modelado para cada trama en concreto no está actualmente desarrollado ni implementado**, es necesario un estudio de los parámetros óptimos de modelado según las características de ruido y modulación AM-FM que pueda presentar cada trama y además **no existe un método o un baremo de medida que nos determine claramente cuando una trama puede considerarse tonal o pase a considerarse transitoria.**

2.2 OBJETIVOS

El **objetivo principal** del presente proyecto es el de **optimizar el modelado no estacionario de forma que para cada trama en concreto se aplique el tipo de modelado y los parámetros de modelado óptimos para ésta**, con el fin de **no producir Overfit ni Subfit** y mejorar así el modelado no estacionario de la parte determinista para ambos tipos de energía, tonal y transitoria en señales musicales. **Optimizando el modelado de la parte determinista mejoraremos su separación de la componente residual.**

Para ello este proyecto se divide en **3 tareas u objetivos diferentes:**

- **Objetivo 1:** Realizar un estudio de los parámetros óptimos de modelado según las características de modulación AM-FM y ruido que nos podemos encontrar en una trama.
- **Objetivo 2:** Realizar un estudio de las posibles mejoras en cuanto a mitigación del Overfit de aplicar un modelo híbrido de modelado y clasificación de picos, empleando un modelado polinomial de parámetros fijos.
- **Objetivo 3:** Investigar y desarrollar un método o baremo de medida que nos determine el grado de transitoriedad o tonalidad de una trama en concreto, con el fin de poder definir una aplicación directa entre el grado de tonalidad/transitoriedad obtenido, el nivel de ruido a unos parámetros óptimos de modelado.

2.3 MOTIVACIÓN Y APLICACIONES

El modelado no estacionario es una vertiente nueva en el estudio de la separación del ruido y la parte determinista en señales musicales con prometedores resultados y mejoras respecto al camino que se había seguido hasta ahora basado en la detección de picos en el dominio frecuencial.

Además de lo novedoso el nuevo método tiene un rango de aplicación mayor que la detección de picos ya que **con el modelado no estacionario podemos separar el ruido de la energía transitoria y no solo de la armónica.**

El presente proyecto además tiene como reto desarrollar un método de medida del grado de armonicidad de una trama con aplicación directa en la optimización del modelado y con múltiples **aplicaciones en otras ramas de la ingeniería y la biomedicina en las que se tiene señales con partes tonales y partes transitorias e interese separarlas para estudiarlas por separado, además de limpiarlas de ruido.**

3. Descripción de las tramas

Para el desarrollo del actual proyecto se trabajará con tramas pertenecientes a secuencias/archivos musicales reales, con características en cuanto a modulación y nivel de ruido diferentes entre si, con el fin de tener **un ejemplo representativo de cada tipo de trama** que nos podamos encontrar en secuencias musicales reales.

Utilizaremos **tramas de archivos musicales monofónicos** (un solo instrumento) por simplicidad y para poder comparar posteriormente con la detección de picos (el modelado se puede utilizar igualmente con señales polifónicas).

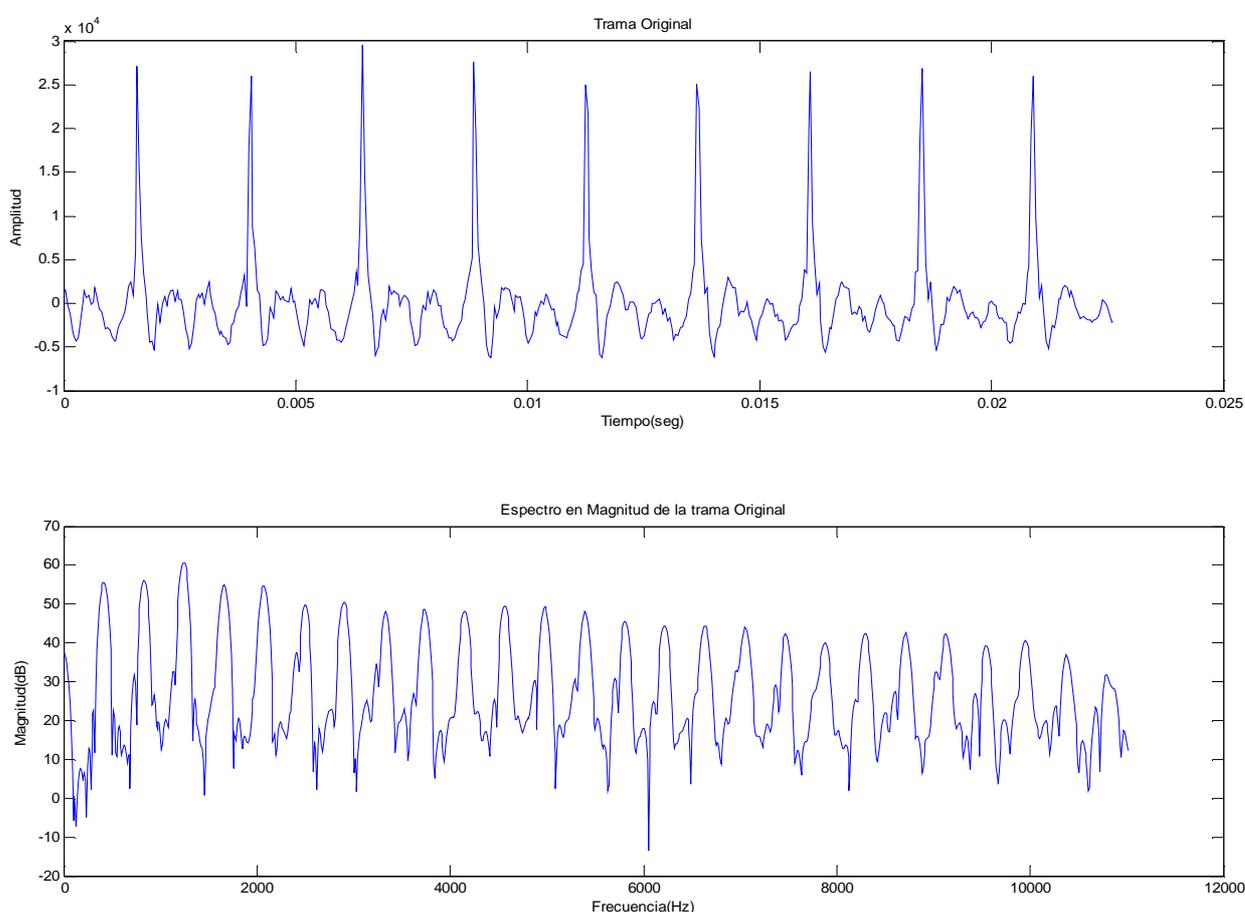
Pasamos a describir y enumerar las distintas tramas que vamos a emplear para experimentar con los algoritmos que desarrollaremos y estudiaremos².

²Para todas las tramas recogemos sus características en cuanto a nivel de modulación, ruido, número de muestras, frecuencia de muestreo y las representamos en el tiempo y en la frecuencia. Para calcular sus espectros aplicamos la FFT de **2048 bins** sobre las tramas inventanadas con ventana Hanning para tener una buena resolución espectral. Para todas ellas **representaremos la mitad de su espectro** por ser señales reales y lo representaremos **en decibelios** para evaluar mejor el nivel de ruido presente en la señal. Las secuencias originales son diezmadas a la mitad para mejorar la eficiencia computacional, las tramas son tomadas de las secuencias tras diezmar, por tanto tienen una frecuencia de muestreo **SR (Sample Rate)** mitad que las secuencias originales.

3.1 TRAMA 1: cuasi estacionaria con poco ruido y poca modulación. Trumpet

- Trama muy limpia de ruido, con poca modulación AM-FM, por tanto **dentro del límite de lo cuasi estacionario**.
- Pertenece al archivo audio **'trumpet4.aiff'** en el que suena una trompeta. Es la cuarta trama de la secuencia con ventana de **500 muestras**, muestras 1501:2000 (tras diezmar).
- La secuencia original tiene una frecuencia de muestreo $SR=44100$ muestras/seg, tras diezmar tendremos **$SR=22.050$ muestras/seg** (la mitad de muestras que la original) y una frecuencia máxima en el espectro de **$f_{max}=11.025$ Hz** (la mitad de SR).

Representamos la trama en el tiempo y en la frecuencia:



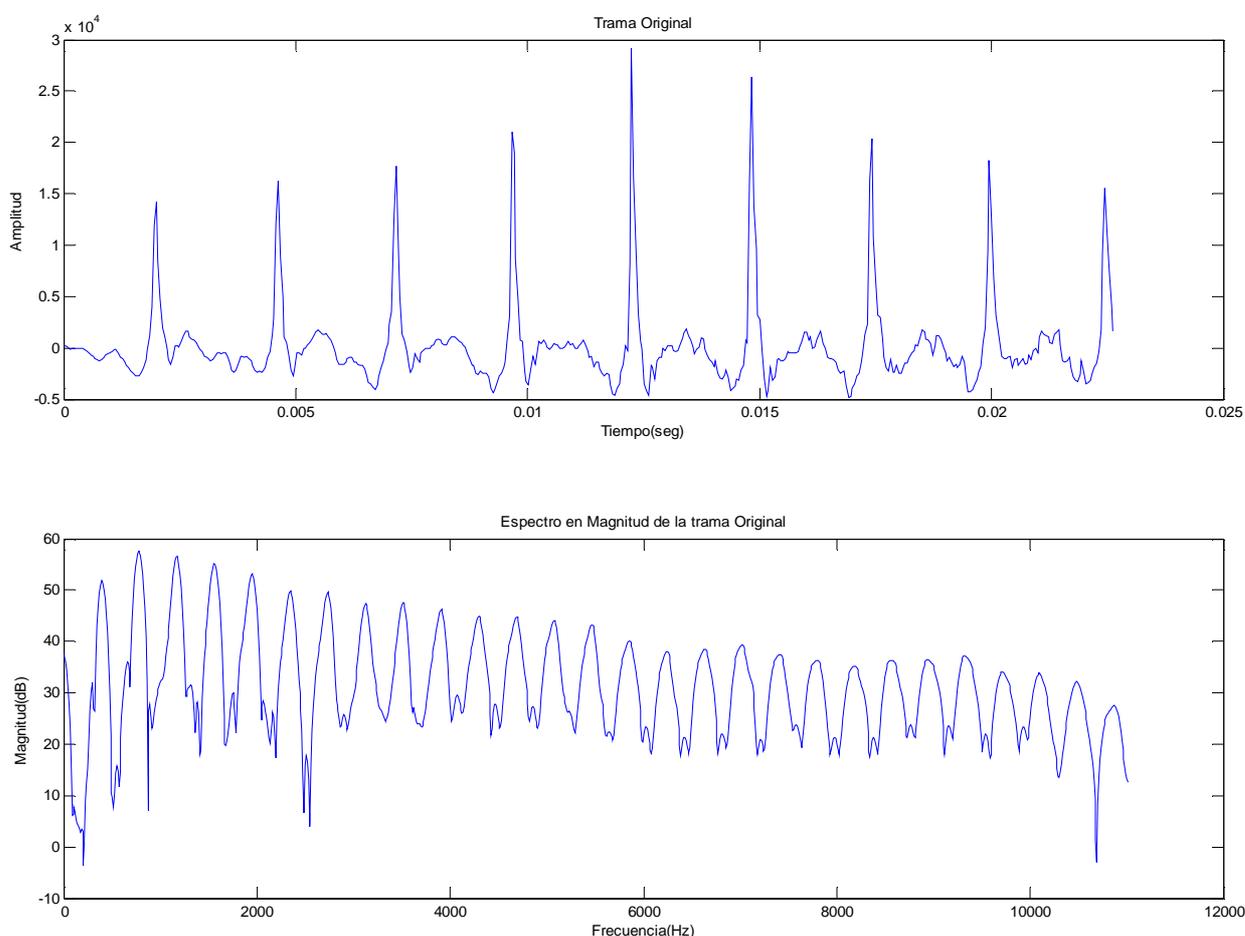
(Fig. 3.1 Trama 1)

Se observa en el dominio temporal la **poca modulación AM-FM**. Del espectro se deduce que el Pitch o frecuencia fundamental de la trama es cercana a los **400 Hz**, el resto de parciales están presentes en las posiciones armónicas, se trata de una **trama tonal**. Se observa como los **parciales están muy definidos** por el poco ruido presente en la trama, existe un margen promedio de **30 decibelios entre el nivel de los parciales y el nivel del ruido**.

3.2 TRAMA 2: cuasi estacionaria con poco ruido y mucha modulación. Trumpet

- Trama con poco ruido y una **fuerte componente de modulación AM en el límite** pero considerada **cuasi estacionaria**.
- Pertenece al archivo audio **'trumpet4.aiff'** en el que suena una trompeta. Es la segunda trama de la secuencia con ventana de **500 muestras**, muestras 501:1000 (tras diezmar).
- La secuencia original tiene una frecuencia de muestreo $SR=44100$ muestras/seg, tras diezmar tendremos $SR=22.050$ muestras/seg y una frecuencia máxima en el espectro de $f_{max}=11.025$ Hz.

Representamos la trama en el tiempo y en la frecuencia:



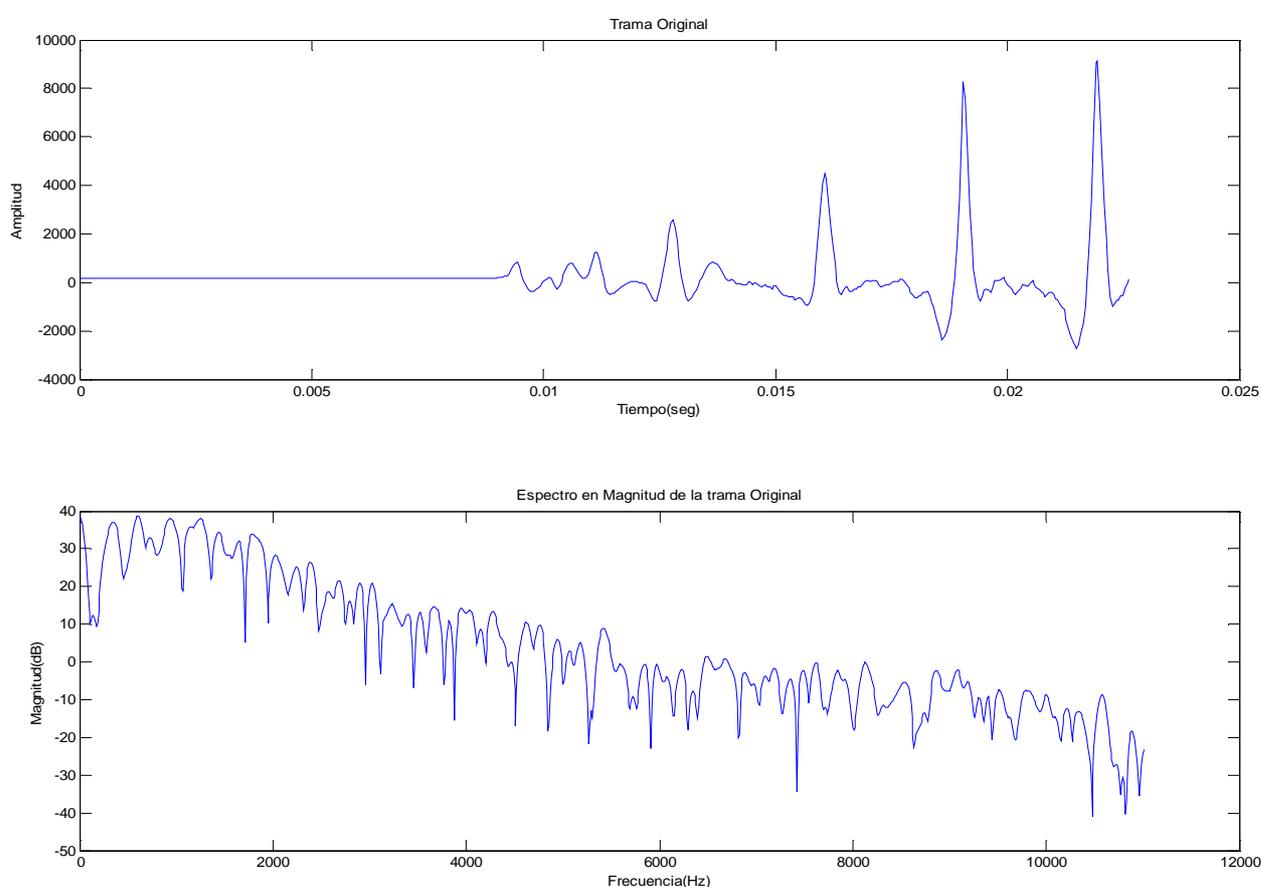
(Fig. 3.2 Trama 2)

Se observa en el dominio temporal la **fuerte modulación AM**. Del espectro se deduce que el Pitch o frecuencia fundamental de la trama es cercana a los **400 Hz**, el resto de parciales están presentes en las posiciones armónicas, se trata de una **trama tonal**. Los **parciales están muy definidos**, existe un margen promedio de **30 decibelios** entre el nivel de los parciales y el nivel del ruido.

3.3 TRAMA 3: transitoria con poco ruido. Trumpet

- Trama con poco ruido y una modulación **fuera del límite de lo cuasi estacionario** y por tanto **transitoria**.
- Pertenece al archivo audio **‘trumpet4.aiff’** en el que suena una trompeta. Es la primera trama de la secuencia con ventana de **500 muestras**, muestras 1:500 (tras diezmar). Es el **ataque inicial**.
- La secuencia original tiene una frecuencia de muestreo $SR=44100$ muestras/seg, tras diezmar tendremos **$SR=22.050$ muestras/seg** y una frecuencia máxima en el espectro de **$f_{max}=11.025$ Hz**.

Representamos la trama en el tiempo y en la frecuencia:



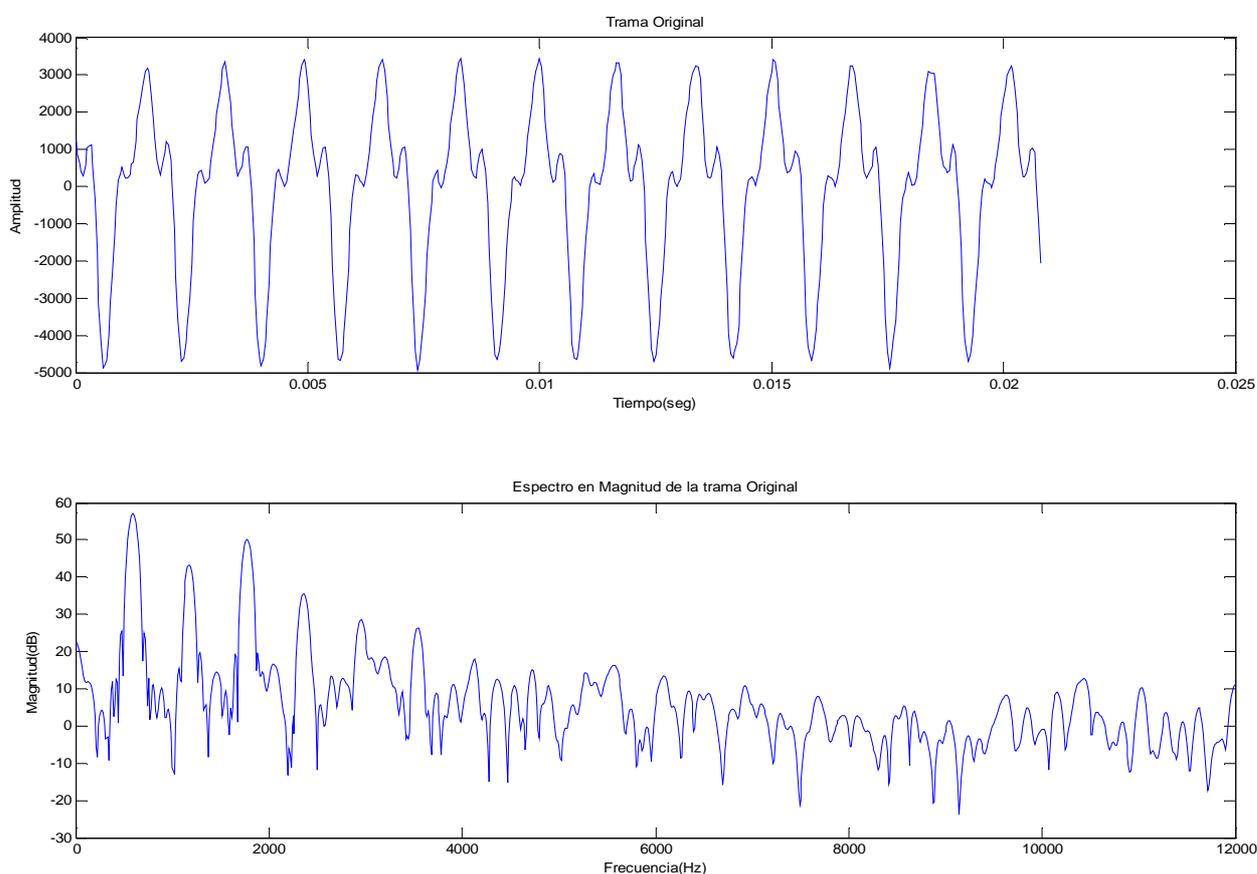
(Fig. 3.3 Trama 3)

Se observa en el dominio temporal como la **fuerte modulación AM-FM correspondiente al ataque inicial se sale del límite de lo cuasi estacionario**. En el espectro se observa como los parciales no presentan patrones de forma similar entre ellos y es difícil situarlos en sus posiciones armónicas a partir del quinto parcial. Se trata de una **trama transitoria y por tanto no armónica**.

3.4 TRAMA 4: cuasi estacionaria con mucho ruido y poca modulación. Shakuhachi

- Trama con **mucho ruido convolutivo** característico del instrumento de viento madera asiático Shakuhachi, asociado al roce producido por el aire al atravesar la lengüeta al soplar. Con poca modulación AM-FM, por tanto **dentro del límite de lo cuasi estacionario**.
- Pertenece al archivo audio ‘Shakuhachi.aiff’ en el que suena dicho instrumento, similar a una flauta. Es la **trama 4321** con ventana de **500 muestras**, muestras 2160001:2160500 (tras diezmar) correspondiente al minuto 1:30 de la pista original.
- La secuencia original tiene una frecuencia de muestreo $SR=48000$ muestras/seg, tras diezmar tendremos **$SR=24.000$ muestras/seg** y una frecuencia máxima en el espectro de **$f_{max}=12.000$ Hz**.

Representamos la trama en el tiempo y en la frecuencia:



(Fig. 3.4 Trama 4)

Se observa en el dominio temporal como la trama tiene una componente de modulación AM-FM pequeña, es una **trama cuasi estacionaria**. Del espectro se deduce que el Pitch de la trama corresponde a una frecuencia de **585 Hz**, el resto de parciales están presentes en las posiciones armónicas pero muy **enmascarados por el ruido a partir del séptimo parcial**. Es una **trama tonal con mucho ruido**.

En la trama 4 observamos como **la envolvente espectral decae rápidamente**, al contrario de lo que ocurría en las tramas tonales correspondientes a la trompeta. Es una **característica particular de cada instrumento**.

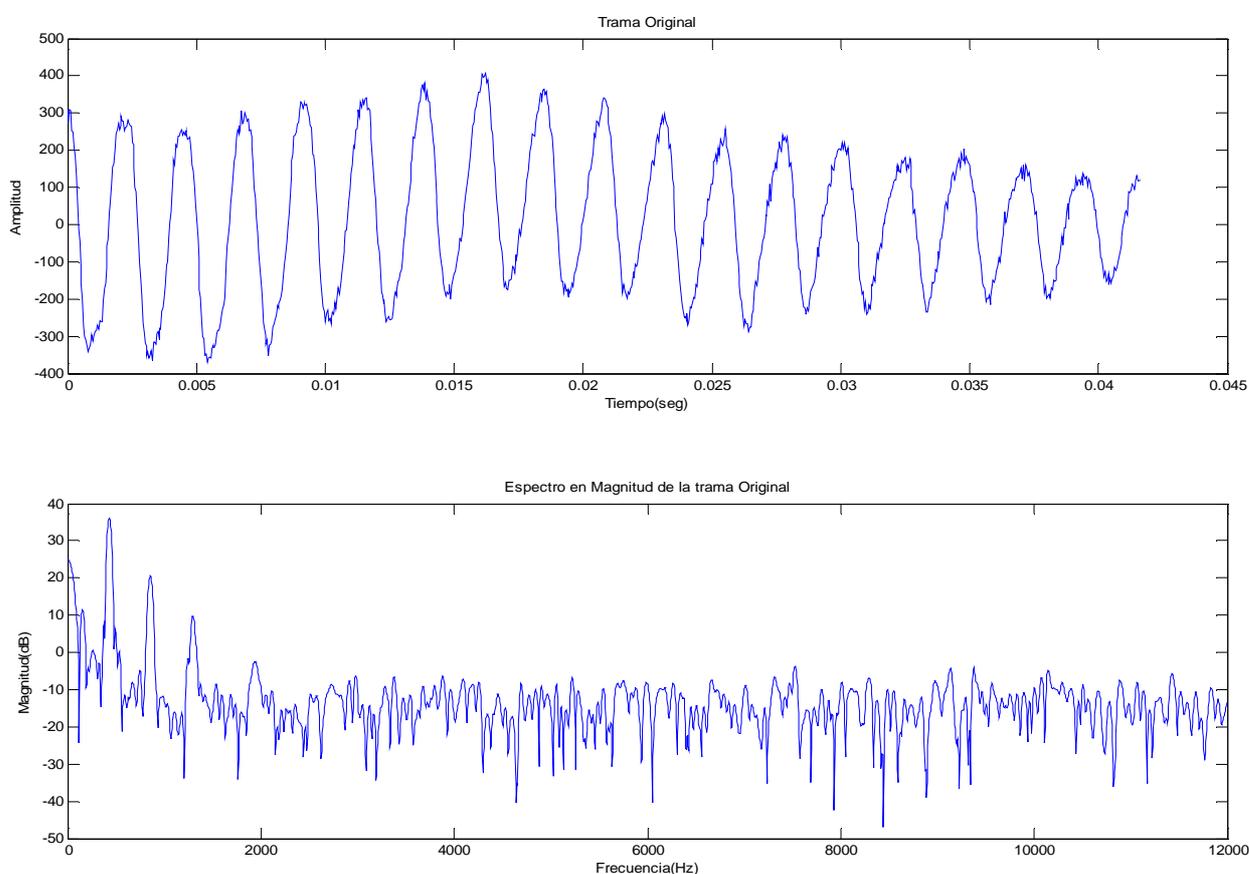
Observando los **3 primeros parciales** podemos decir que para esta trama existe un margen promedio entre el **nivel de los parciales y el del ruido de unos 50 decibelios**. El margen va decayendo hasta hacerse cero a medida que cae la potencia de los parciales y se camuflan con el ruido siguiendo la envolvente.

En las tramas tonales correspondientes a la trompeta, el margen entre el nivel de los parciales y el ruido era menor, en torno 30 decibelios, pero en ese caso **la energía de la señal estaba distribuida a lo largo de muchos parciales**, ya que su envolvente caía muy lentamente. En las **tramas correspondientes al instrumento Shakuhachi la mayor parte de la energía se concentra en los 5-6 primeros parciales** y eso hace que para éstos su potencia respecto al nivel del ruido sea mayor.

3.5 TRAMA 5: cuasi estacionaria con mucho ruido y mucha modulación. Shakuhachi

- Trama con **mucho ruido convolutivo** y con **mucha modulación AM-FM dentro del límite de lo cuasi estacionario**.
- Pertenece al archivo audio **'Shakuhachi.aiff'** en el que suena dicho instrumento. Es la **trama 1777** con ventana de **1000 muestras**, muestras 888001:889000 (tras diezmar) correspondiente al minuto 0:37 de la pista original.
- La secuencia original tiene una frecuencia de muestreo $SR=48000$ muestras/seg, tras diezmar tendremos **$SR=24.000$ muestras/seg** y una frecuencia máxima en el espectro de **$f_{max}=12.000$ Hz**.

Representamos la trama en el tiempo y en la frecuencia:



(Fig. 3.5 Trama 5)

En dominio temporal se aprecia como la señal tiene una gran modulación AM-FM (junto a una variación en la componente DC) pero la consideramos dentro del límite de lo cuasi estacionario. En el dominio frecuencial observamos la gran cantidad de ruido presente en la señal, a partir del cuarto parcial los demás están muy enmascarados por pulsos ruidosos. El Pitch de la señal está en torno a los **433 Hz**. Estamos ante un ejemplo del **peor caso de trama tonal que nos podemos encontrar**, con mucho ruido y mucha variación en amplitud y frecuencia.

Además de las tramas representativas aquí descritas, en algunos puntos del desarrollo se tomarán otras tramas adicionales con el fin de corroborar y contrastar resultados además de tramas sintéticas que nos servirán de referencia. En cada caso cuando llegue el momento se explicará el uso de cada una de ellas.

4. Estudio del Overfit en el Modelado no estacionario

En este estudio ilustraremos los efectos de **Overfit** y **Subfit** que se producen en el modelado no estacionario aplicándolo a las tramas representativas descritas anteriormente cuando los parámetros y el tipo de modelado no se eligen de manera adecuada a las características de la trama.

4.1 ESTUDIO DEL OVERFIT EN EL MODELADO POLINOMIAL

Emplearemos primeramente el algoritmo clásico de modelado no estacionario polinomial para modelar la parte determinista de las tramas siguiendo los pasos explicados en la descripción matemática de dicho método.

Aplicamos la función que implementa el método a cada trama con diferentes parámetros de entrada, **grado del polinomio** y **número de armónicos**, para **aproximarlas polinomialmente** (modelar su parte determinista) **en el dominio temporal**³.

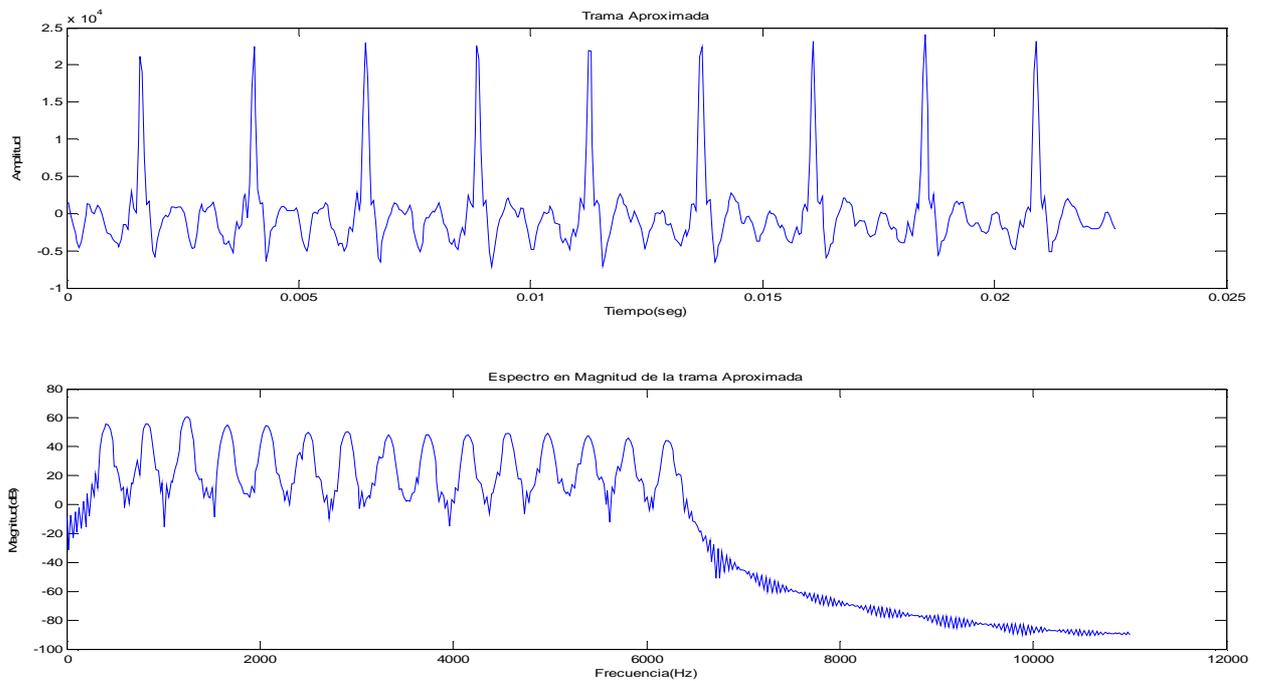
³ El modelado polinomial está implementado en la función **“PolyApproximation.m”** y se aplica sobre las tramas sin inventanar. Tendremos: **“x”** trama original, **“xapp_gradoP”** trama aproximada con polinomio de grado P y **“z”** residuo o ruido ($z=x-xapp_gradoP$). Aplicamos la FFT sobre dichas señales inventanadas y obtenemos: **“X”** espectro de la trama original, **“Xapp_gradoP”** espectro de la trama aproximada y **“Z”** espectro del residuo. Todas las variables y resultados obtenidos son guardados en archivos de variables Matlab, uno para cada trama.

4.1.1 MODELADO POLINOMIAL TRAMA 1:

Trama cuasi estacionaria poco ruidosa con poca modulación. Trumpet

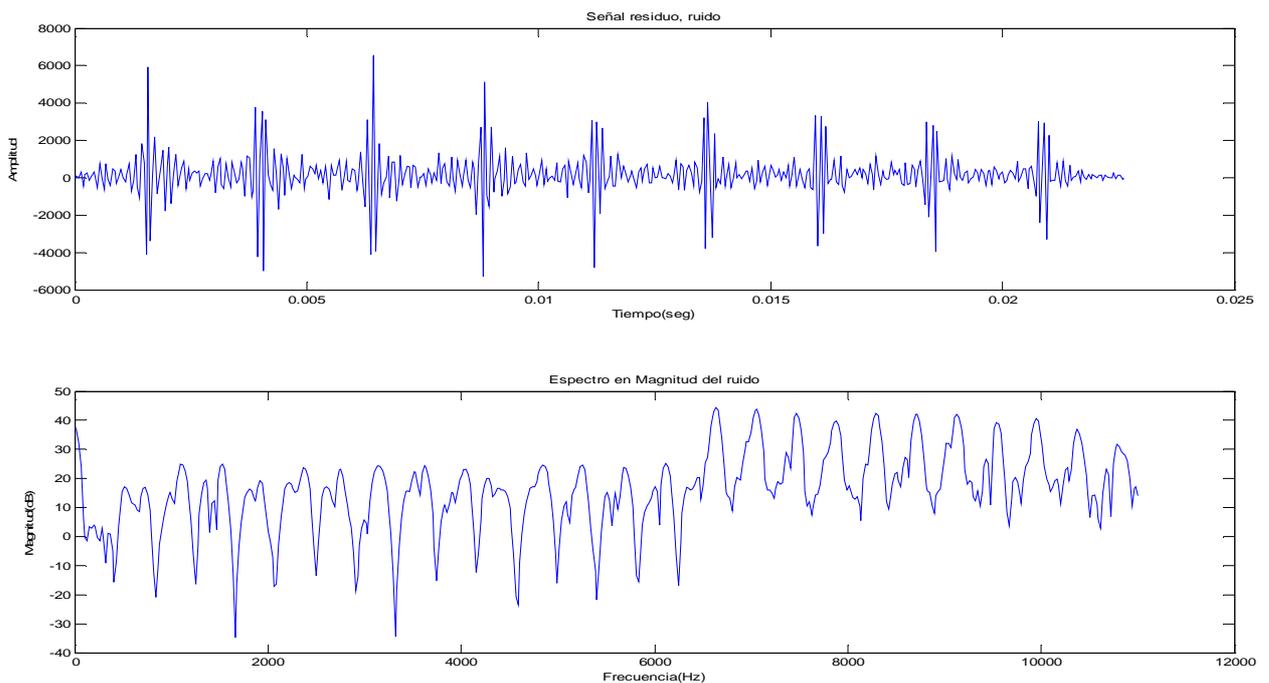
4.1.1.1 - Grado del polinomio= 6, número de armónicos=15.

Representamos la trama aproximada/modelada polinomialmente y su espectro:



(Fig. 4.1.1.1.1 Trama 1 Modelada polinomialmente, $P=6$, $N_{harm}=15$)

Para el residuo obtenemos lo siguiente:



(Fig. 4.1.1.1.2 Residuo o ruido de la Trama 1, $P=6$, $N_{harm}=15$)

Se observa en la trama aproximada como se han modelado solo los 15 primeros armónicos (es un parámetro de entrada en la función) pero en el espectro de la trama original (*Fig. 3.1*) se observa como son importantes un mayor número de parciales.

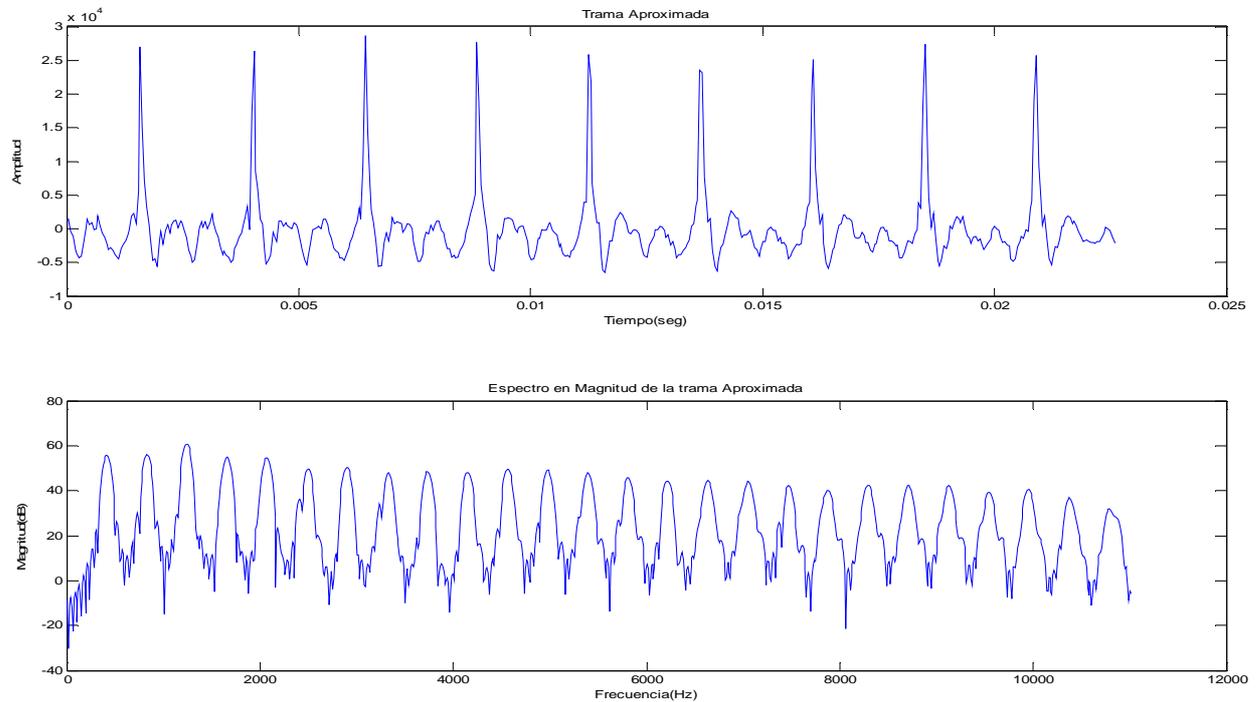
Con solo 15 armónicos se puede observar en el espectro de la parte residual como se produce un salto brusco a partir del último armónico aproximado, que se traduce en una componente determinista a partir de dicho armónico y en el tiempo tenemos para el ruido cierta periodicidad que no debería existir (el ruido es aleatorio).

El número de armónicos a tener en cuenta para la aproximación dependerá del nivel de ruido y de la envolvente concreta que tenga el instrumento a estudiar, que en unos casos podrá decaer más rápidamente o más lentamente, y por tanto se tendrán en cuenta menos o más parciales respectivamente. Para la trompeta que es el instrumento que suena en esta trama **la envolvente decae lentamente y además se tiene poco ruido**, por ello la parte determinista tiene más peso y **debemos tener en cuenta un mayor número de armónicos a la hora de modelar (entre 25 y 30)**.

Realizamos el mismo análisis pero incrementando el número de parciales a tener en cuenta en la aproximación para eliminar la periodicidad del ruido y quitarle así la componente determinista que teníamos al modelar con 15 armónicos.

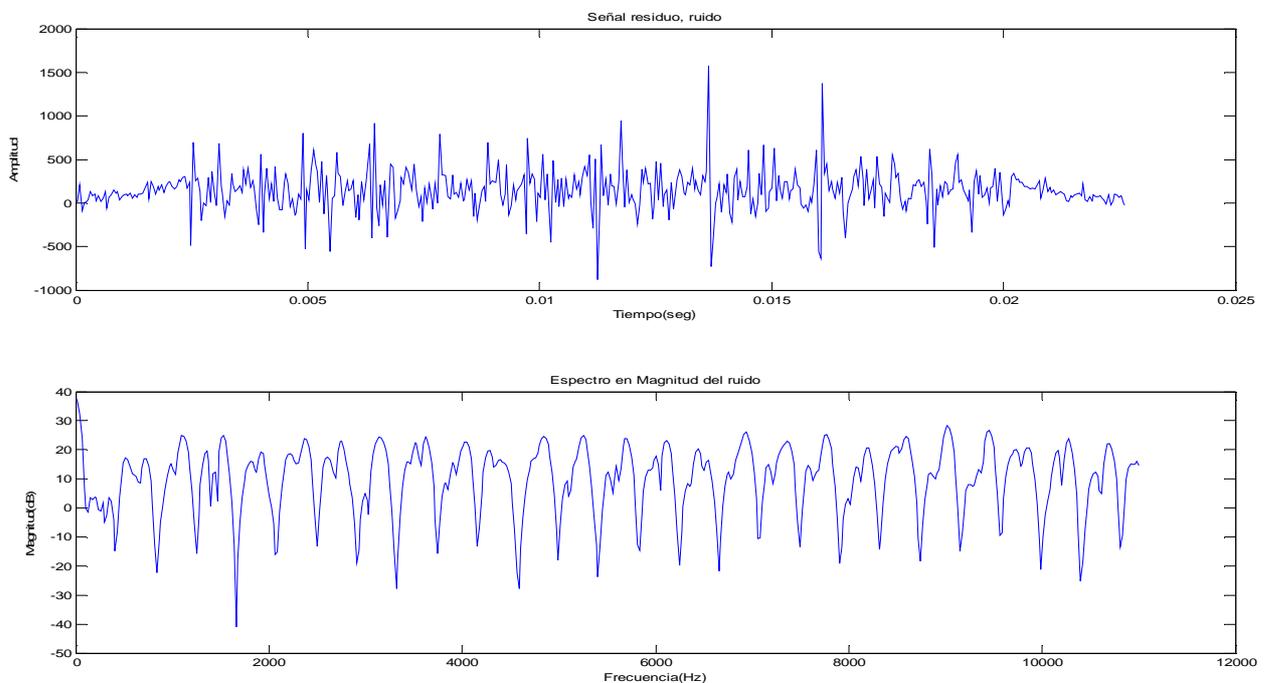
4.1.1.2 - Grado del polinomio= 6, número de armónicos=28.

Trama aproximada en el tiempo y su espectro:



(Fig. 4.1.1.2.1 Trama 1 Modelada polinomialmente, $P=6$, $N_{harm}=28$)

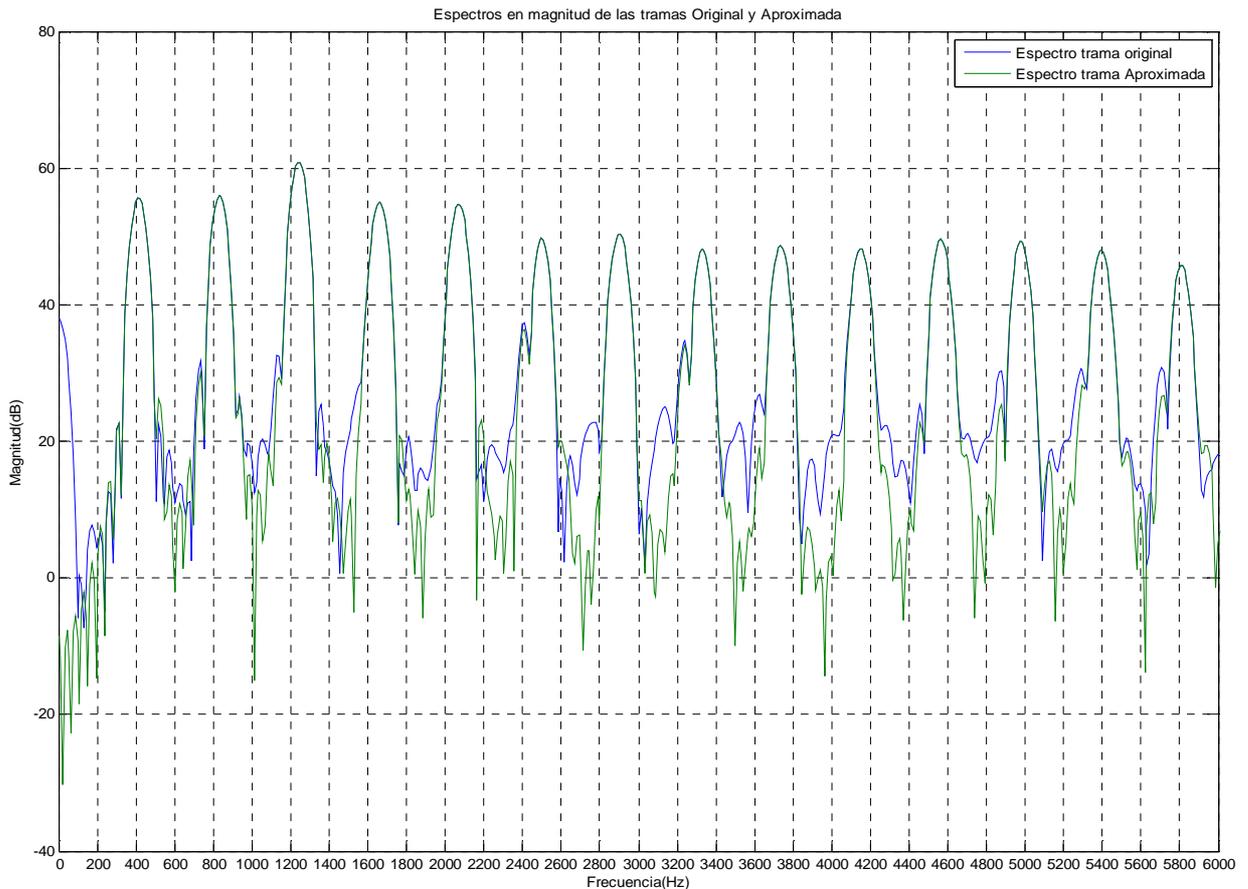
Residuo:



(Fig. 4.1.1.1.2 Residuo o ruido de la Trama 1, $P=6$, $N_{harm}=28$)

En este caso (aproximación con 28 armónicos) se observa como el espectro del residuo es mas uniforme y en el tiempo no existe periodicidad como en el caso anterior, **se esta modelando bien la parte determinista y separándola del ruido.**

Ampliamos los espectros de las tramas original y aproximada:

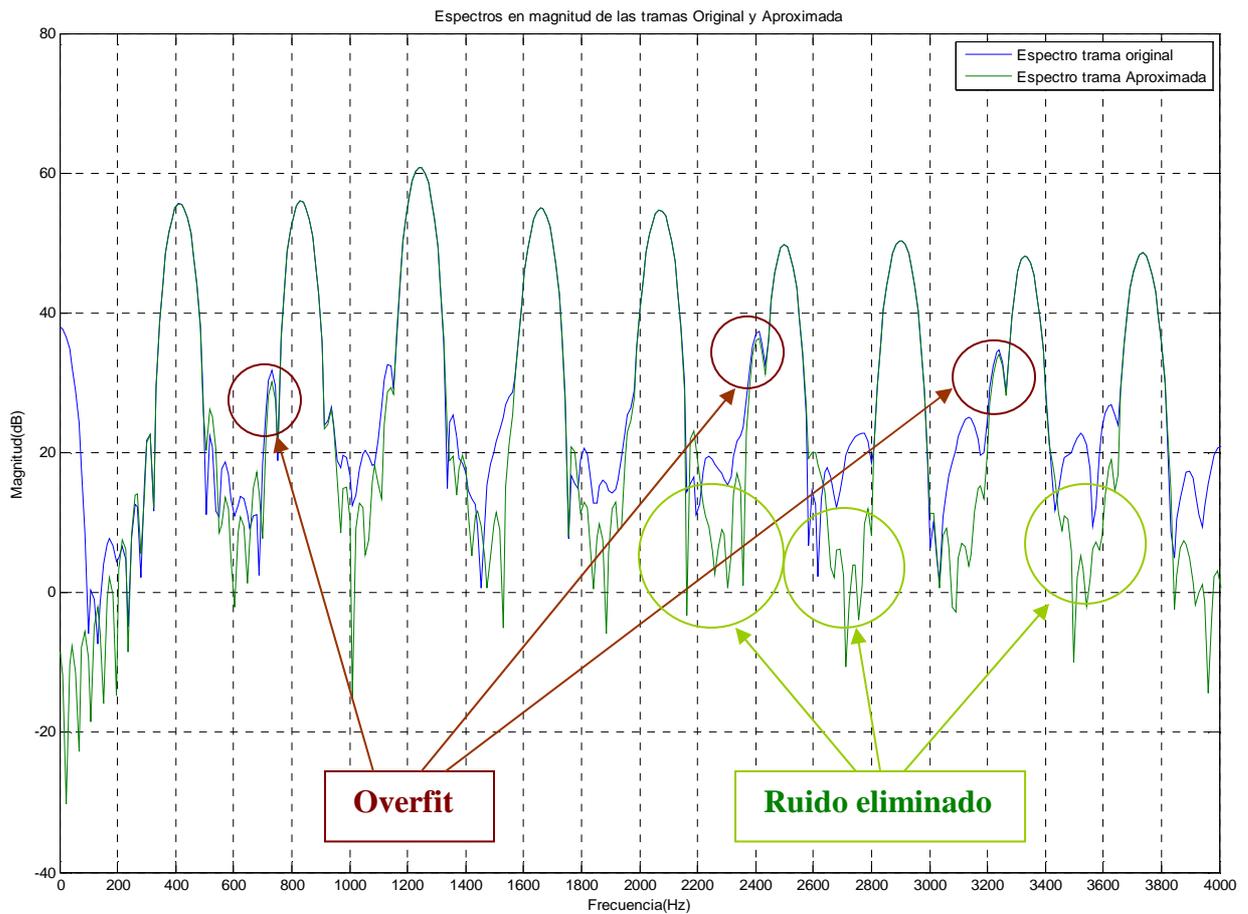


(Fig. 4.1.1.1.3 Espectros de la Trama 1 original y aproximada, $P=6$, $N_{harm}=28$)

Se observa en la trama aproximada como **se ha reducido parte del ruido de la trama original**, los valles son mas profundos y los parciales están mas definidos (existe un margen promedio de unos 40 decibelios frente a los 30 de la original) **el método de modelado de la parte determinista está funcionando.**

En la trama aproximada se puede observar además que no existe la componente DC de la original, la considera parte del ruido.

Ampliamos más para estudiar la presencia de **Overfit** en la trama aproximada:



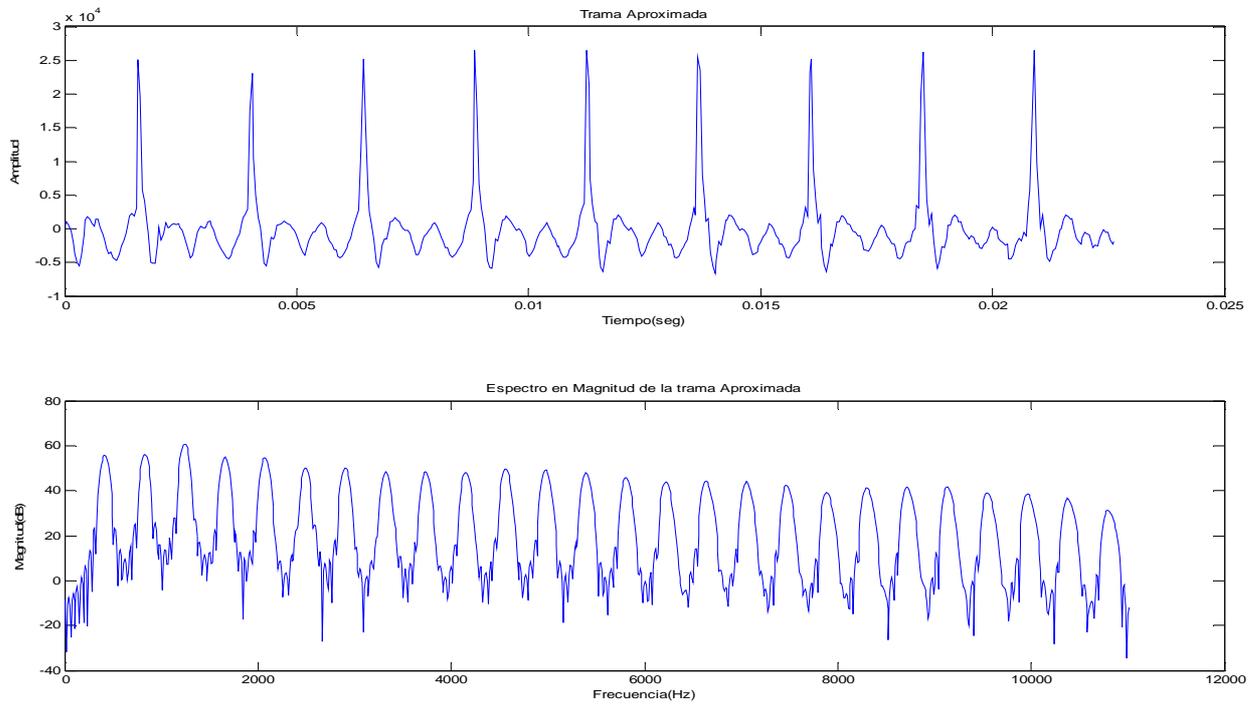
(Fig. 4.1.1.1.4 Espectros ampliados de la Trama 1 original y aproximada, $P=6$, $N_{harm}=28$)

Se observa como a pesar de haber eliminado en la aproximación gran cantidad de ruido de la trama original, debido al alto grado del polinomio de aproximación (**grado 6**) se produce el llamado efecto **Overfit** o de **Sobreestima**, en el que el **polinomio comienza a modelar pulsos adosados a los parciales que corresponden a ruido** y no a la parte determinista (en el tiempo se modelaría rizado).

Veremos que sucede si sobre la misma trama bajamos el grado del polinomio de aproximación.

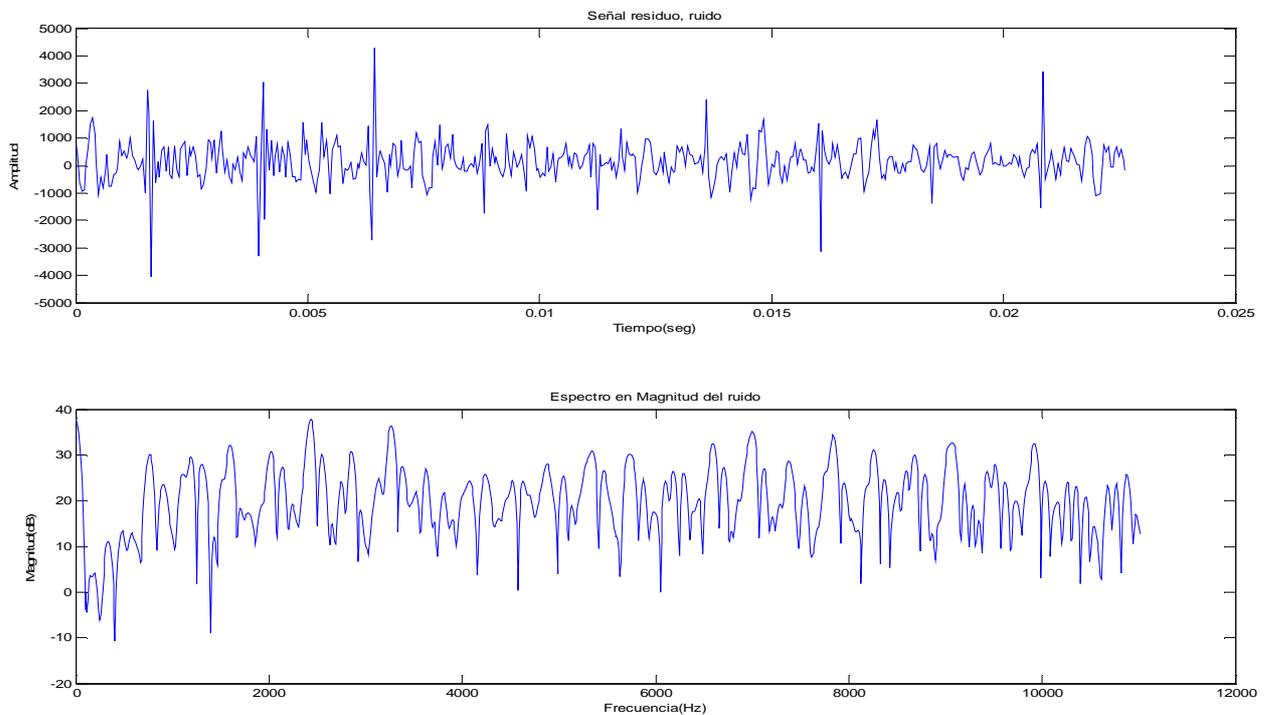
4.1.1.3 - Grado del polinomio= 3, número de armónicos=28.

Obtenemos la siguiente trama aproximada en el tiempo y su espectro:



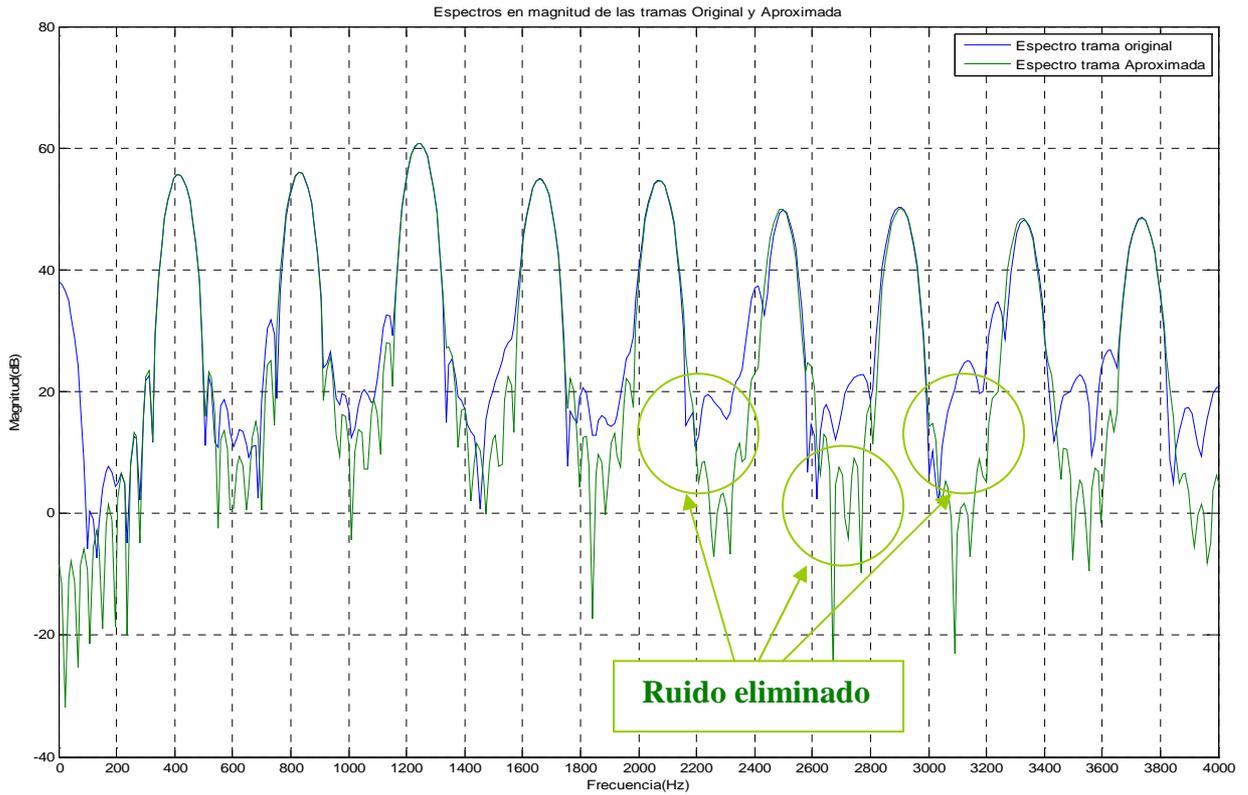
(Fig. 4.1.1.3.1 Trama 1 Modelada polinomialmente, $P=3$, $N_{harm}=28$)

Para el residuo tenemos:

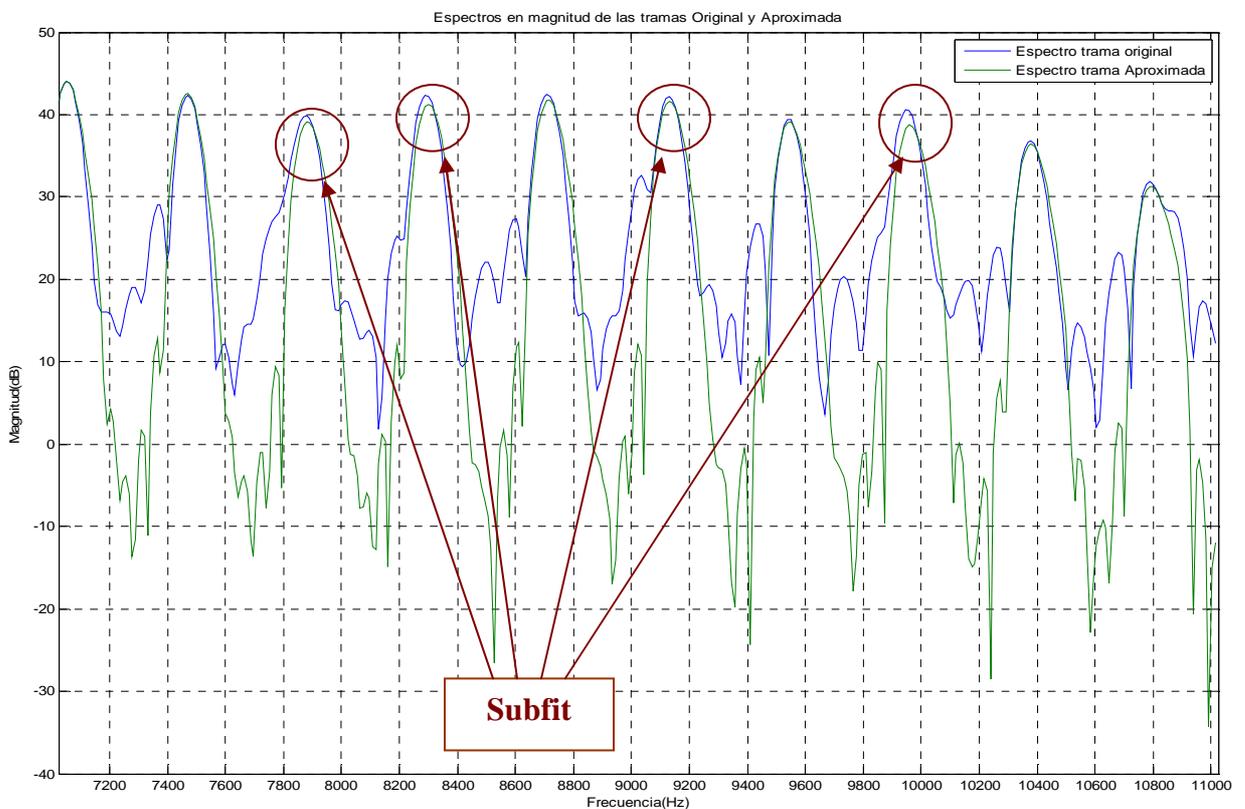


(Fig. 4.1.1.3.2 Residuo o ruido de la Trama 1, $P=3$, $N_{harm}=28$)

Ampliamos los espectros de las tramas original y aproximada:



(Fig. 4.1.1.3.3 Espectros de la Trama 1 original y aproximada, $P=3$, $N_{harm}=28$, Baja frecuencia)



(Fig. 4.1.1.3.4 Espectros de la Trama 1 original y aproximada, $P=3$, $N_{harm}=28$, Alta frecuencia)

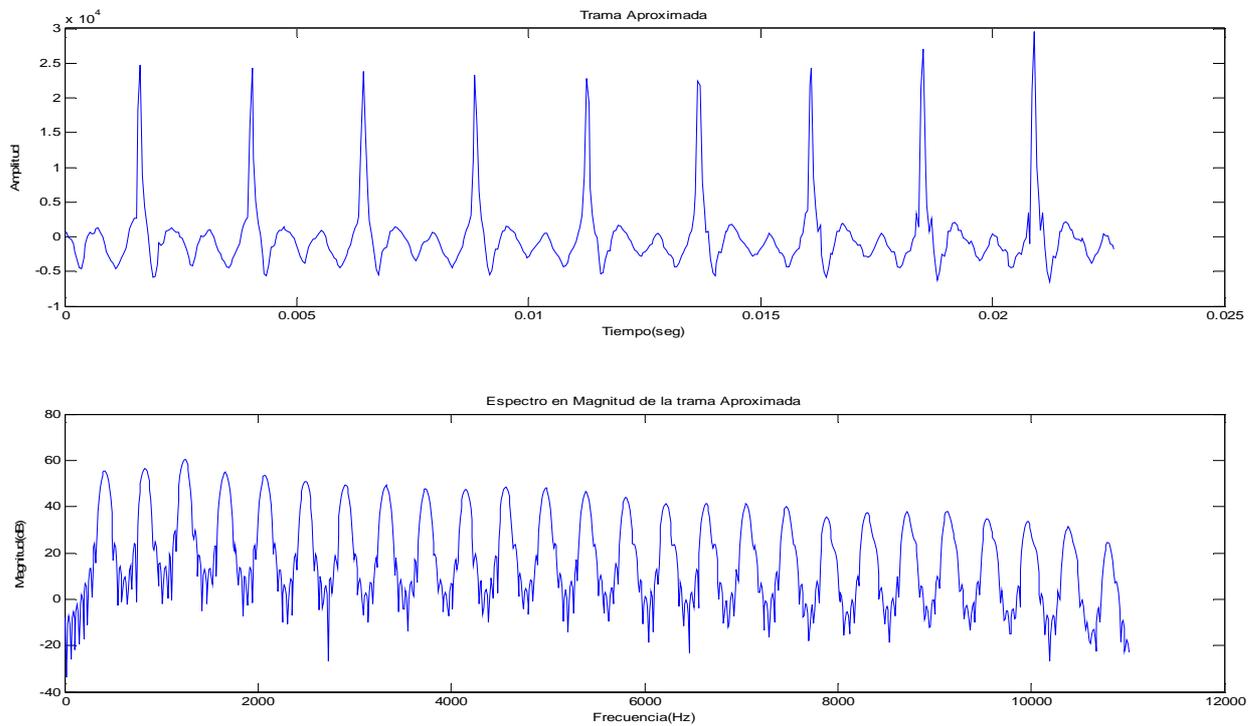
Con **grado 3** en el polinomio de aproximación para esta trama y en baja frecuencia (hasta 4000 Hz) se modela la parte determinista correctamente, se elimina gran cantidad de ruido en la aproximación (en torno a 20 dB) y **no se produce Overfit**.

En cambio los parciales en altas frecuencias son modelados con un **ligero Subfit**, algunos de estos parciales **no llegan a su nivel máximo** y en algunos casos la **forma esta deformada** ligeramente. **El Subfit se manifiesta más en los parciales altos** ya que la modulación FM es más patente en altas frecuencias. En este caso nuestra trama no tenía apenas modulación AM-FM pero aún así el polinomio de grado 3 no llega a seguir del todo dicha modulación en los parciales altos, aunque **el efecto Subfit producido en esta trama es muy ligero y puede despreciarse**.

Veremos que sucede si seguimos bajando el grado del polinomio de aproximación.

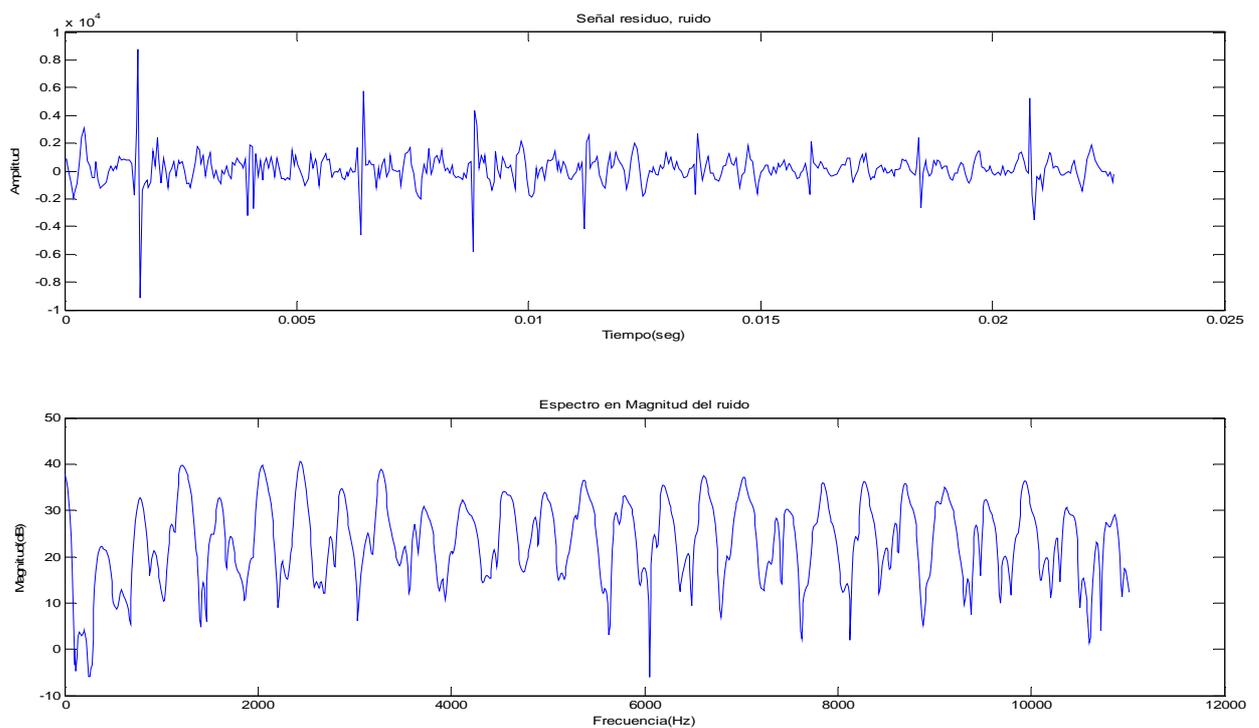
4.1.1.4 - Grado del polinomio= 2, número de armónicos=28.

Obtenemos la siguiente trama aproximada y su espectro:



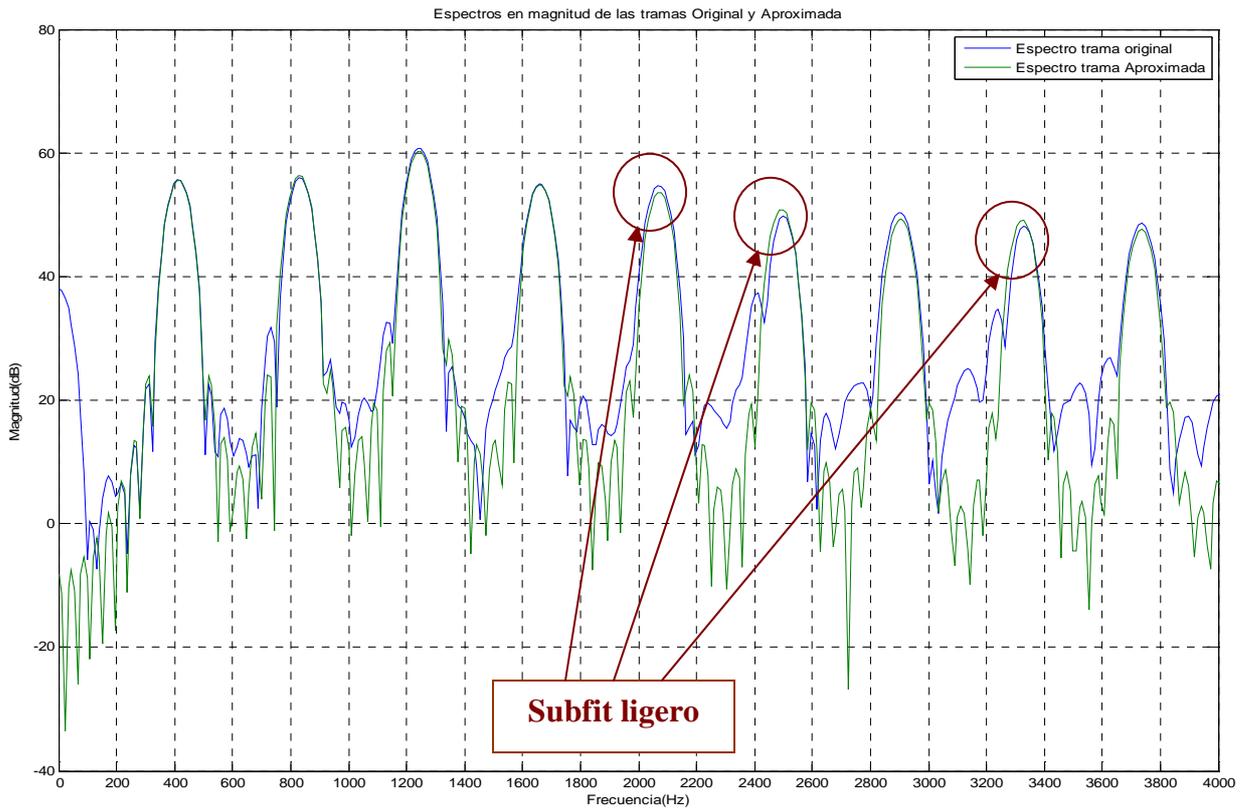
(Fig. 4.1.1.4.1 Trama 1 Modelada polinomialmente, $P=2$, $N_{harm}=28$)

Para el residuo tenemos:

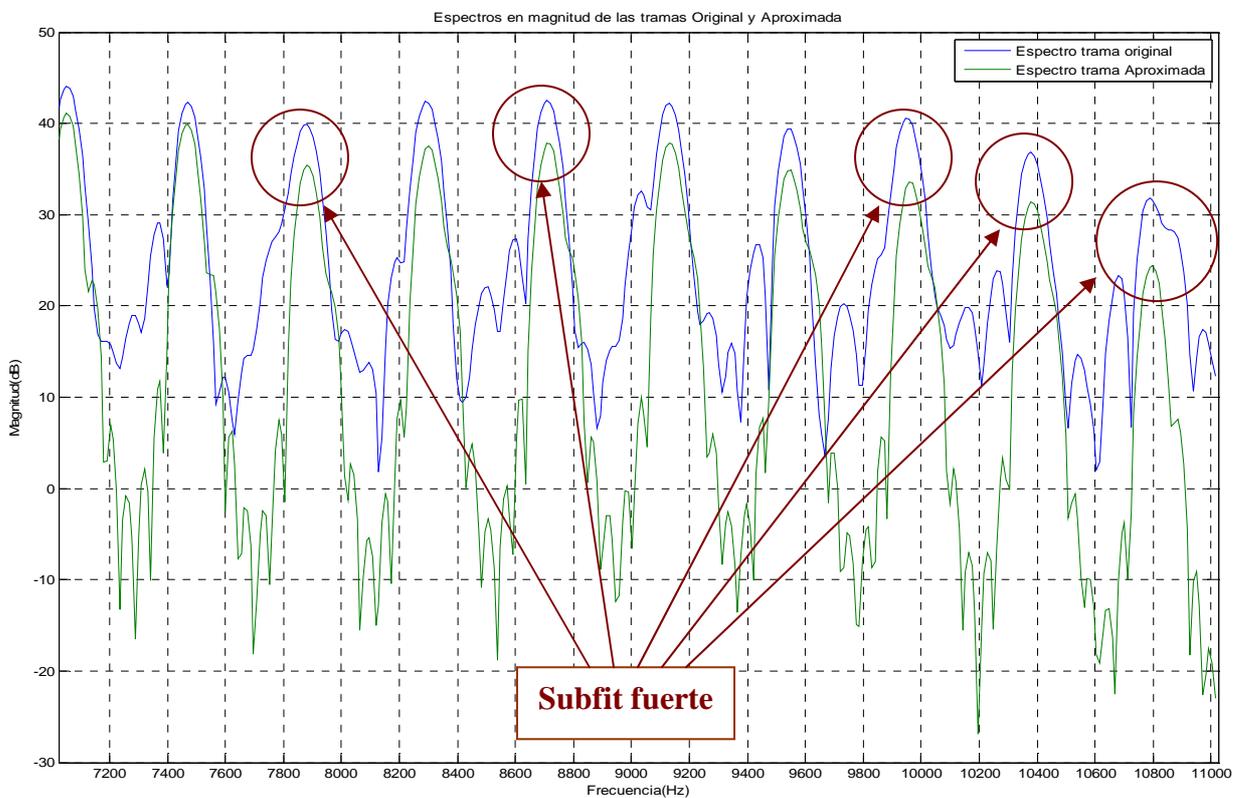


(Fig. 4.1.1.4.2 Residuo o ruido de la Trama 1, $P=2$, $N_{harm}=28$)

Ampliamos los espectros de las tramas original y aproximada:



(Fig. 4.1.1.4.3 Espectros de la Trama 1 original y aproximada, $P=2$, $N_{harm}=28$, Baja frecuencia)



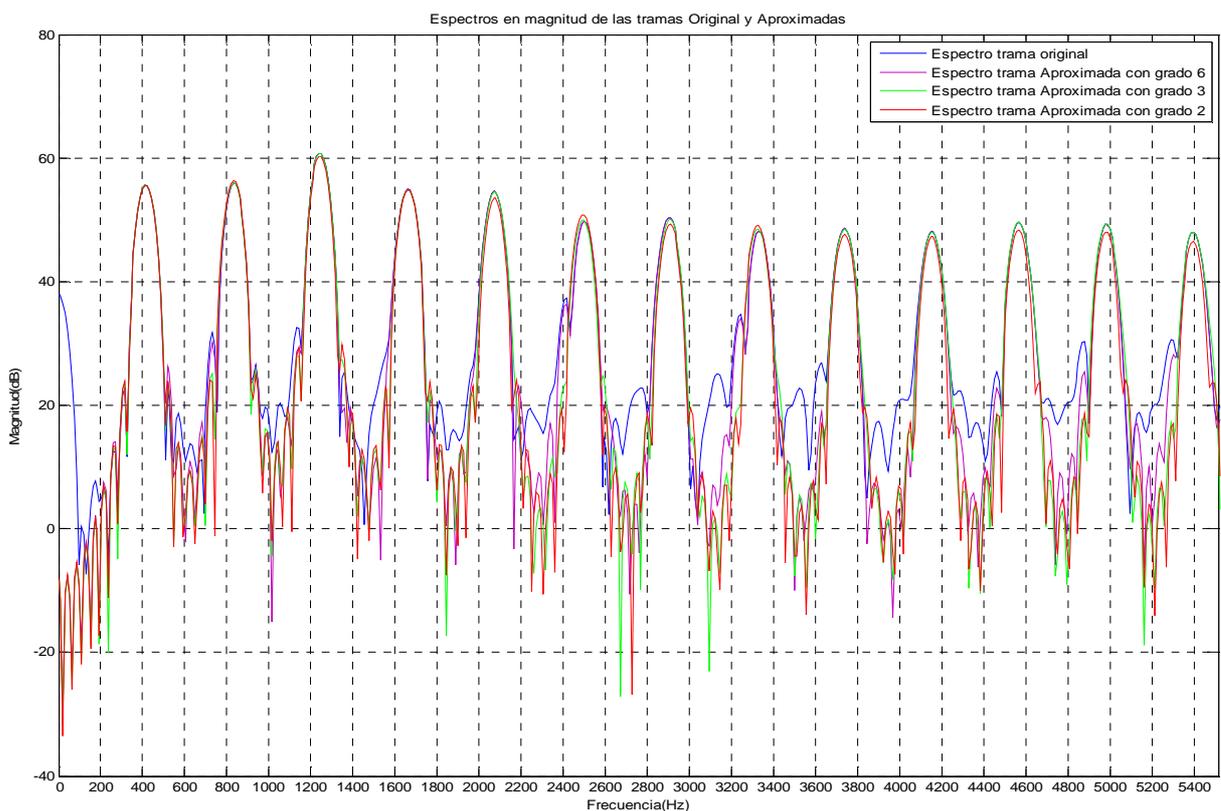
(Fig. 4.1.1.4.4 Espectros de la Trama 1 original y aproximada, $P=2$, $N_{harm}=28$, Alta frecuencia)

En este caso con **grado 2** en el polinomio de aproximación se observa como se elimina gran parte de ruido pero existe **Subfit en todas las frecuencias**, siendo **mayor en los parciales en altas frecuencias** (la modulación FM es más pronunciada en altas frecuencias).

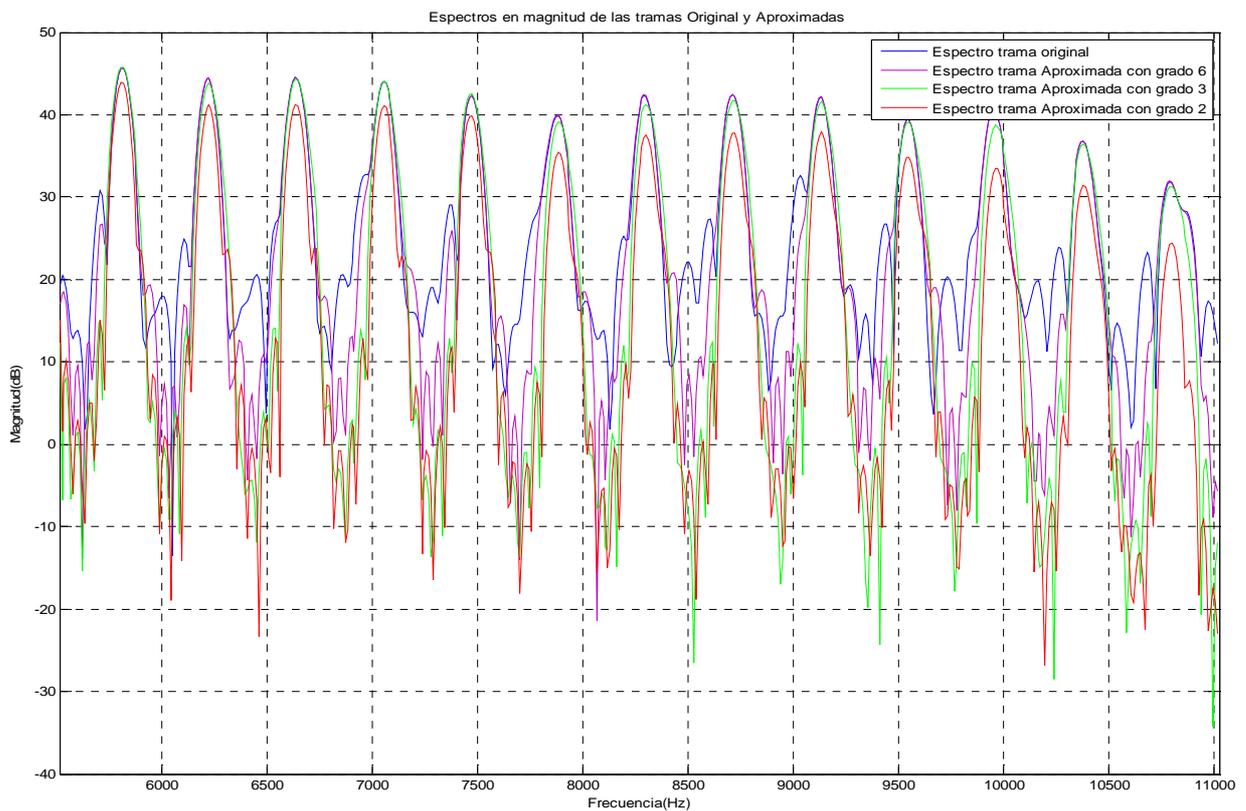
Pese a ser una trama con poca modulación AM-FM, el polinomio de grado 2 no es capaz de seguir correctamente las variaciones en amplitud y frecuencia presentes en la trama y se traduce en una **deformación en la forma y el nivel máximo de los parciales**, por tanto **no se modela bien la parte determinista**.

4.1.1.5 - Conclusiones al modelado polinomial de la Trama 1:

Representamos los espectros de las tramas aproximadas con los distintos grados del polinomio junto al de la trama original para comparar mejor los resultados:



(Fig. 4.1.1.5.1 Espectros de la Trama 1 original y aproximadas, $P=6, 3$ y 2 $N_{harm}=28$, Baja frecuencia)



(Fig. 4.1.1.5.2 Espectros de la Trama 1 original y aproximadas. $P=6, 3$ y 2 $N_{harm}=28$, Alta frecuencia)

Se concluye de este estudio que **para este tipo de tramas, con poco ruido y poca modulación AM-FM, lo adecuado será aproximar con modelado polinomial de grado 3** y un número de armónicos alto, aunque este último parámetro dependerá también del tipo de instrumento y de su envolvente espectral concreta. Con dicho grado de polinomio obtenemos **buenos resultados en cuanto a eliminación de ruido** (en torno a 20 dB) y modelamos los parciales correctamente. Con grado 3 el **Subfit producido en los parciales altos es muy ligero, despreciable**.

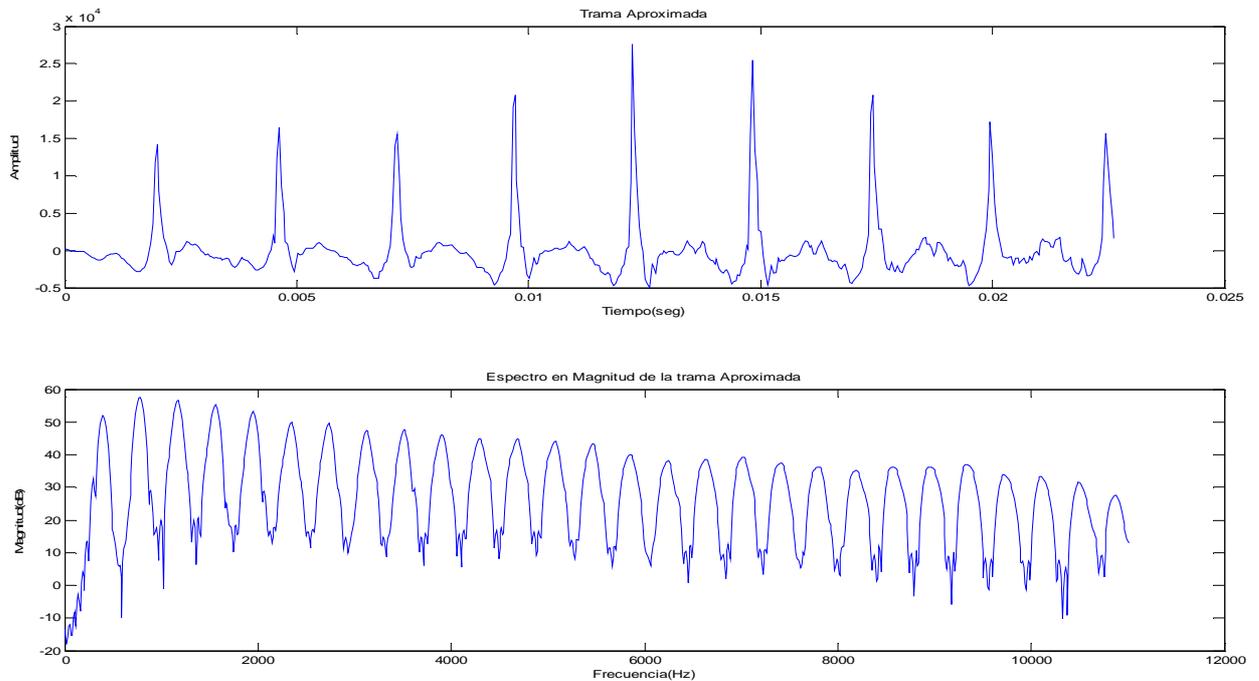
TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3

4.1.2 MODELADO POLINOMIAL TRAMA 2:

Trama cuasi estacionaria poco ruidosa con fuerte modulación. Trumpet

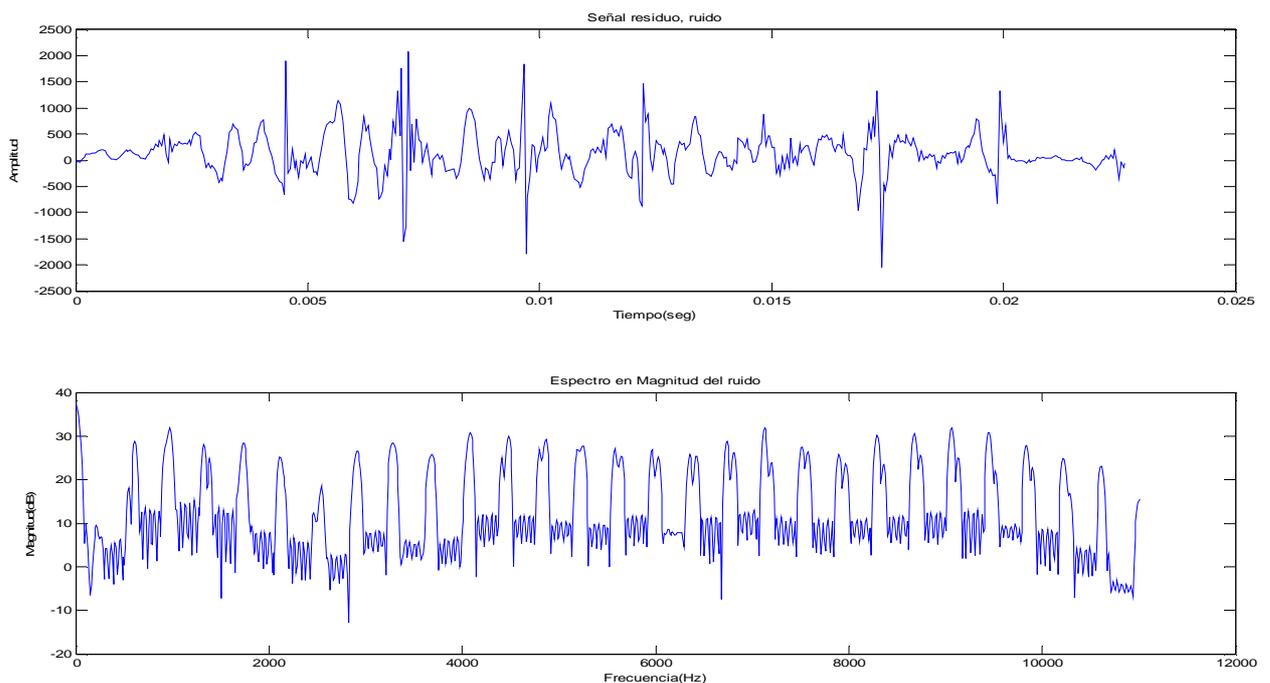
4.1.2.1 - Grado del polinomio= 6, número de armónicos=28.

Representamos la trama aproximada polinomialmente y su espectro:



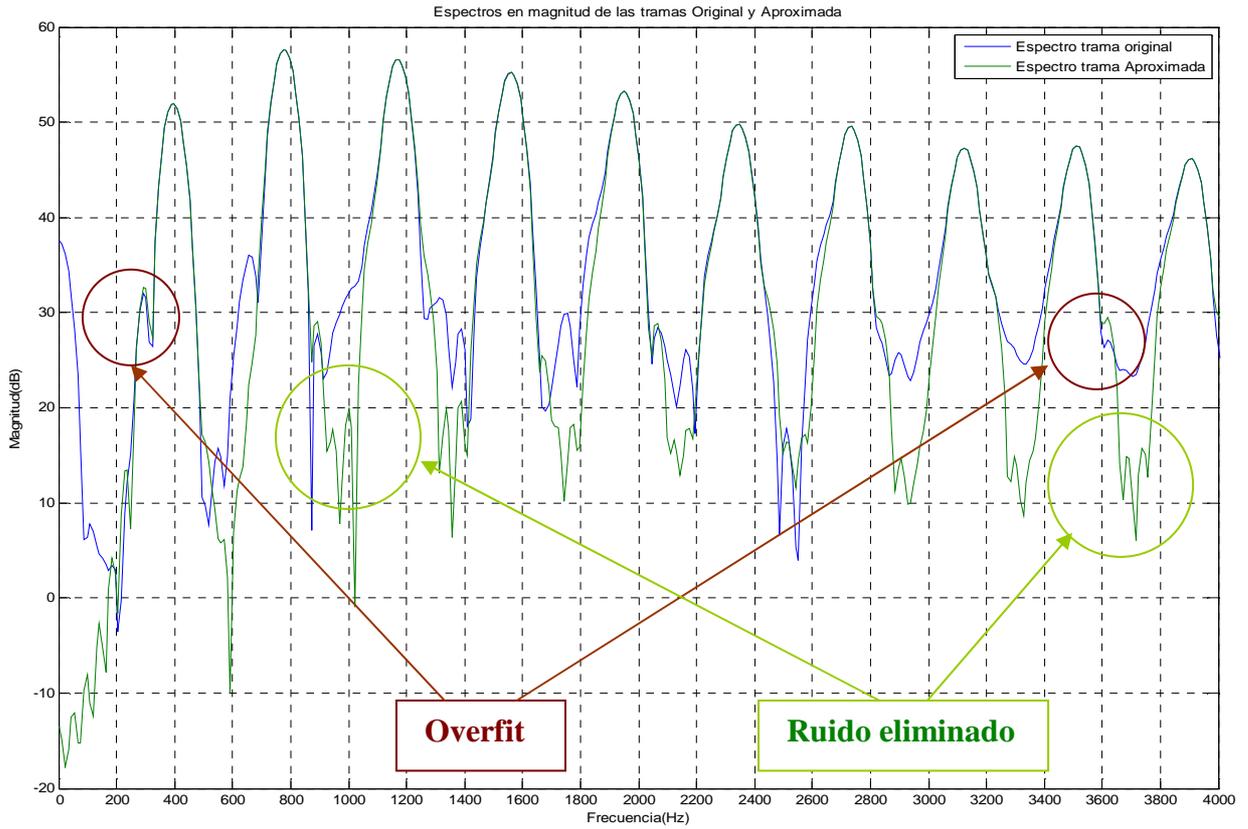
(Fig. 4.1.2.1.1 Trama 2 Modelada polinomialmente, $P=6$, $N_{harm}=28$)

Para el residuo obtenemos lo siguiente:

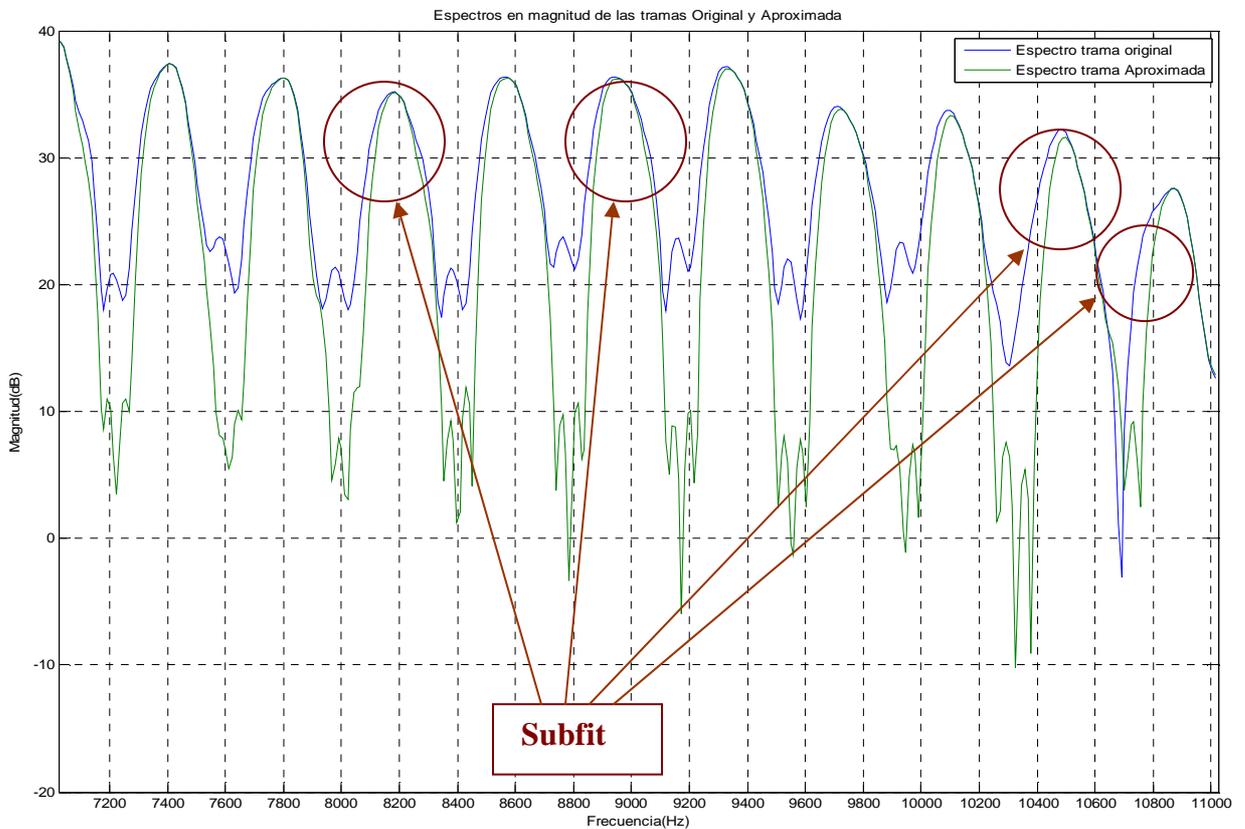


(Fig. 4.1.2.1.2 Residuo o ruido de la Trama 2, $P=6$, $N_{harm}=28$)

Ampliamos los espectros de las tramas original y aproximada:



(Fig. 4.1.2.1.3 Espectros de la Trama 2 original y aproximada, $P=6$, $N_{harm}=28$, Baja frecuencia)



(Fig. 4.1.2.1.4 Espectros de la Trama 2 original y aproximada, $P=6$, $N_{harm}=28$, Alta frecuencia)

Se observa en el espectro de la trama aproximada como se ha reducido parte del ruido de la trama original, los valles son mas profundos sobre todo en altas frecuencias y los parciales están mas definidos. Existe un **margen promedio entre el nivel de los parciales y el ruido de unos 40 decibelios frente a los 30 de la original.**

La componente DC de la trama original es eliminada junto al ruido, no aporta información a la señal y no estará en la parte determinista.

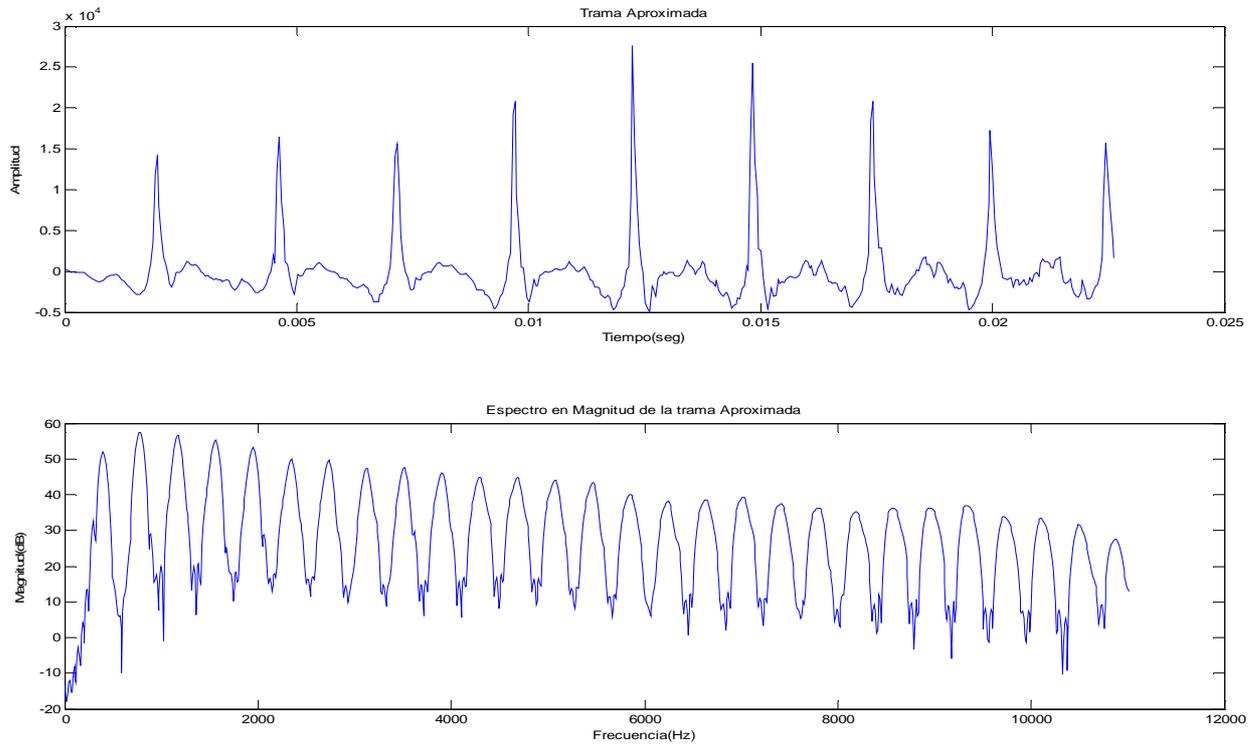
Al modelar con **grado 6 se produce un ligero Overfit despreciable** en bajas frecuencias.

Incluso con grado 6 en el polinomio de aproximación no se consigue seguir del todo bien la variación en amplitud asociada a la modulación. La forma con la que se modelan los **parciales en alta frecuencia presenta cierta distorsión** y no se llega al nivel máximo de los parciales originales. Se produce **Subfit** aunque es **muy ligero y se puede despreciar.**

Veremos que sucede si sobre la misma trama subimos el grado del polinomio de aproximación a un valor aún no estudiado. Nos planteamos si al aumentar el grado a un valor mayor que 6 será capaz de modelar correctamente la modulación de la trama.

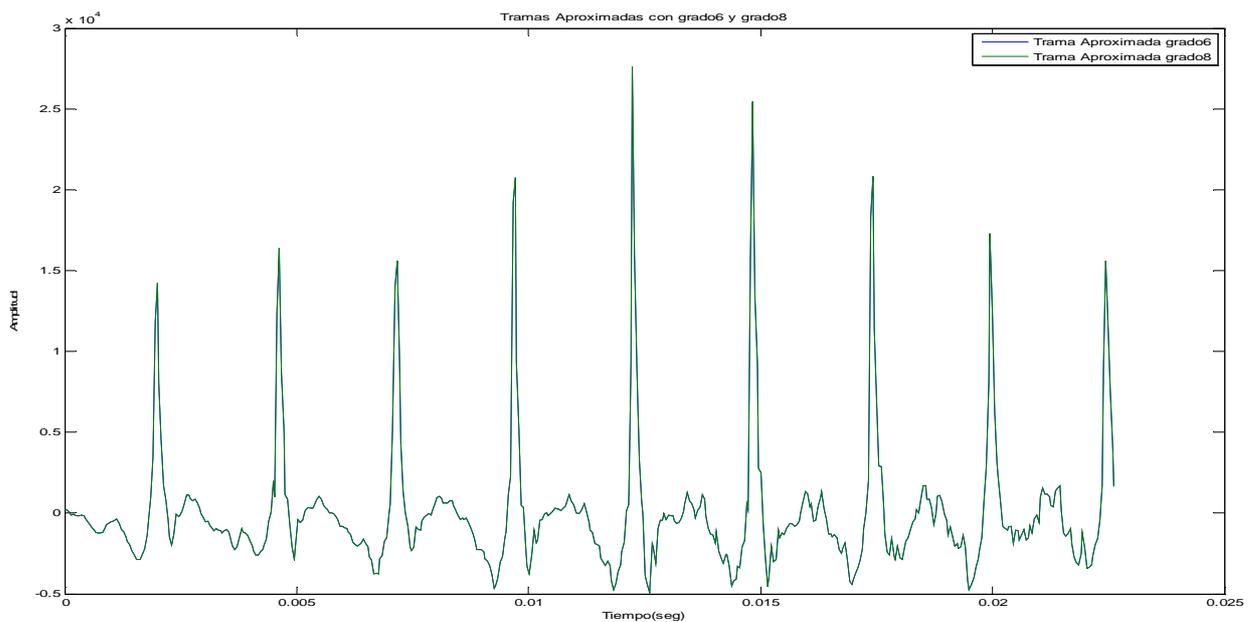
4.1.2.2 - Grado del polinomio= 8, número de armónicos= 28.

Obtenemos la siguiente trama aproximada y su espectro:



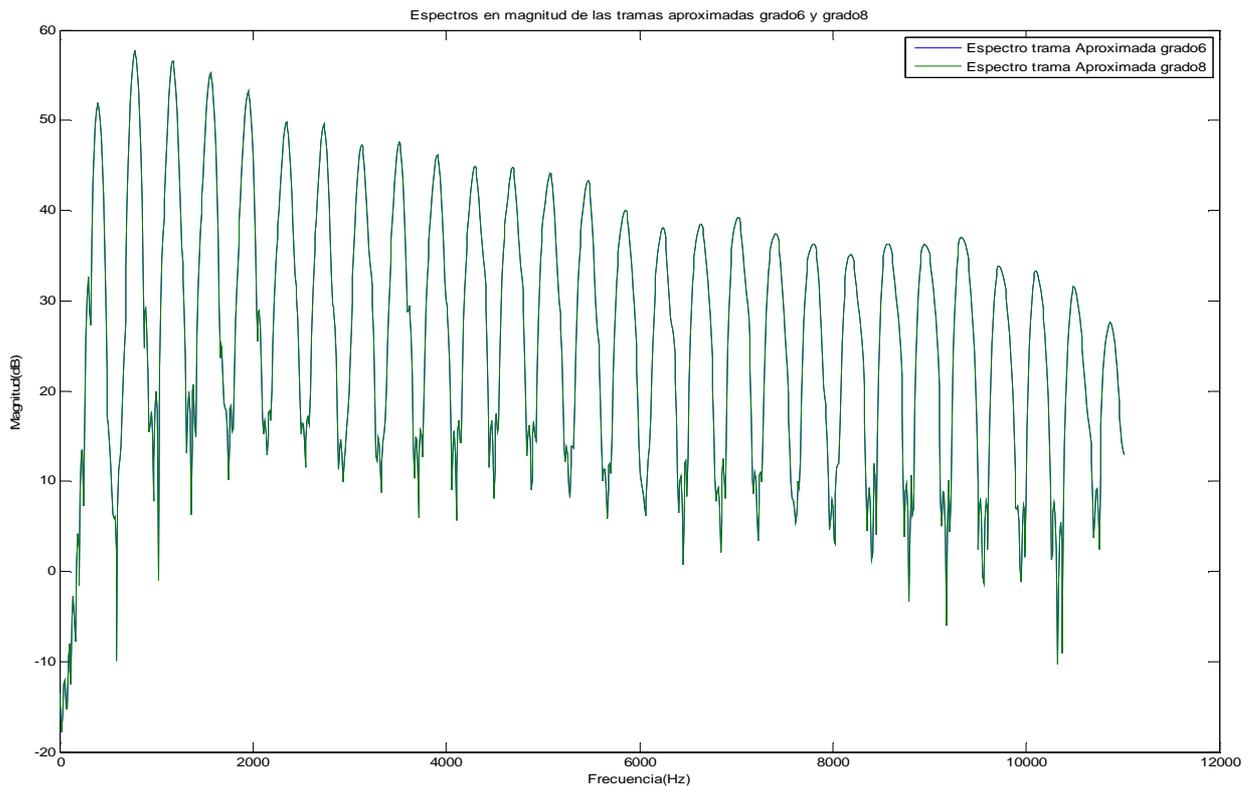
(Fig. 4.1.2.2.1 Trama 2 Modelada polinomialmente, $P=8$, $N_{harm}=28$)

Representamos juntas las tramas aproximadas con grado 6 y grado 8 en el dominio temporal:



(Fig. 4.1.2.2.2 Trama 2 Modelada polinomialmente, $P=6$ y 8 , $N_{harm}=28$)

Comparando los espectros de ambas aproximaciones:



(Fig, 4.1.2.2.3 Espectros de la Trama 2 aproximada, $P=6$ y 8 , $N_{harm}=28$)

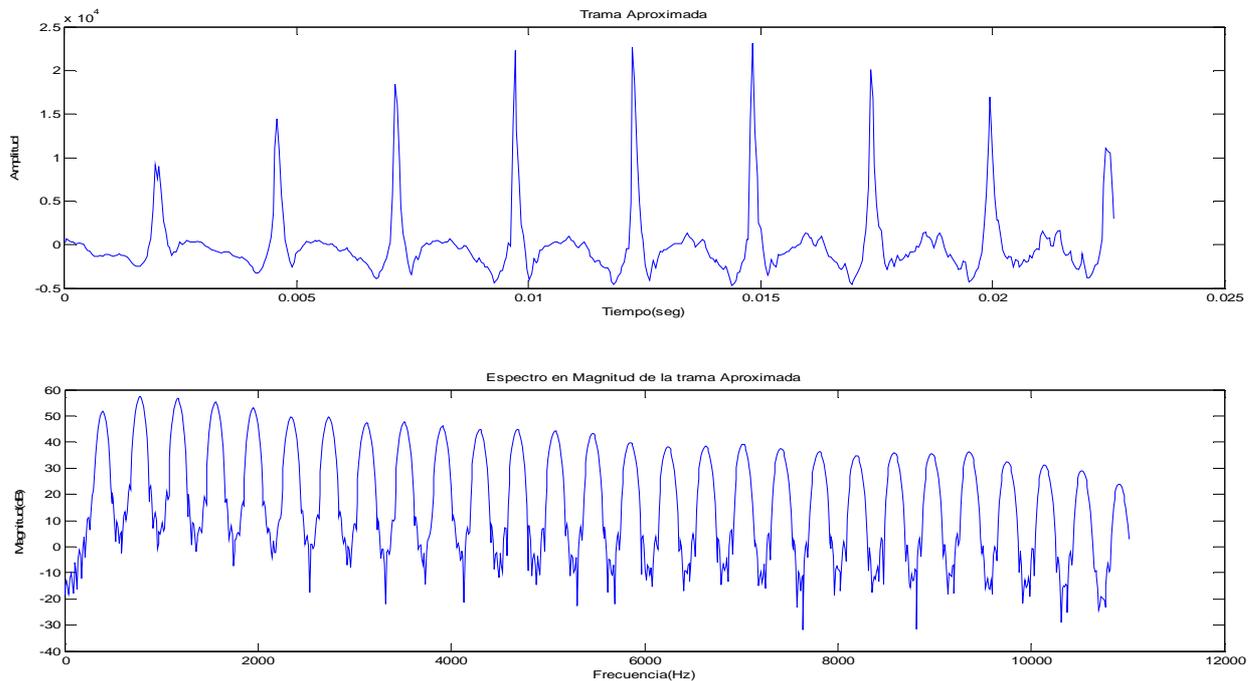
Se observa tanto en el tiempo como en la frecuencia como **al aproximar con grado de polinomio 8 se obtiene la misma trama que al aproximar con grado 6**, se produce una saturación en el modelado. Nos quedamos con el de grado 6 que es más rápido computacionalmente. Se ha probado con grado 7 y se obtienen los mismos resultados.

El polinomio de aproximación tiene un límite superior a partir del cual se satura y no se obtienen mejores resultados, este límite es grado 6.

Veremos que sucede si bajamos el grado en el polinomio de aproximación en cuanto a la eliminación de ruido.

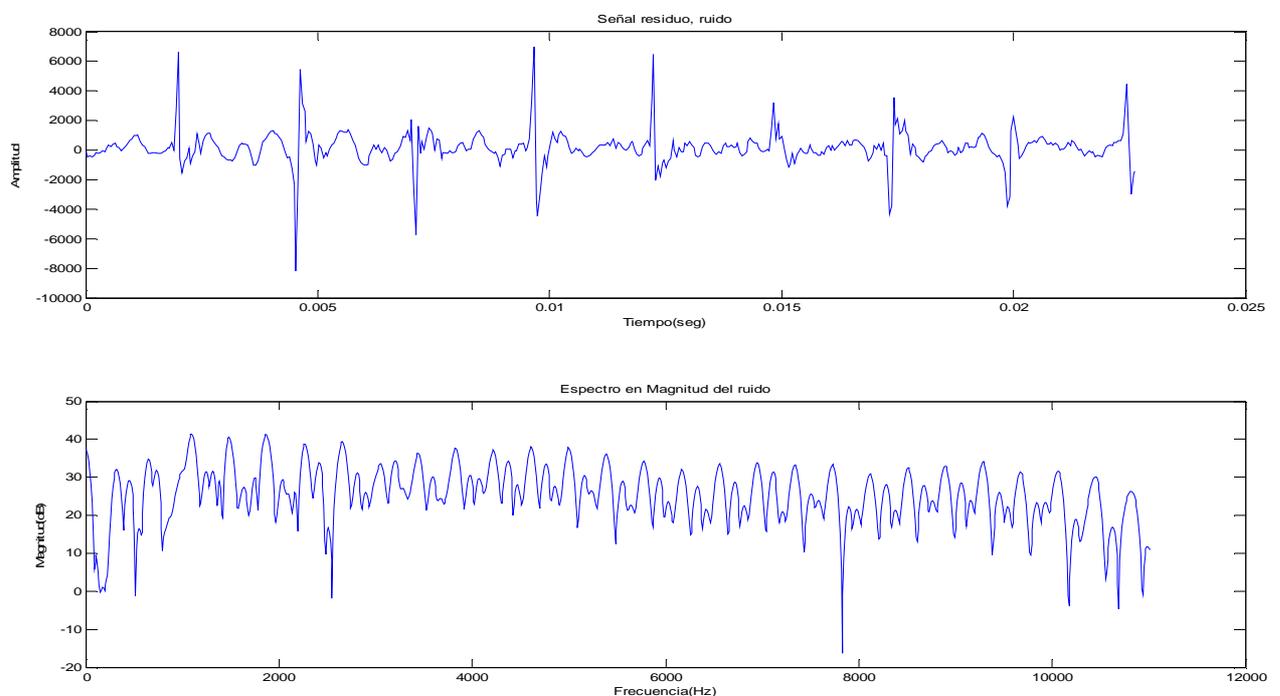
4.1.2.3 - Grado del polinomio= 3, número de armónicos= 28.

Obtenemos la siguiente trama aproximada y su espectro:



(Fig. 4.1.2.3.1 Trama 2 Modelada polinomialmente, $P=3$, $N_{harm}=28$)

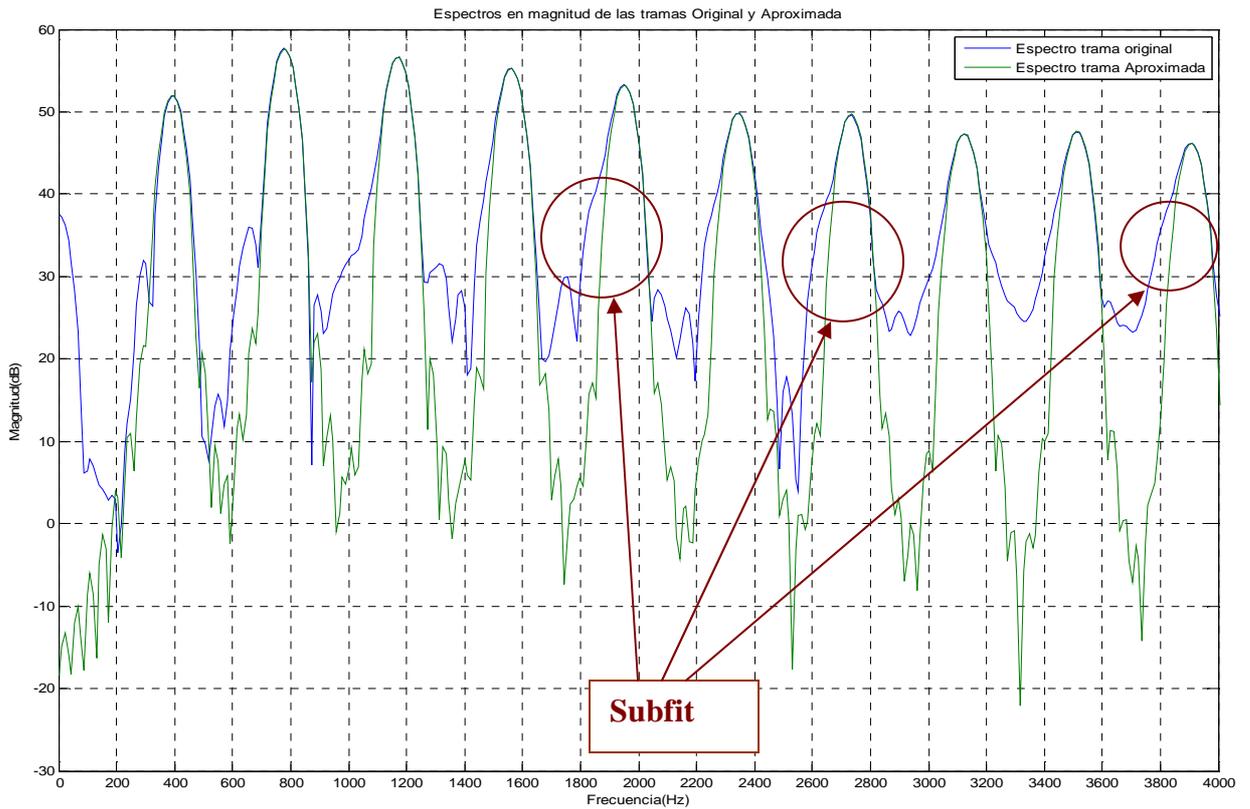
Para el residuo tenemos:



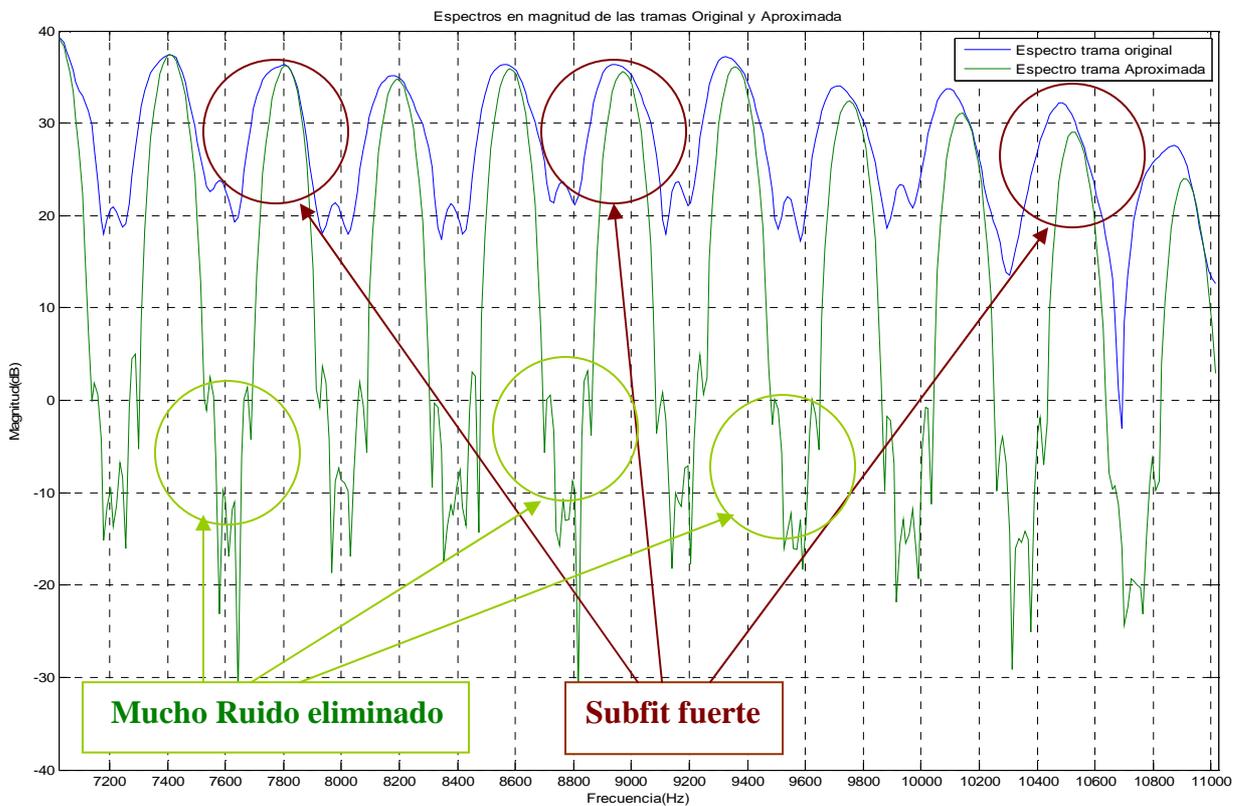
(Fig. 4.1.2.3.2 Residuo o ruido de la Trama 2, $P=3$, $N_{harm}=28$)

En el dominio temporal para la trama aproximada se observa como con **grado 3** en el polinomio **no se consigue modelar bien toda la variación en amplitud** asociada a la modulación AM de la trama original.

Ampliamos los espectros de las tramas original y aproximada:



(Fig. 4.1.2.3.3 Espectros de la Trama 2 original y aproximada, $P=3$, $N_{harm}=28$, Baja frecuencia)



(Fig. 4.1.2.3.4 Espectros de la Trama 2 original y aproximada, $P=3$, $N_{harm}=28$, Alta frecuencia)

Con **grado 3** en el polinomio de aproximación no se produce Overfit pero se produce **Subfit en todo el espectro, siendo muy fuerte en los parciales en alta frecuencia.**

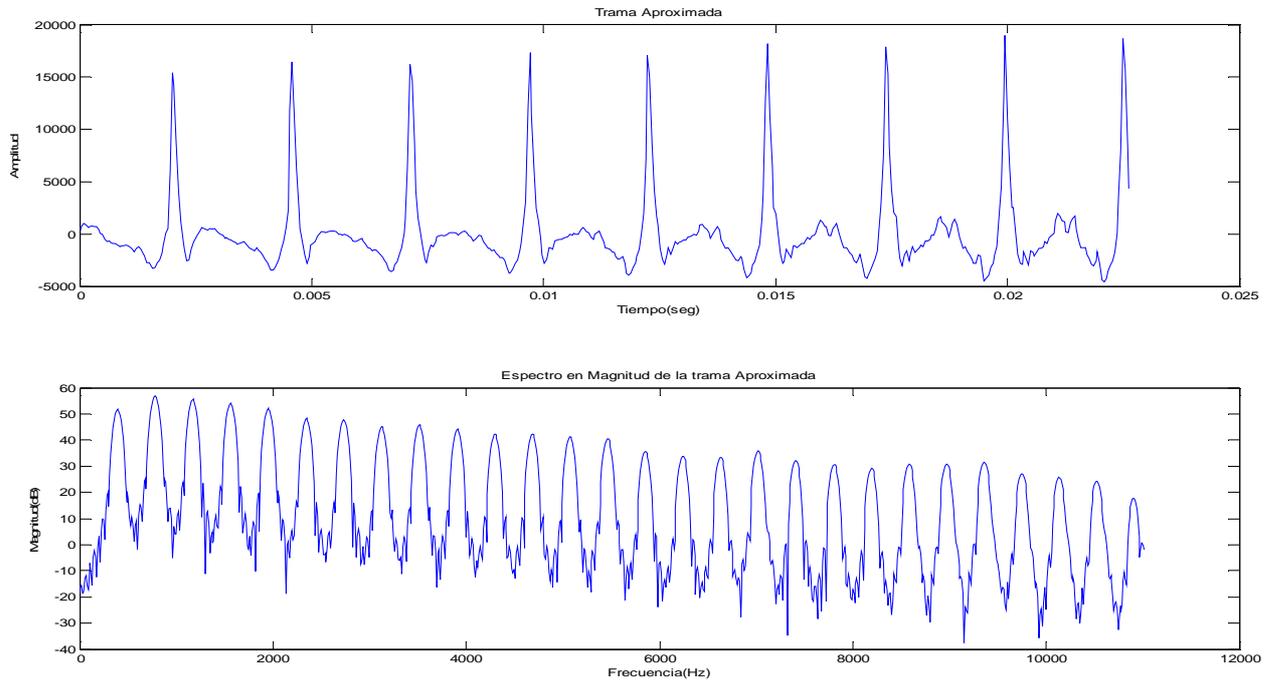
En cuanto al ruido nos damos cuenta que **cuanto más bajo es el grado del polinomio de aproximación más ruido se elimina.** En este caso tenemos un margen promedio de unos **50 decibelios entre el nivel de los parciales y le ruido frente a los 30 de la trama original.**

Como ya se comentó el efecto de Subfit es un problema grave ya que **distorsiona la información de la señal,** debido a ello aunque con grado 3 se consiga eliminar más ruido, no podemos modelar esta trama con un grado tan bajo, ya que **el Subfit producido comienza a ser importante y no se modela bien la parte determinista de la trama.**

Veremos por ultimo que sucede si modelamos con grado 2 en el polinomio de aproximación

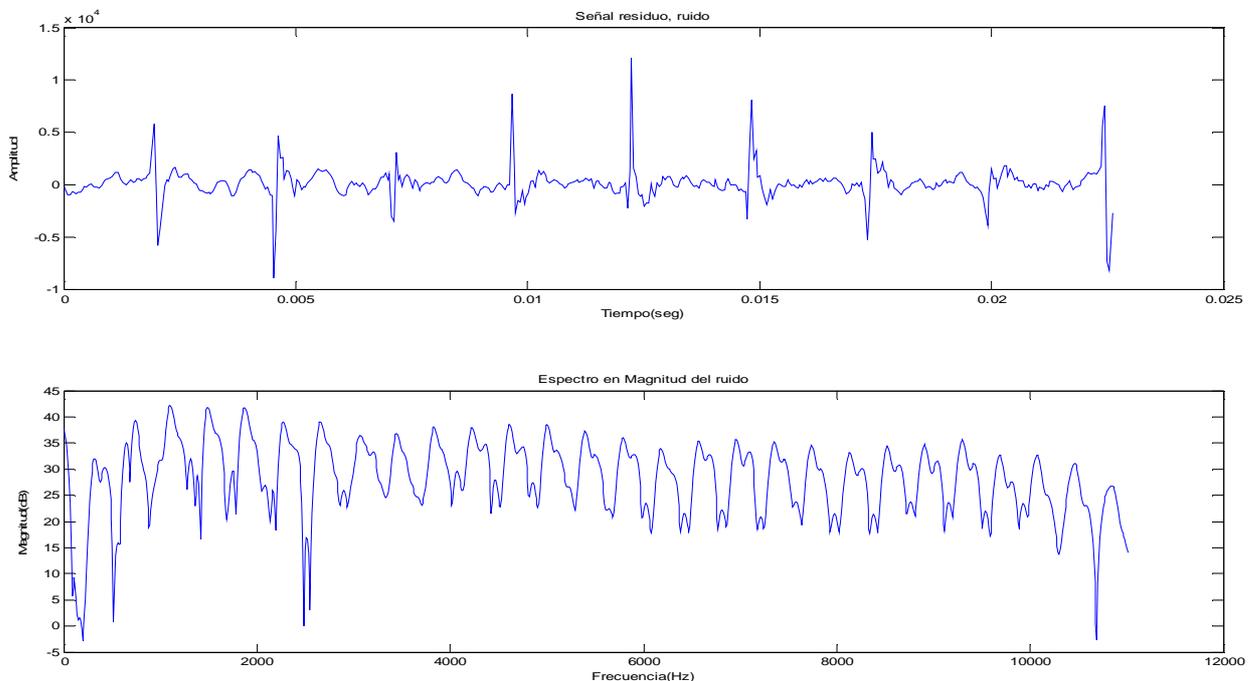
4.1.2.4 - Grado del polinomio= 2, número de armónicos=28.

Obtenemos la siguiente trama aproximada:



(Fig. 4.1.2.4.1 Trama 2 Modelada polinomialmente, $P=2$, $N_{harm}=28$)

Para el residuo tenemos:

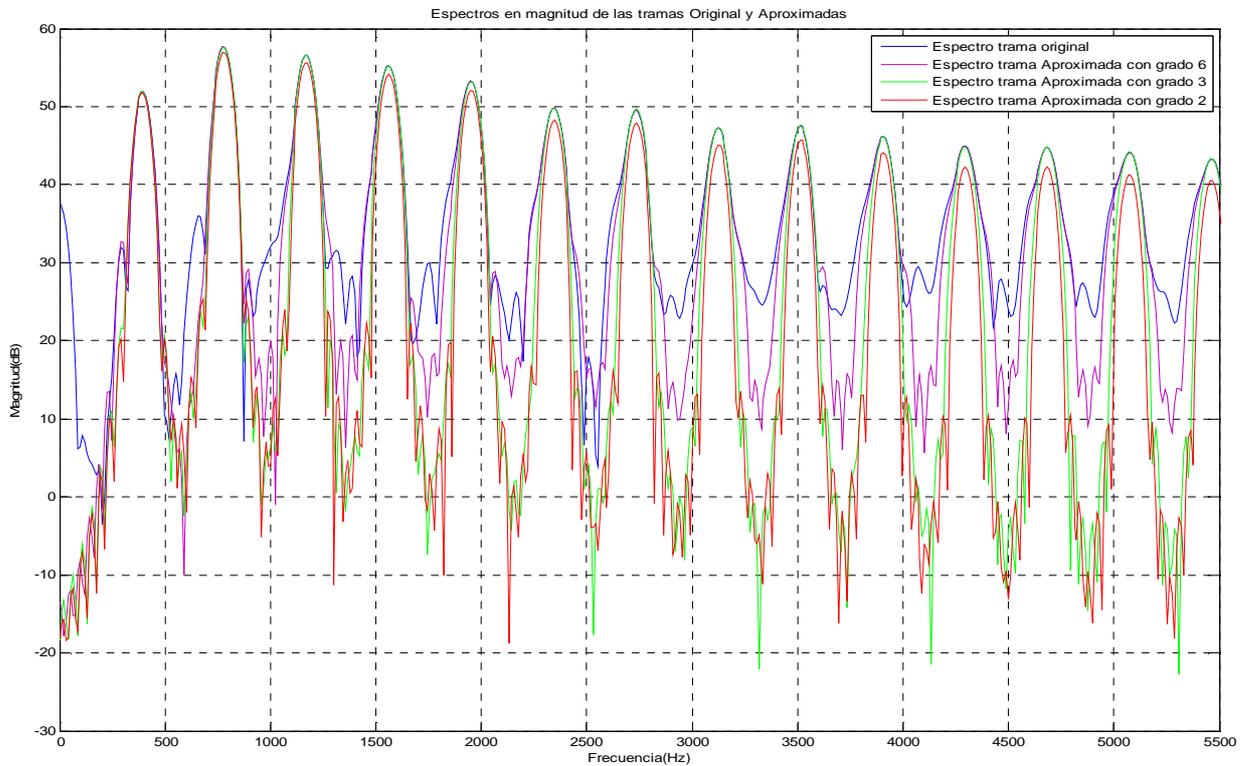


(Fig. 4.1.2.4.2 Residuo o ruido de la Trama 2, $P=2$, $N_{harm}=28$)

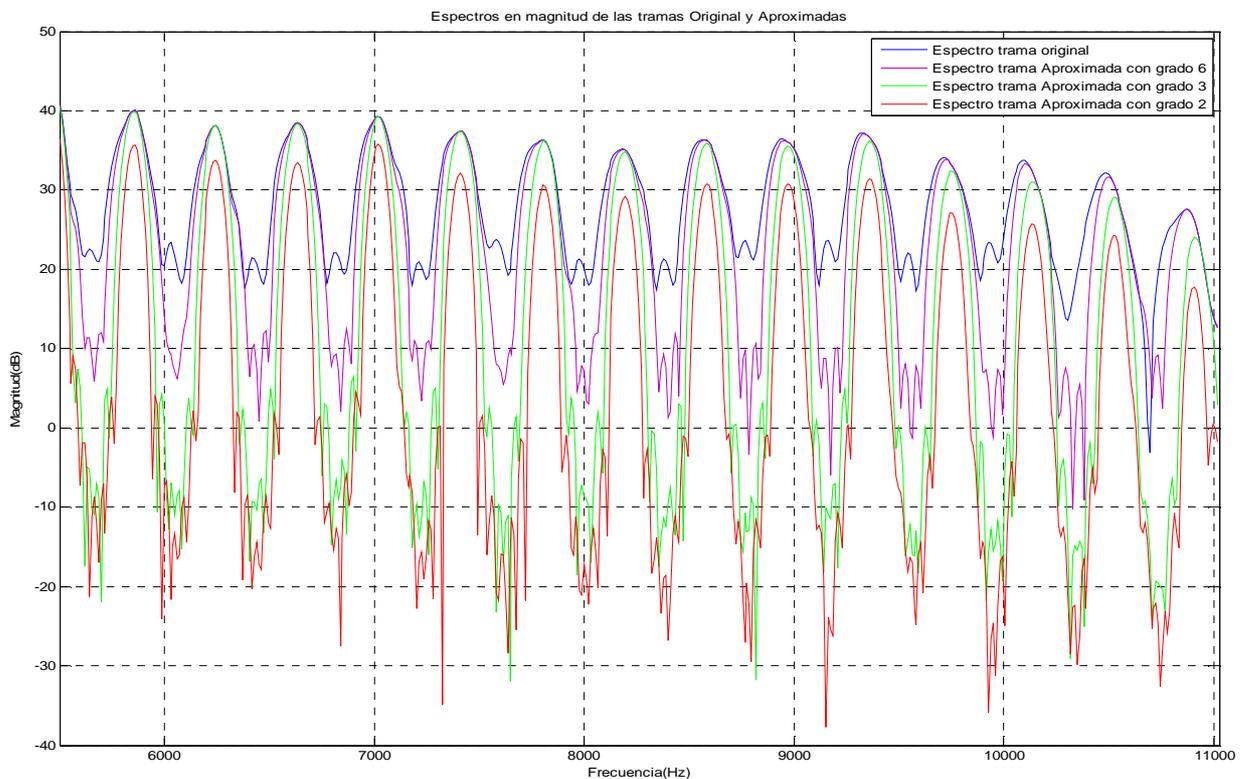
Se observa en el dominio temporal como trama aproximada no consigue modelar la variación en amplitud asociada a la fuerte modulación AM de la trama original y en el residuo tenemos mucha parte determinista. El polinomio de **grado 2 para este tipo de tramas es muy bajo, produce mucho Subfit y no modela bien.**

4.1.2.5 - Conclusiones al modelado polinomial de la Trama 2:

Representamos los espectros de las señales aproximadas junto al de la original para comparar mejor los resultados:



(Fig. 4.1.2.5.1 Espectros de la Trama 2 original y aproximadas, $P=6, 3$ y 2 $N_{harm}=28$, Baja frecuencia)



(Fig. 4.1.2.5.1 Espectros de la Trama 2 original y aproximadas, $P=6, 3$ y 2 $N_{harm}=28$, Alta frecuencia)

Se concluye que **para este tipo de tramas, con poco ruido y fuerte modulación, lo adecuado será aproximar con modelado polinomial de grado 6**, para seguir y modelar mejor la variación asociada a la fuerte modulación. Con dicho grado de polinomio obtenemos **buenos resultados en cuanto a eliminación de ruido** (en torno a 10 dB) y el Subfit producido en alta frecuencia es despreciable.

Grados menores eliminan más ruido pero distorsionan los parciales por el Subfit producido y no modelan bien. **Grados mayores no obtienen mejores resultados** ya que se produce saturación en el modelado.

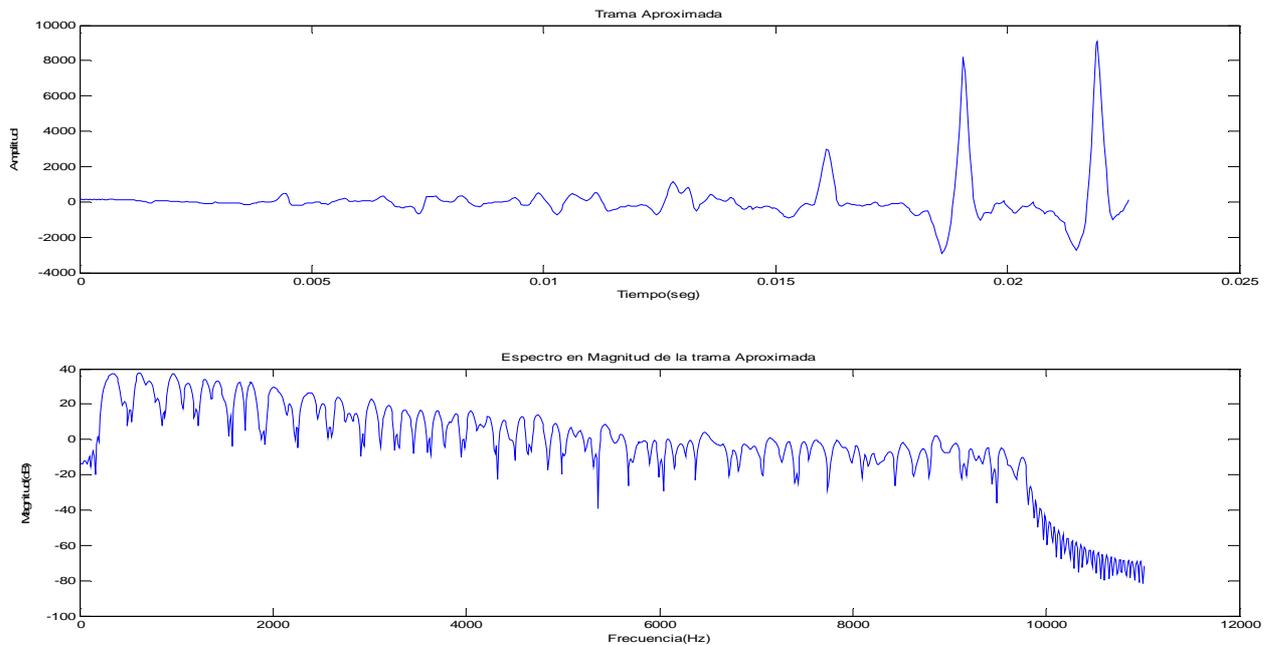
TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3
Mucha modulación y poco ruido	Modelado no estacionario polinomial grado 6

4.1.3 MODELADO POLINOMIAL TRAMA 3:

Trama transitoria poco ruidosa. Trumpet

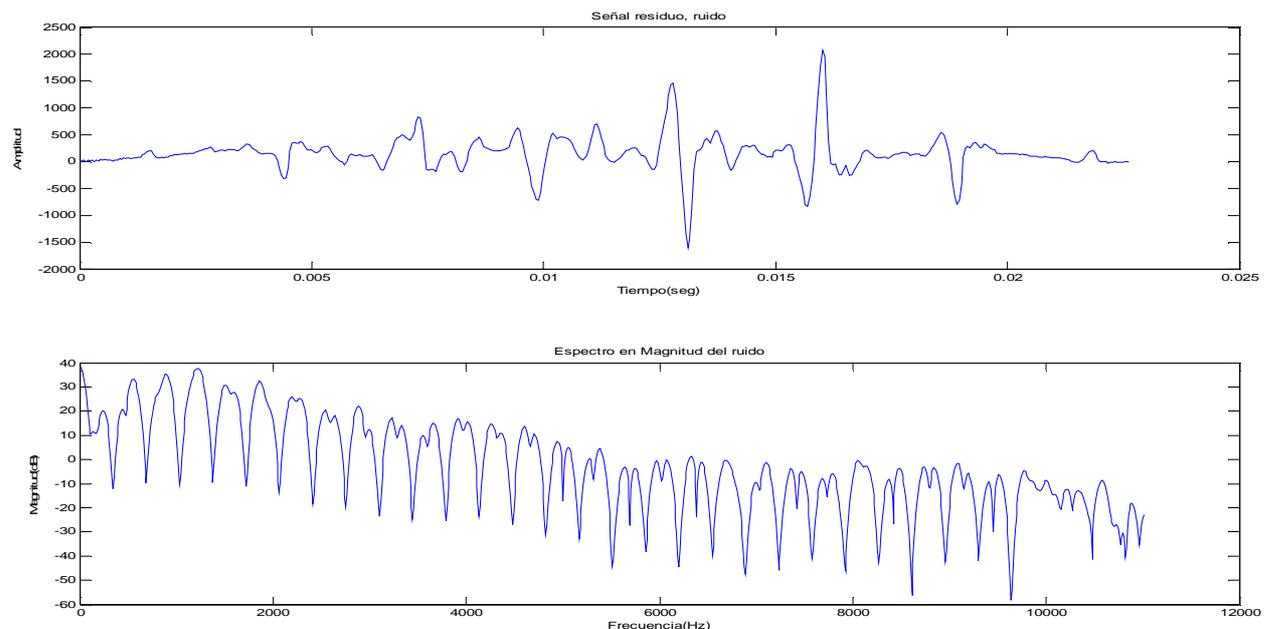
4.1.3.1 - Grado del polinomio= 6, número de armónicos=28.

Representamos la trama aproximada polinomialmente y su espectro:



(Fig. 4.1.3.1.1 Trama 3 Modelada polinomialmente, $P=6$, $N_{harm}=28$)

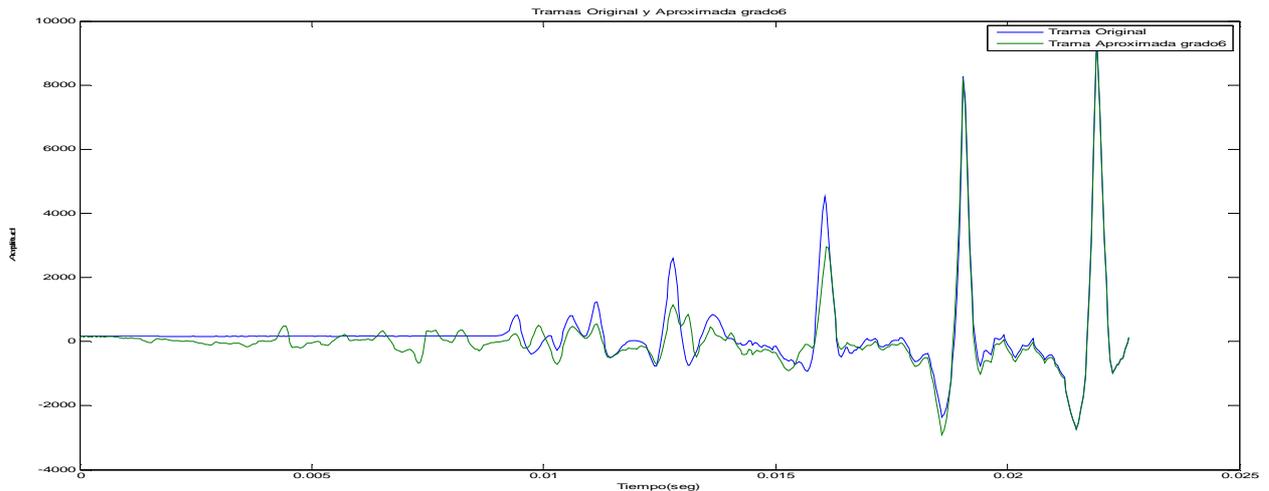
Para el residuo tenemos:



(Fig. 4.1.3.1.2 Residuo o ruido de la Trama 3, $P=6$, $N_{harm}=28$)

Se observa como la señal residuo obtenida tiene parte de los picos correspondientes a la parte determinista del transitorio, **no se aproxima bien.**

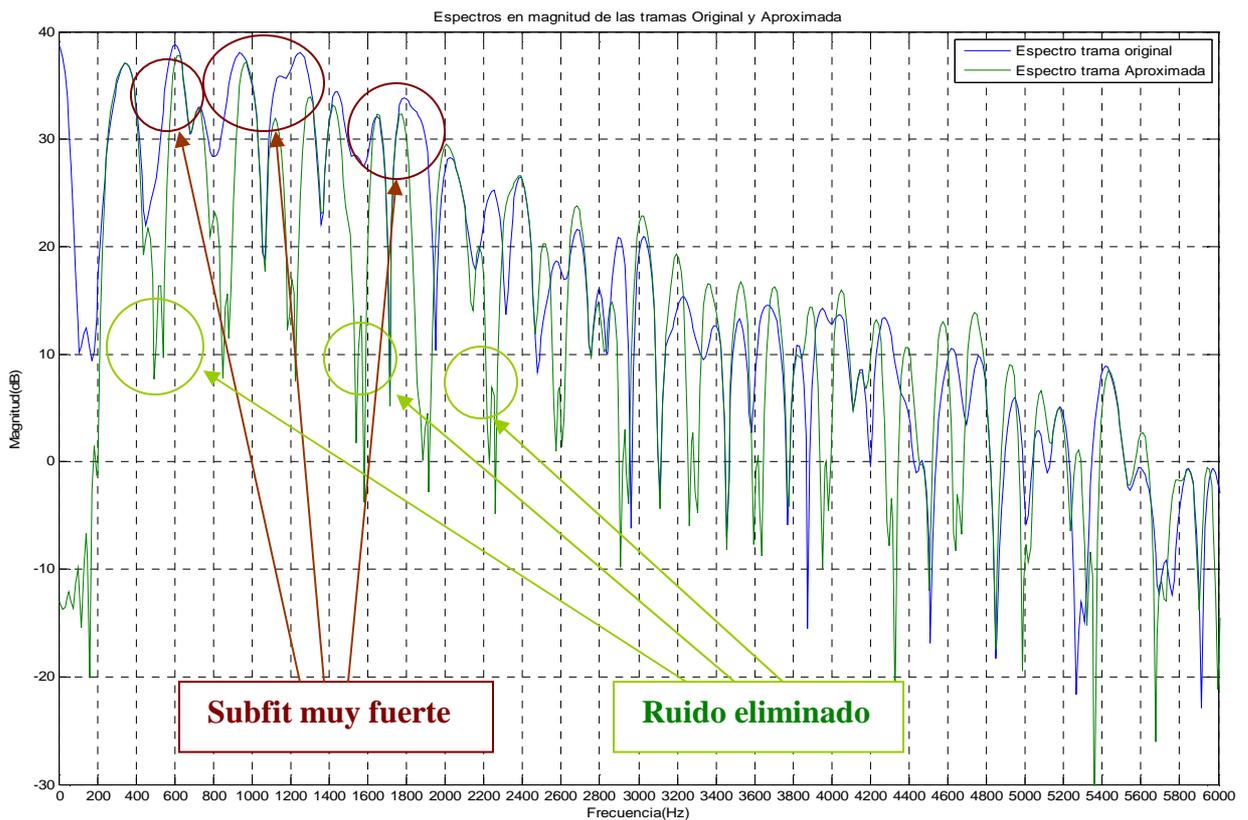
Representamos en el dominio temporal la trama original junto a la aproximada:



(Fig. 4.1.3.1.3 Trama 3 original y modelada polinomialmente, $P=6$, $N_{harm}=28$)

Se observa en el dominio temporal como la aproximación en verde **no es capaz de seguir correctamente el ataque inicial, no se modela del todo bien el transitorio.**

Representamos los espectros de las tramas original y aproximada:



(Fig. 4.1.3.1.4 Espectros de la Trama 3 original y aproximada, $P=6$, $N_{harm}=28$)

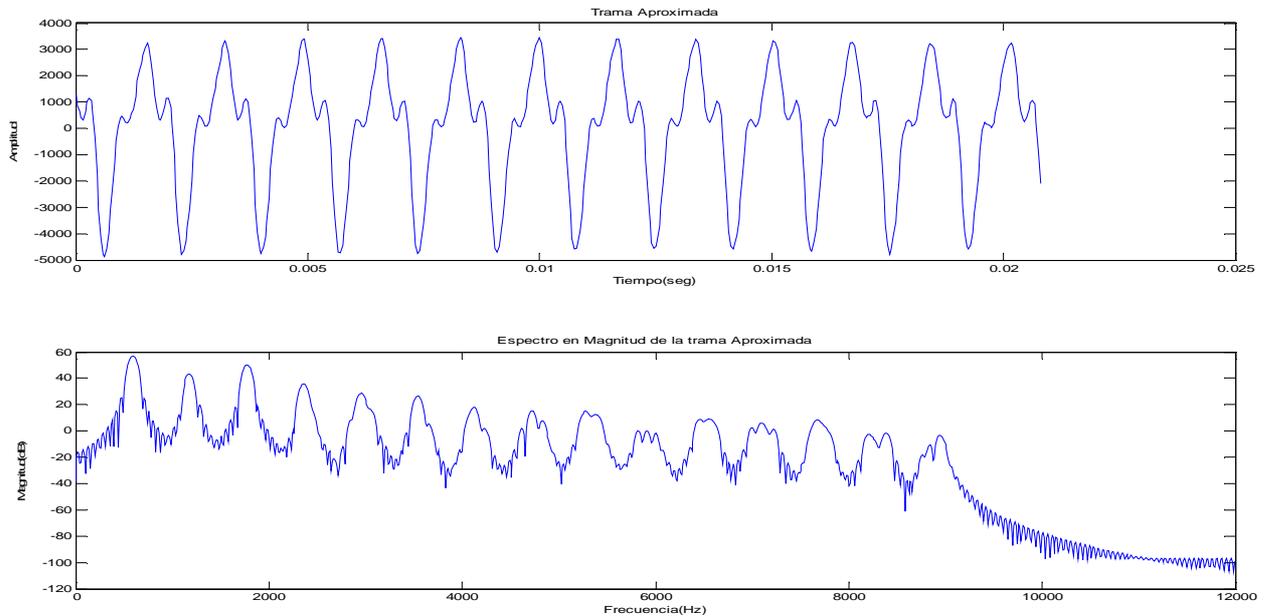
En el dominio frecuencial se observa como apenas se modela correctamente el primer parcial, para el resto se produce un **fuerte Subfit**. El polinomio de **grado 6 no consigue modelar tramas con tanta modulación**, en las que sus parciales se van ensanchando y deforman a medida que se sube en frecuencia hasta el punto de producirse dobles picos e inarmonicidad. **El modelado polinomial no es adecuado para modelar este tipo de tramas, consideradas transitorias.**

4.1.4 MODELADO POLINOMIAL TRAMA 4:

Trama cuasi estacionaria ruidosa y con poca modulación. Shakuhachi

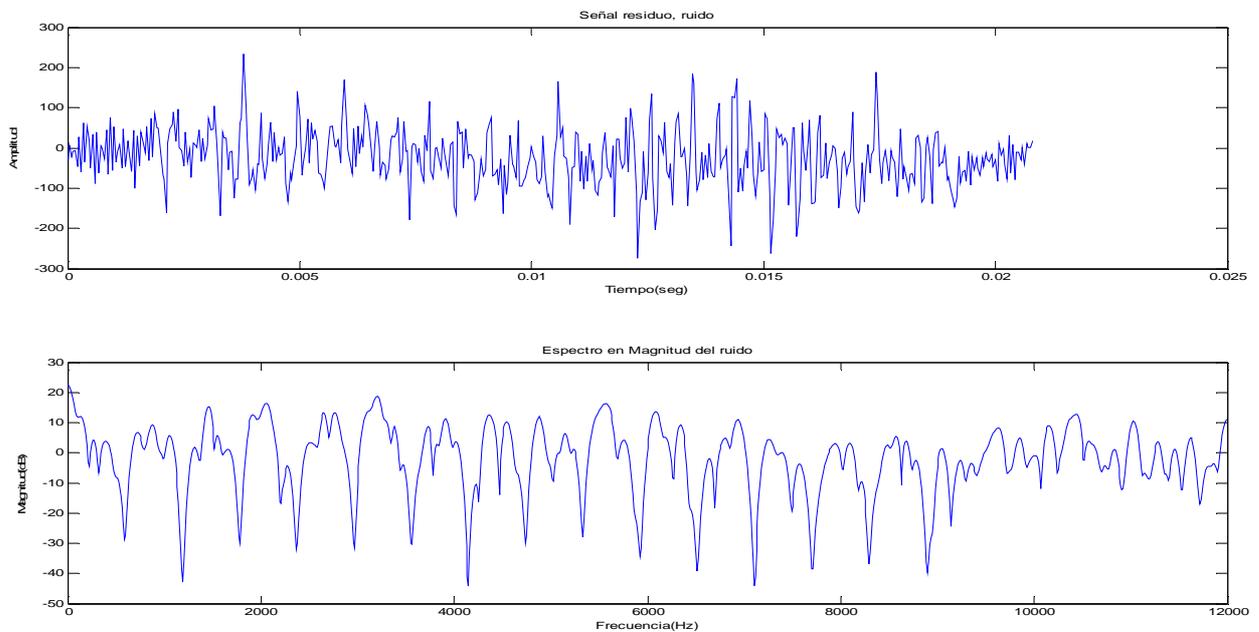
4.1.4.1 - Grado del polinomio= 6, número de armónicos=15.

Trama aproximada en el tiempo y su espectro:



(Fig. 4.1.4.1.1 Trama 4 Modelada polinomialmente, $P=6$, $N_{harm}=15$)

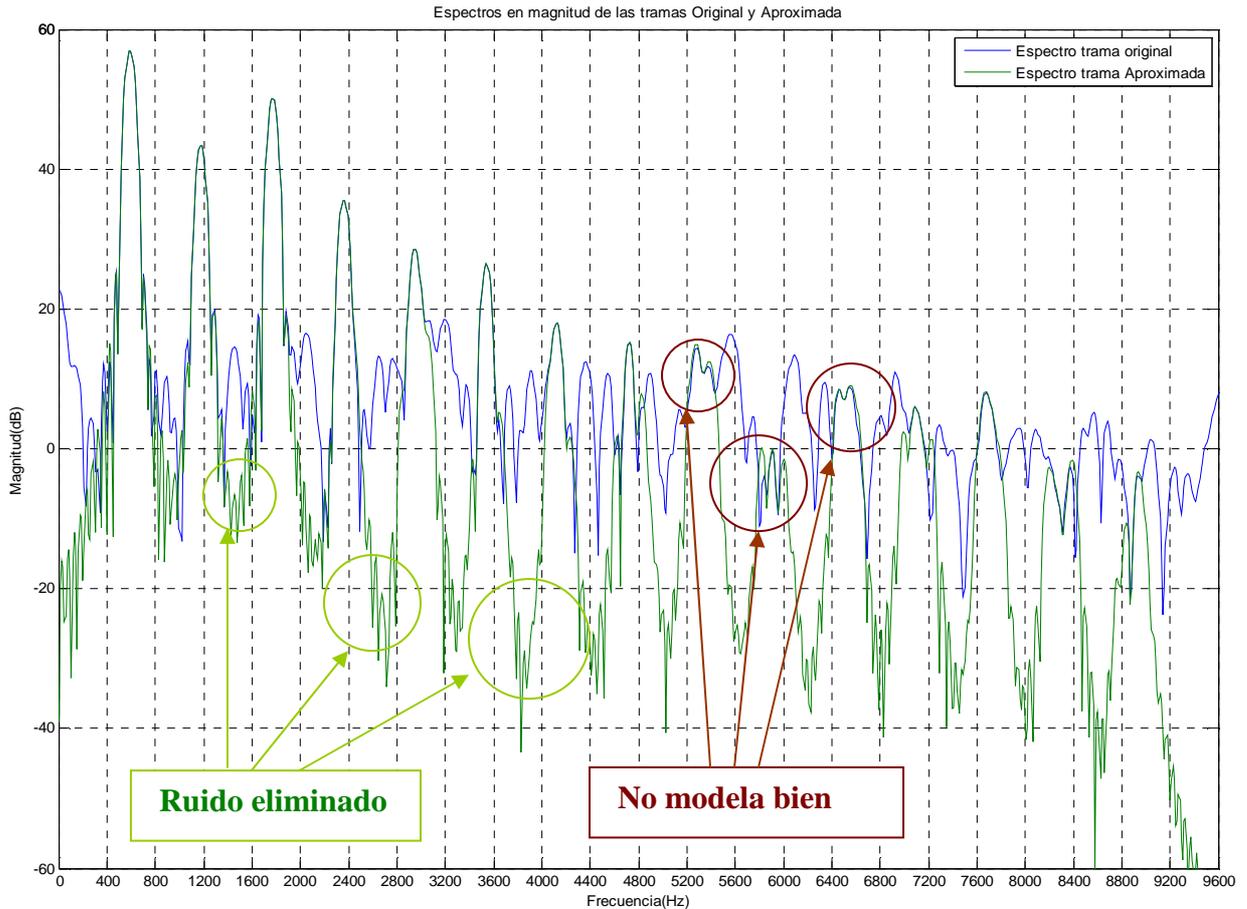
Residuo:



(Fig. 4.1.4.1.2 Residuo o ruido de la Trama 4, $P=6$, $N_{harm}=15$)

Modelamos con 15 armónicos porque la envolvente espectral decae rápidamente. La componente residual obtenida carece de periodicidad por tanto es **suficiente con 15 armónicos para modelar** la componente determinista.

Representamos juntos los espectros de la trama original y la aproximada:



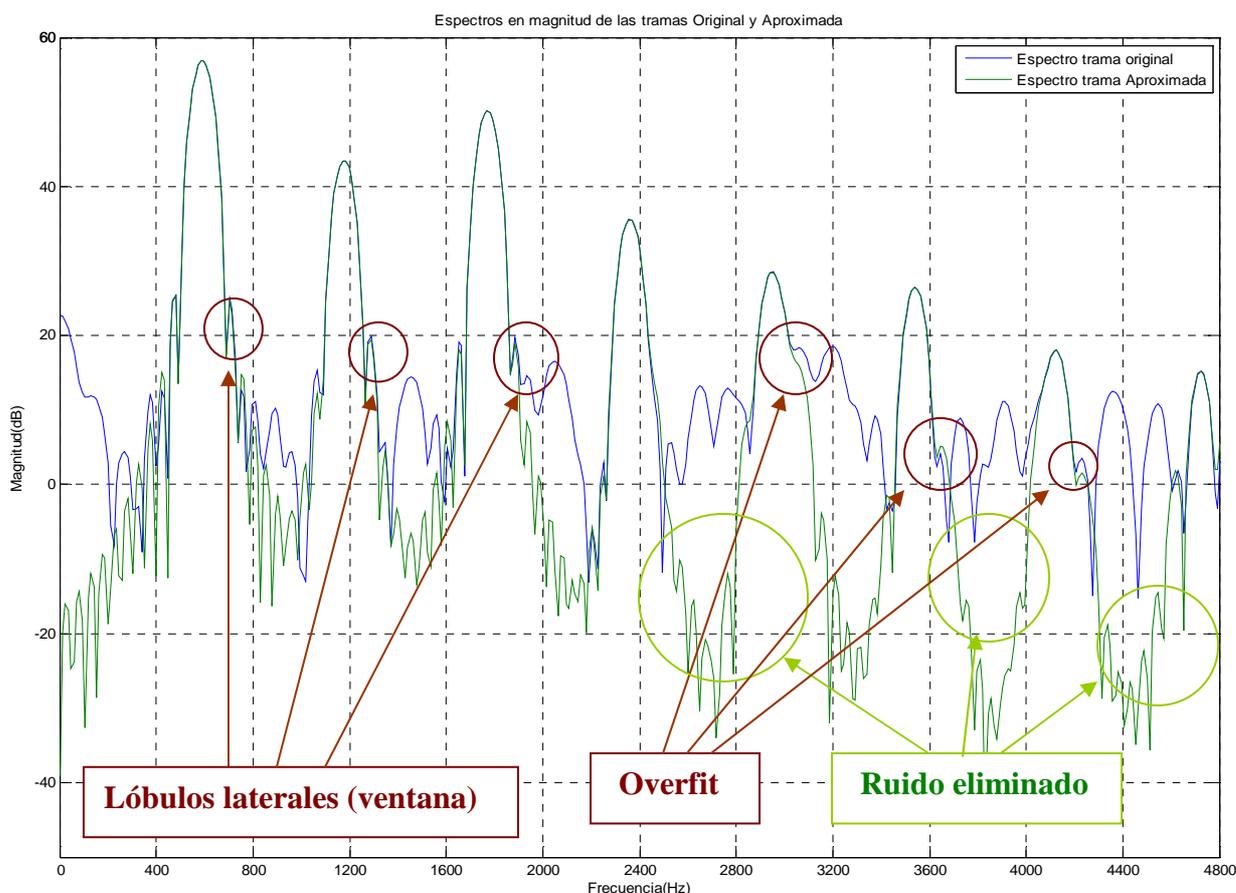
(Fig. 4.1.4.1.3 Espectros de la Trama 4 original y aproximada, $P=6$, $N_{harm}=15$)

Se observa como al modelar conseguimos eliminar gran cantidad de ruido en la zona de los 8 primeros parciales pero **a partir del octavo los parciales modelados tienen formas no muy fiables**. Estos parciales en la trama original se encuentran tan inmersos en el ruido que no es posible extraer la información determinista. Además algunos instrumentos en algunas tramas pueden producir cierta inarmonicidad en la posición de sus parciales que es más pronunciada en alta frecuencia como se intuye que pueda ocurrir en esta trama. Esto dificulta aún más su modelado.

El polinomio de **grado 6 tiene tanto grado de libertad que al modelar a partir del octavo parcial lo hace siguiendo la forma de varios picos simultáneamente** en un efecto similar al **Overfit**.

La mayor parte de la energía se encuentra concentrada en los primeros 8 parciales, por ello el mal modelado a partir del octavo parcial no influye demasiado sobre la energía determinista modelada total.

Ampliamos la zona de los primeros parciales que están bien modelados y es donde se concentra la mayor parte de la energía determinista para estudiar la existencia de Overfit:



(Fig. 4.1.4.1.4 Espectros ampliados de la Trama 4 original y aproximada, $P=6$, $N_{harm}=15$)

La mayor parte de la energía esta concentrada en los **3 primeros parciales**, se observa que el margen entre nivel de éstos y el del ruido es de **más de 60 decibelios**. Los pequeños picos adosados a estos parciales **no son efecto del Overfit**, sino que son **lóbulos laterales producido al enventanar**, visibles en éstos primeros parciales por lo bajo que se encuentra el nivel del ruido respecto al nivel de éstos (no estamos diciendo que el nivel de ruido sea bajo, sino que al estar casi toda la energía determinista concentrada en los 3 primeros parciales, la relación SNR es muy grande en esta zona).

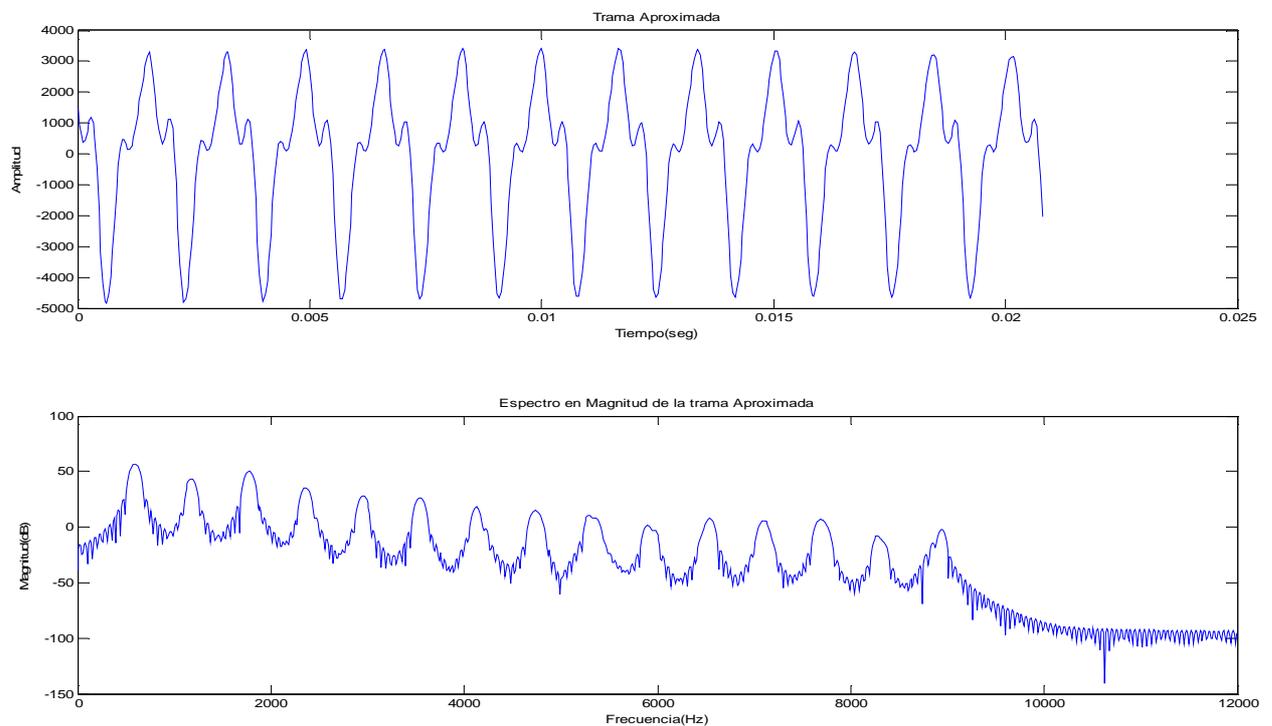
En anteriores tramas no se visualizaban los lóbulos laterales porque siempre quedaban por debajo o al mismo nivel que el ruido.

En el quinto, sexto y séptimo parcial se observa un **ligero Overfit despreciable**. Destaca la gran cantidad de pulsos ruidosos eliminados entre los parciales de esta zona. Podemos decir que **se modelan bien los parciales donde se concentra la mayor parte de la energía**, eliminando gran cantidad de ruido y **presentando en los 8 primeros parciales un margen promedio entre el nivel de éstos y el ruido de unos 60 decibelios**. En la trama original solo se encontraban así de limpios los 3 primeros parciales.

Veremos que sucede si sobre la misma trama bajamos el grado del polinomio de aproximación.

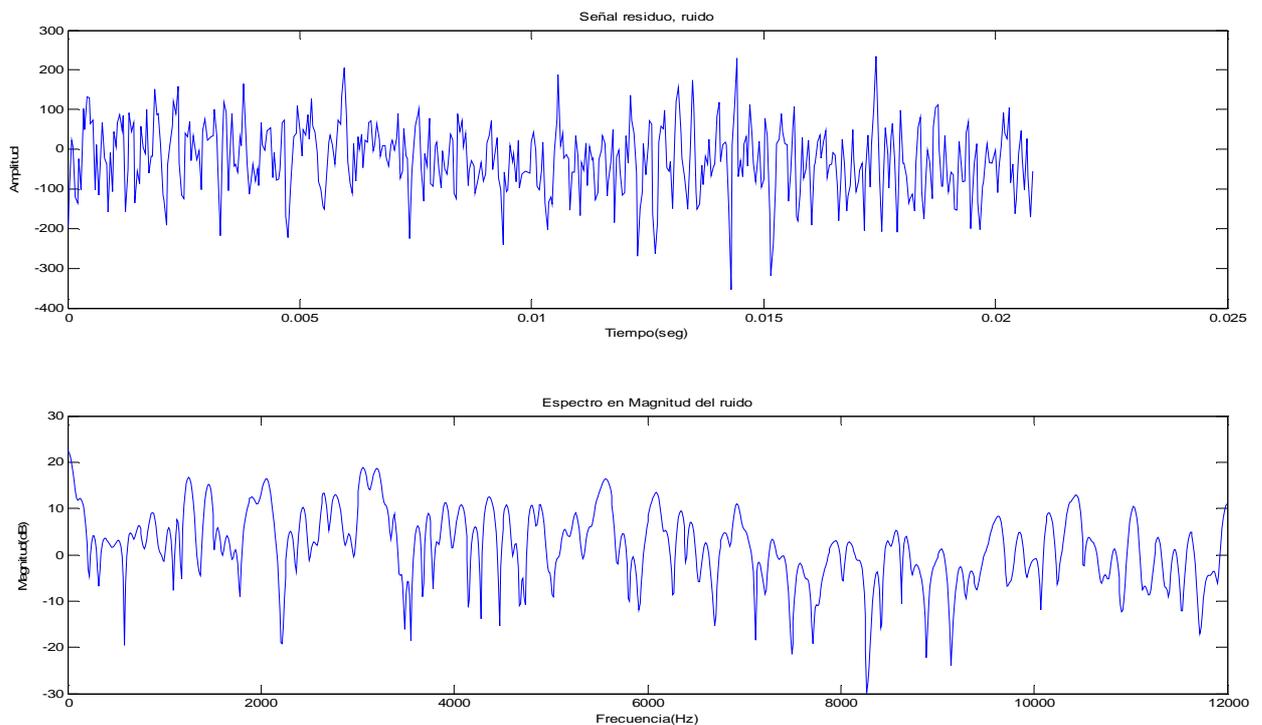
4.1.4.2 - Grado del polinomio= 3, número de armónicos=15.

Obtenemos la siguiente trama aproximada en el tiempo y su espectro:



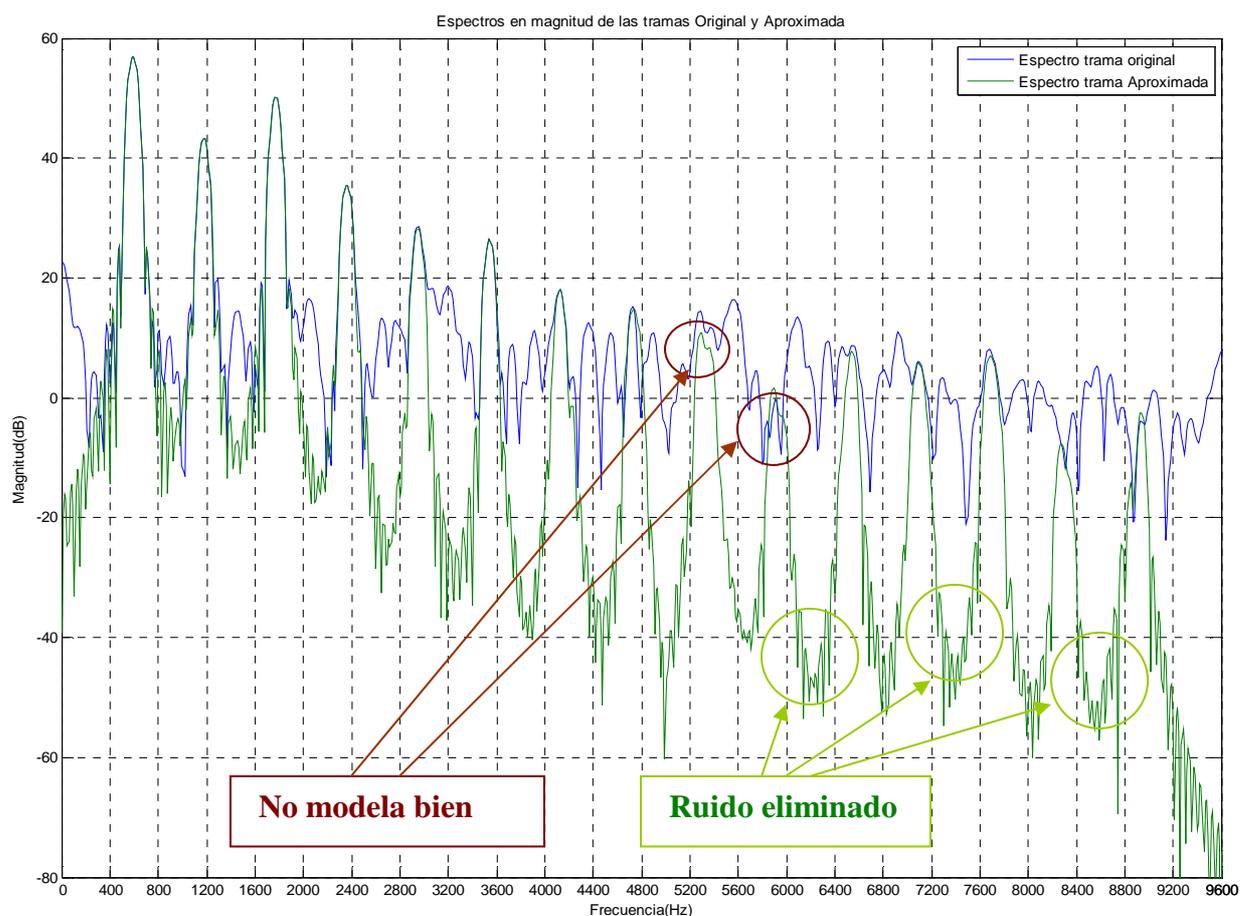
(Fig. 4.1.4.2.1 Trama 4 Modelada polinomialmente, $P=3$, $N_{harm}=15$)

Para el residuo tenemos:



(Fig. 4.1.4.2.2 Residuo o ruido de la Trama 4, $P=3$, $N_{harm}=15$)

Representamos los espectros de las tramas original y aproximada:

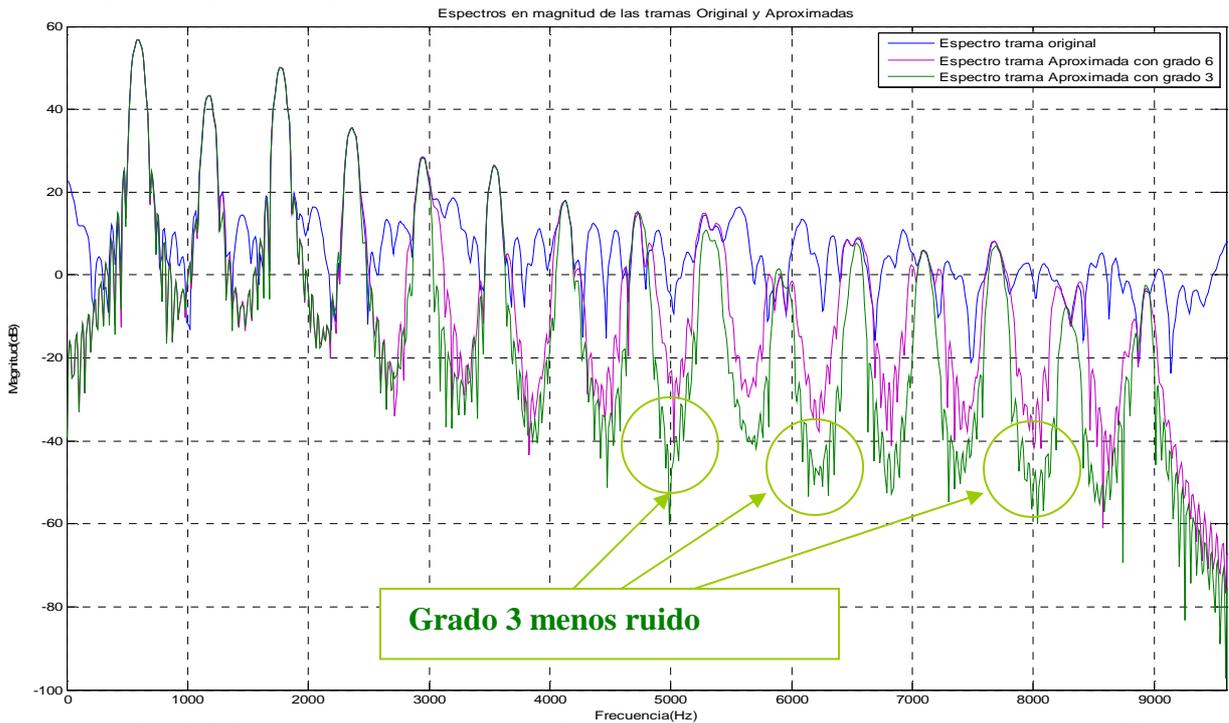


(Fig. 4.1.4.2.3 Espectros de la Trama 4 original y aproximada, $P=3$, $N_{harm}=15$)

Se observa en este caso como con **grado 3** en el polinomio de aproximación se consigue modelar con **menos ruido** en la zona de los parciales altos. Respecto a la forma con la que se modelan los parciales, hasta el octavo parece que es fiable, el noveno tiene doble pico y no es correcto, **a partir del noveno se modelan los parciales con formas coherentes pero sus posiciones no son fiables**, ya que se intuye que en la trama original los parciales no estén armónicamente distribuidos y los parciales modelados puede que no caigan en las posiciones correctas, no se puede saber con certeza porque estos parciales en la trama original están muy inmersos en el ruido. Como ya se comentó **la zona importante es la de los 8 primeros parciales que es donde se concentra la mayor parte de la energía determinista**.

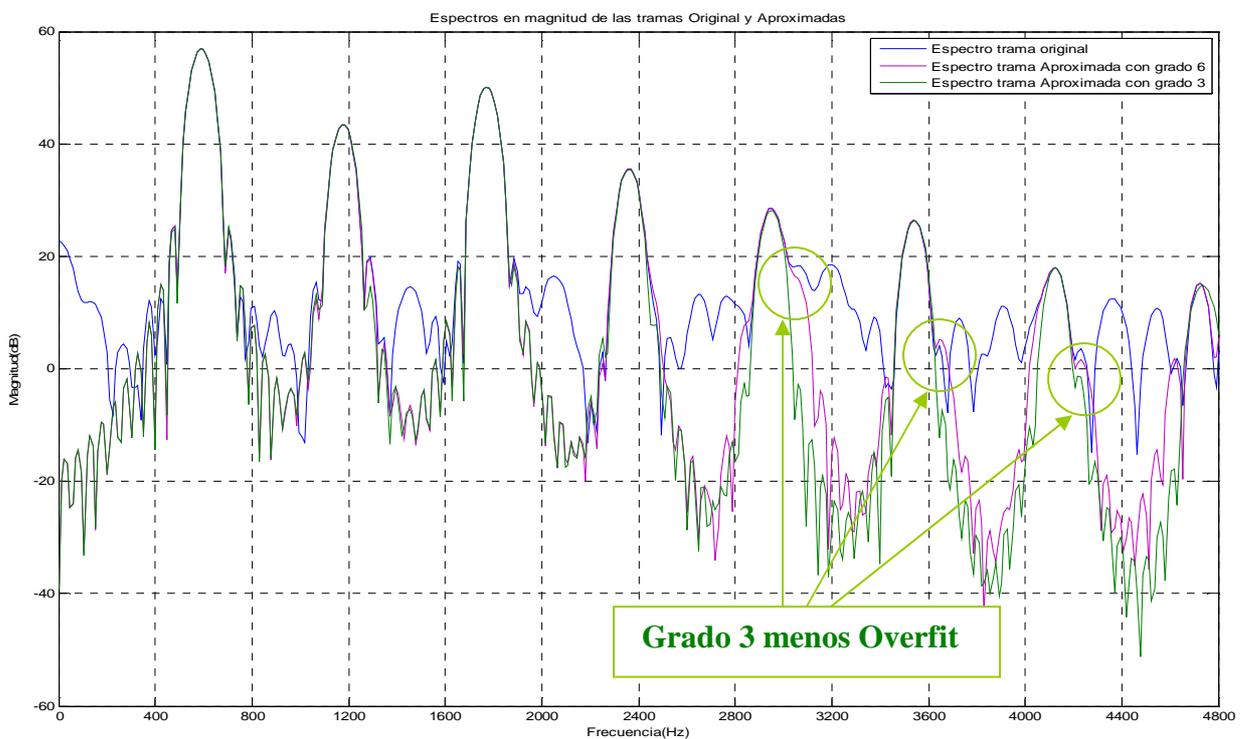
Si la envolvente espectral de este instrumento fuese más plana como en el caso de la trompeta, el ruido afectaría a todos los parciales de forma más equitativa. En este caso al decaer rápidamente la envolvente, a partir del noveno parcial se tiene muy poca energía determinista y el efecto del ruido es más devastador sobre los parciales altos que en el caso de la trompeta.

Representamos el espectro de la trama original junto a los de las aproximaciones de grado 6 y grado 3 para comparar:



(Fig. 4.1.4.2.4 Espectros de la Trama 4 original y aproximada, $P=6$ y 3 , $N_{harm}=15$)

Ampliando los primeros parciales:



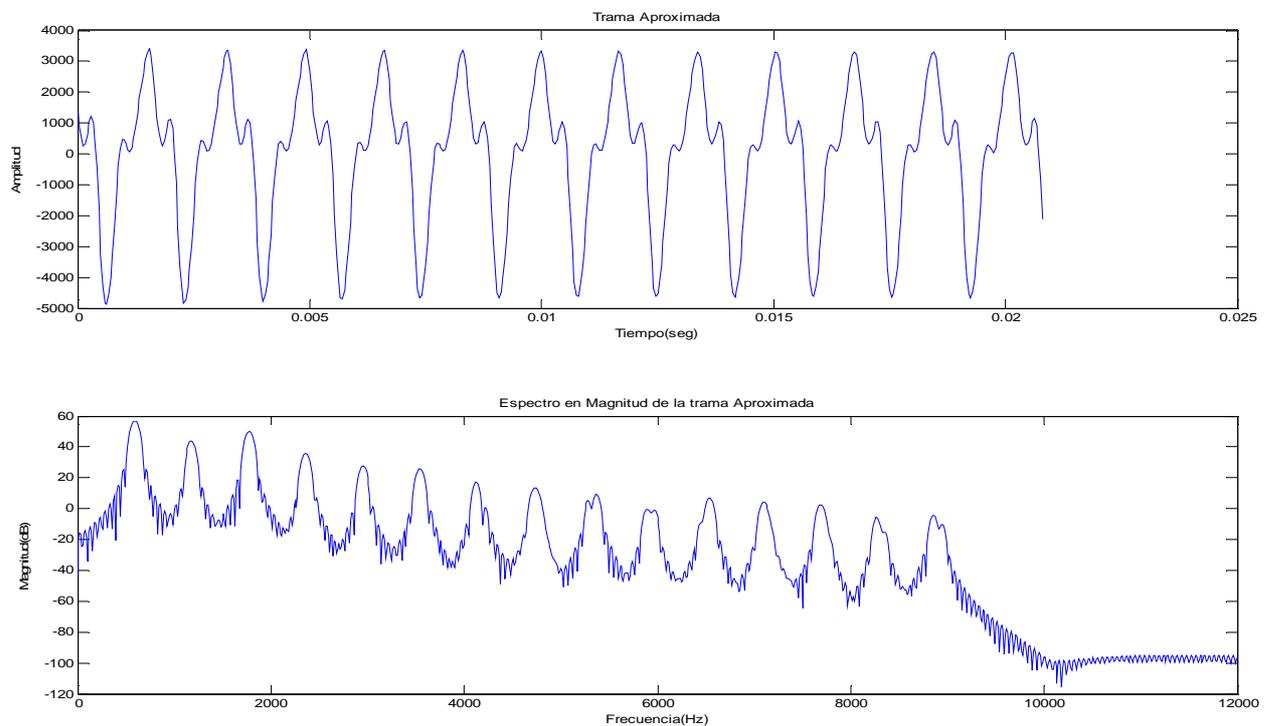
(Fig. 4.1.4.2.5 Espectros ampliados de la Trama 4 original y aproximada, $P=6$ y 3 , $N_{harm}=15$)

Observando ambas aproximaciones se aprecia como la trama modelada **con grado 3 elimina más ruido y produce menos Overfit** que la modelada con grado 6.

Veremos que sucede si bajamos a 2 el grado del polinomio de aproximación.

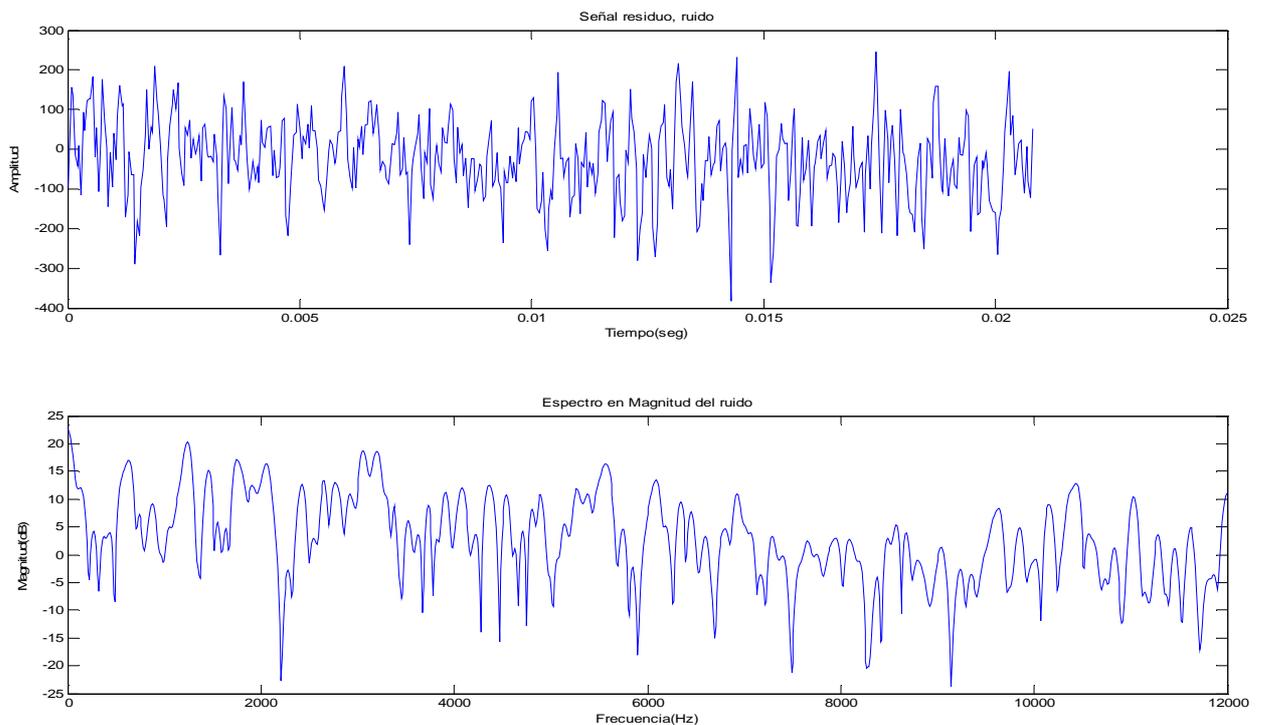
4.1.4.3 - Grado del polinomio= 2, número de armónicos=15.

Obtenemos la siguiente trama aproximada y su espectro:



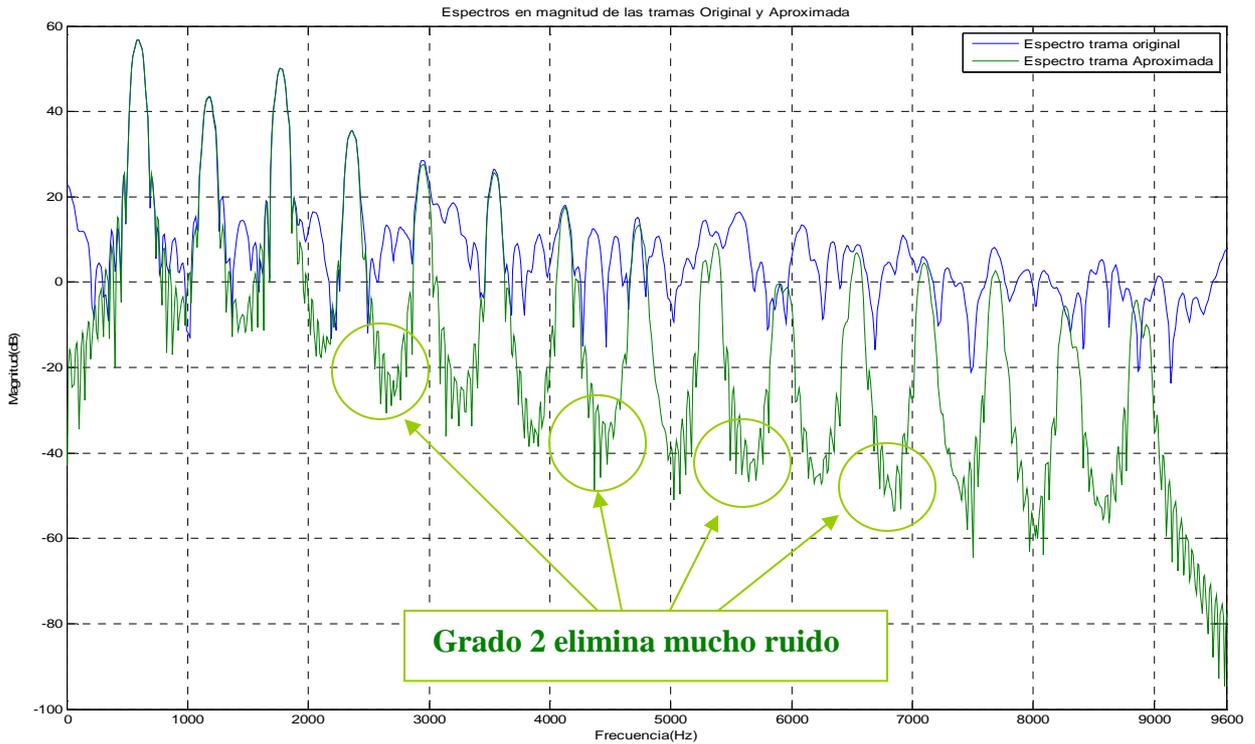
(Fig. 4.1.4.3.1 Trama 4 Modelada polinomialmente, $P=2$, $N_{harm}=15$)

Para el residuo tenemos:



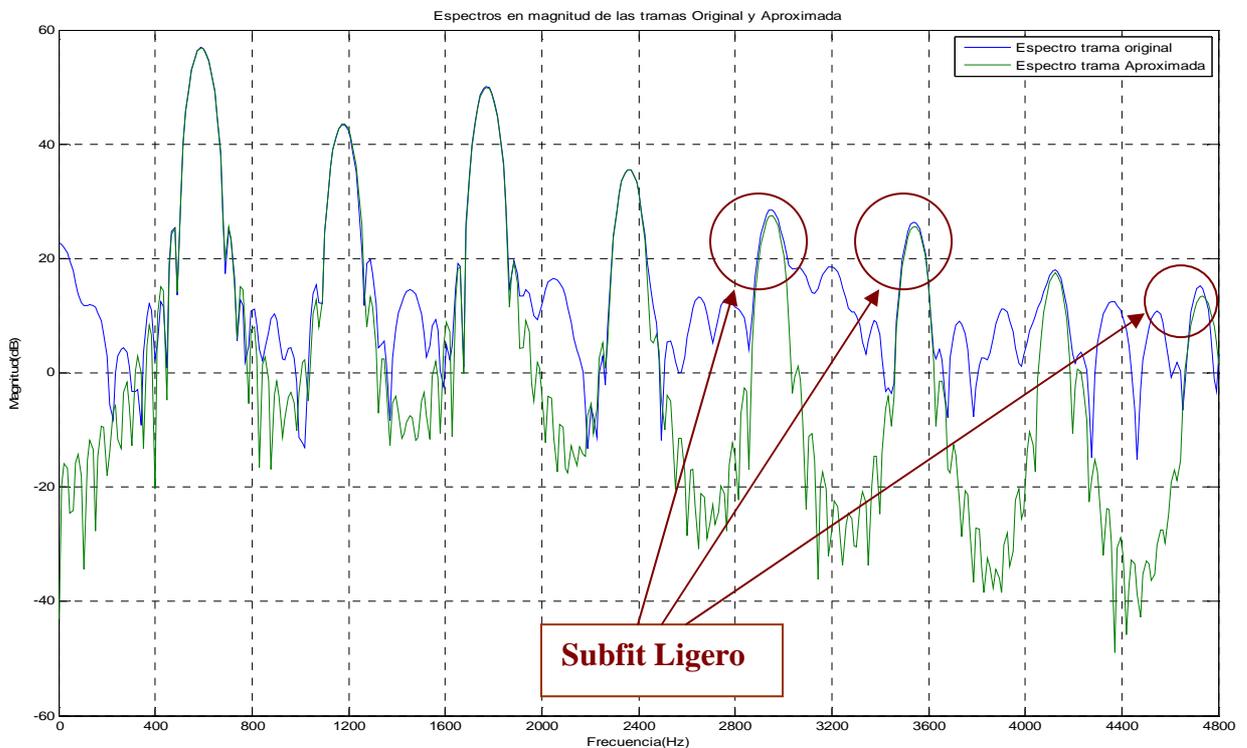
(Fig. 4.1.4.3.2 Residuo o ruido de la Trama 4, $P=2$, $N_{harm}=15$)

Representamos el espectro de la trama aproximada con grado 2 y el de la original:



(Fig. 4.1.4.3.3 Espectros de la Trama 4 original y aproximada, $P=3$, $N_{harm}=15$)

Ampliando los primeros parciales:

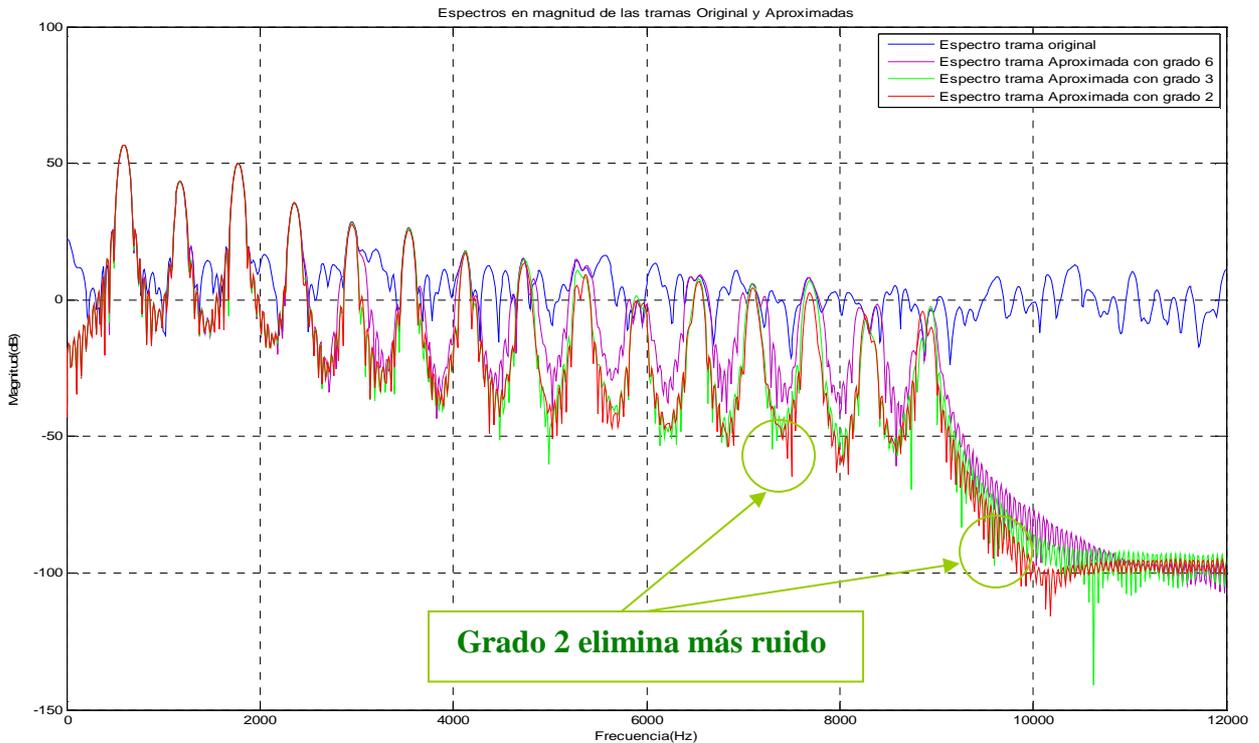


(Fig. 4.1.4.3.4 Espectros ampliados de la Trama 4 original y aproximada, $P=3$, $N_{harm}=15$)

Se observa como con **grado 2** en el polinomio de aproximación **se elimina mucho ruido** y comienza a producirse un **ligero Subfit que puede considerarse despreciable**.

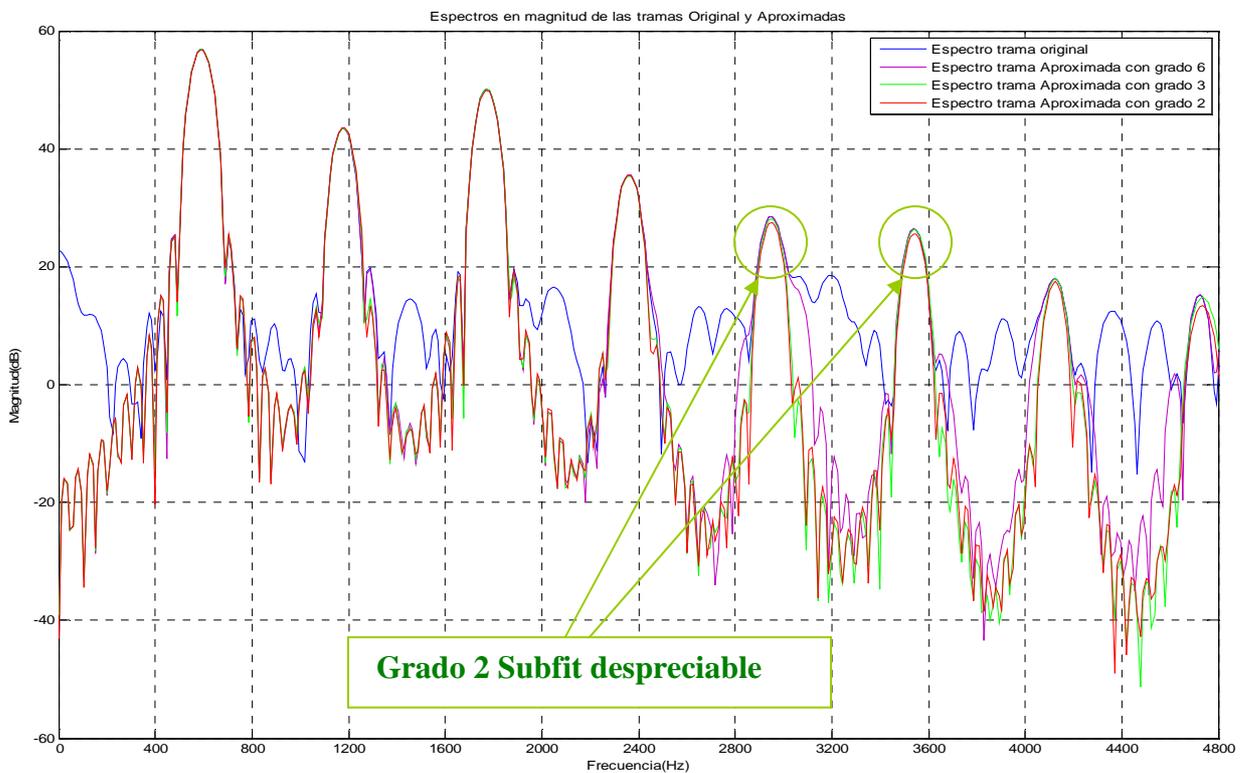
4.1.4.4 - Conclusiones al modelado polinomial de la Trama 4:

Representamos el espectro de la trama original junto a los de las aproximaciones estudiadas para comparar:



(Fig. 4.1.4.4.1 Espectros de la Trama 4 original y aproximadas, $P=6, 3$ y 2 $N_{harm}=15$)

Ampliando:



(Fig. 4.1.4.4.1 Espectros ampliados de la Trama 4 original y aproximadas, $P=6, 3$ y 2 $N_{harm}=15$)

Se observa como con **grado 2** en el polinomio de aproximación **se elimina mayor cantidad de ruido** aunque no existe mucha diferencia con el polinomio de grado 3. El polinomio de **grado 2 produce un ligero Subfit que puede considerarse despreciable**, por ello, ante más o menos iguales resultados, nos quedamos con el polinomio de grado 2 que siempre será computacionalmente más eficiente que el de grado 3.

Se concluye que **para este tipo de tramas, muy ruidosas y con poca modulación AM-FM, lo adecuado será modelar con modelado polinomial de grado 2**, ya que al tener poca modulación **apenas produce Subfit, no produce Overfit y es el que más ruido elimina**, además de ser **computacionalmente más eficiente**.

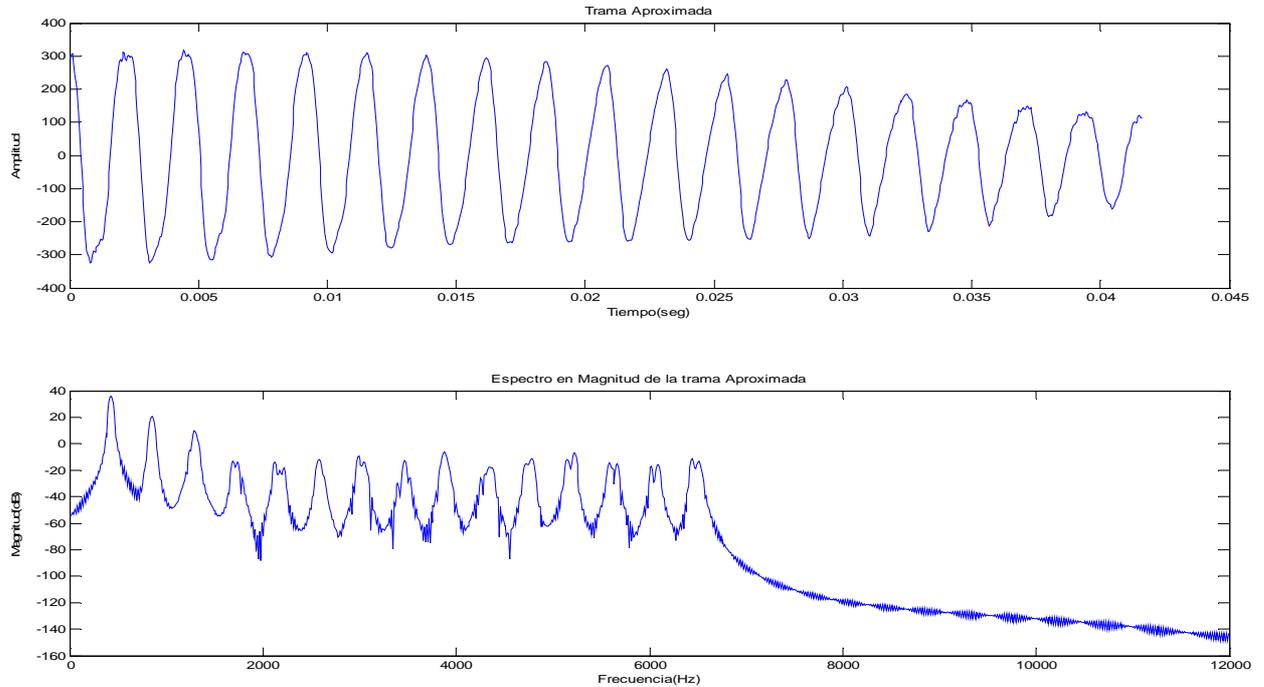
TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3
Mucha modulación y poco ruido	Modelado no estacionario polinomial grado 6
Poca modulación y mucho ruido	Modelado no estacionario polinomial grado 2

4.1.5 MODELADO POLINOMIAL TRAMA 5:

Trama cuasi estacionaria ruidosa y con mucha modulación. Shakuhachi

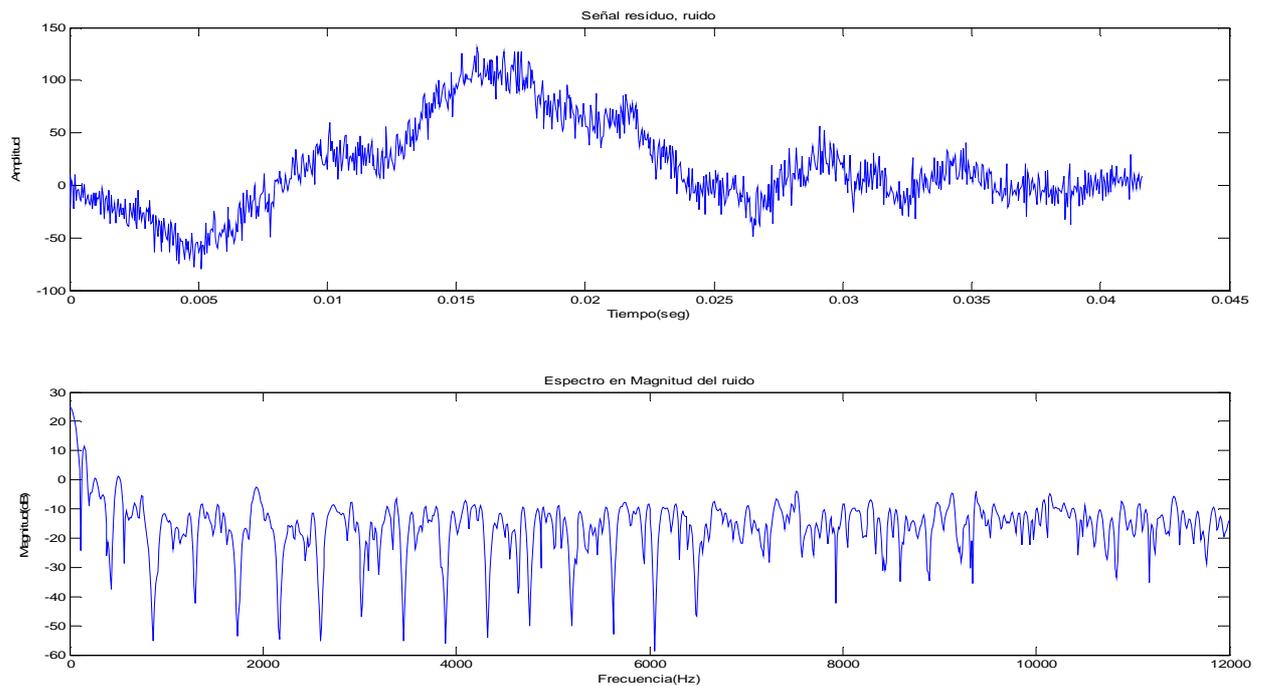
4.1.5.1 - Grado del polinomio= 6, número de armónicos=15.

Representamos la trama aproximada polinomialmente y su espectro:



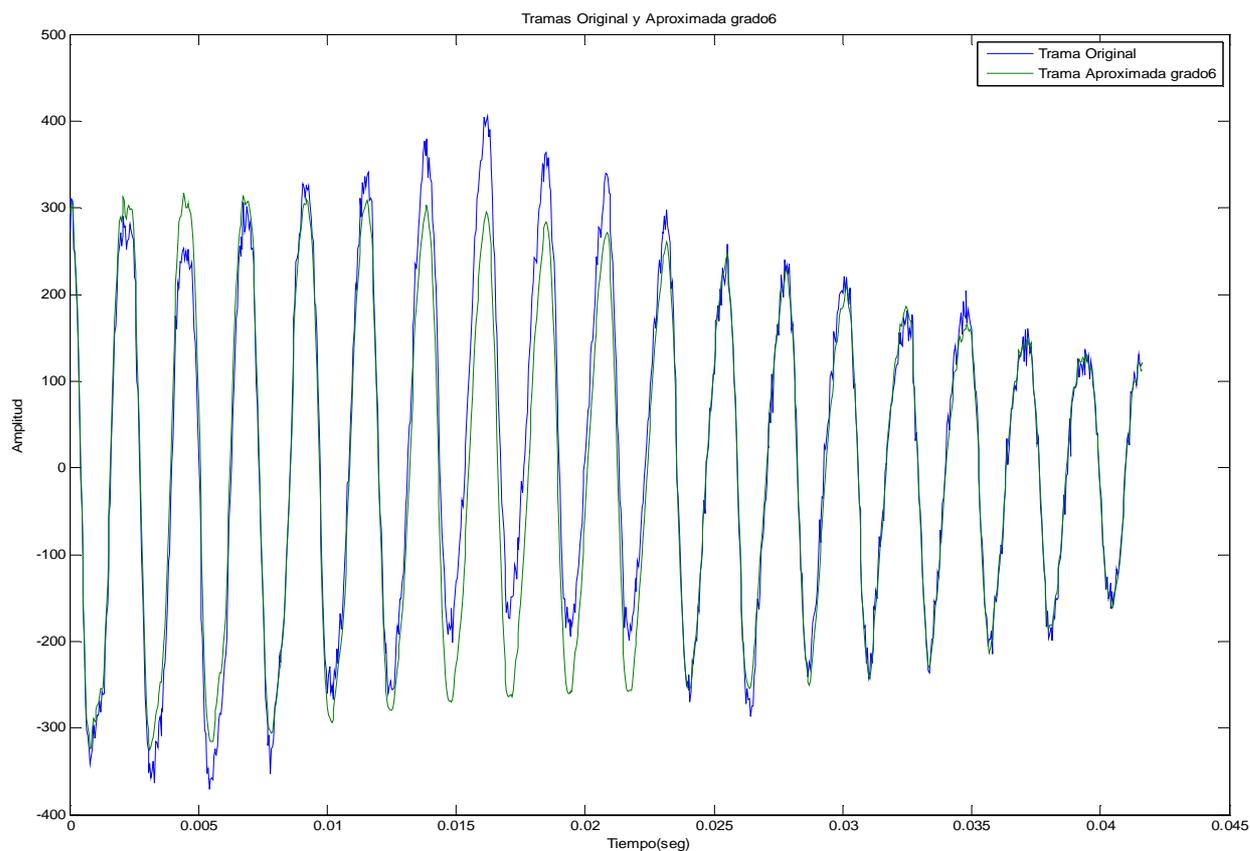
(Fig. 4.1.5.1.1 Trama 5 Modelada polinomialmente, $P=6$, $N_{harm}=15$)

Para el residuo obtenemos lo siguiente:



(Fig. 4.1.5.1.2 Residuo o ruido de la Trama 5, $P=6$, $N_{harm}=15$)

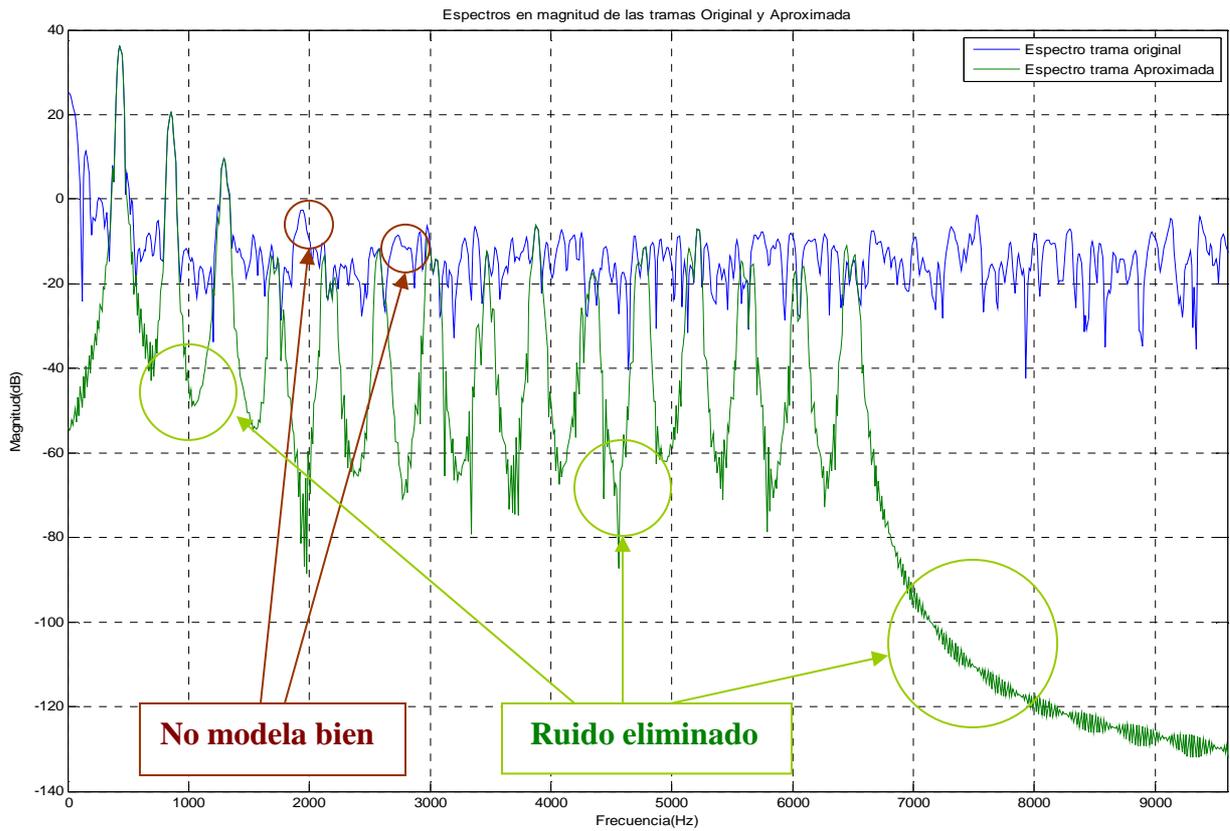
Además de estudiar comparativamente los espectros de las tramas original y aproximada, es interesante estudiar también las tramas en el dominio temporal:



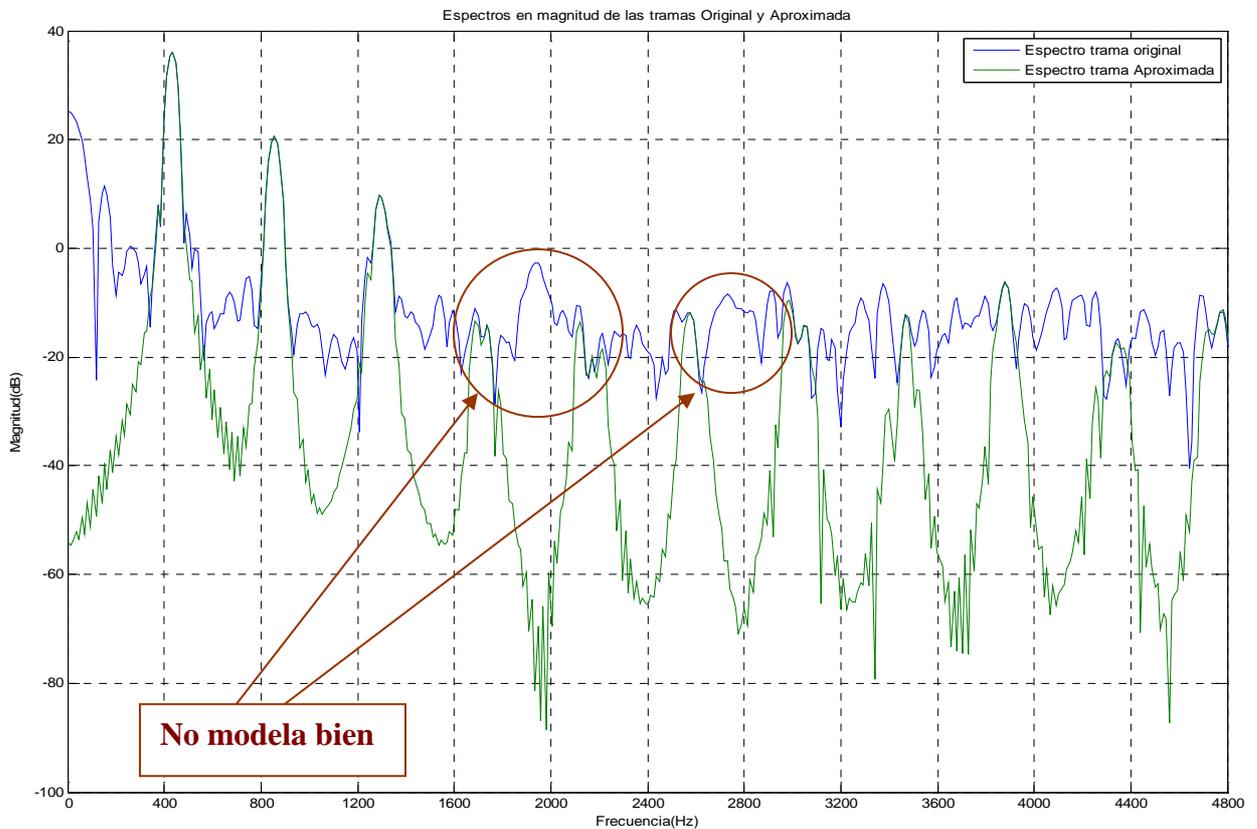
(Fig. 4.1.5.1.3 Trama 5 original y modelada polinomialmente, $P=6$, $N_{harm}=15$)

Se observa claramente en el dominio temporal como la trama modelada con polinomio de aproximación de **grado 6 no es capaz de seguir del todo la fuerte modulación AM-FM** de la trama original, se produce **Subfit**, pese a ser el grado más alto que podemos aplicar sin que se produzca saturación. Este tipo de trama se encontraría en el **límite de lo tonal**, casi es transitoria, además el alto nivel de ruido hace de esta trama el **peor caso que nos podamos encontrar dentro de las tramas consideradas armónicas**.

Veremos que sucede en el dominio frecuencial, representamos juntos los espectros de la trama original y la aproximada:



(Fig. 4.1.5.1.4 Espectros de la Trama 5 original y aproximada, $P=6$, $N_{harm}=15$)



(Fig. 4.1.5.1.5 Espectros ampliados de la Trama 5 original y aproximada, $P=6$, $N_{harm}=15$)

Se observa en el dominio frecuencial como los **3 primeros parciales son modelados correctamente** eliminando ruido en sus valles, **a partir del cuarto parcial no se modela correctamente**, debido a que **tenemos una fuerte inarmonicidad** en dichos parciales y no se encuentran en sus posiciones armónicas.

La inarmonicidad se puede producir en algunos instrumentos y para algunas tramas, pero es un fenómeno poco probable. El modelado polinomial es capaz de aproximar tramas en las que exista cierta inarmonicidad pero falla si la inarmonicidad es fuerte.

Realmente estamos ante el **peor caso de trama tonal** que nos podamos encontrar, se intuye que el cuarto y quinto parcial comienzan a tener una **fuerte inarmonicidad que el modelado no es capaz de seguir**. A este efecto se suma el ensanchamiento de los parciales a medida que se sube en frecuencia debido a la FM. **El cuarto parcial es interpretado como un gran pico ruidoso**, lo mismo ocurre con el quinto parcial que se intuye en una posición aún más alejada de la armónica, aunque no se puede saber con certeza por la gran cantidad de ruido.

Por tanto el **cuarto parcial modelado y los siguientes no pueden considerarse correctos**, ya no solo por no estar en las posiciones en las que se intuyen están los parciales de la trama original, sino también por su **forma con dobles picos producidos por el alto grado del polinomio**.

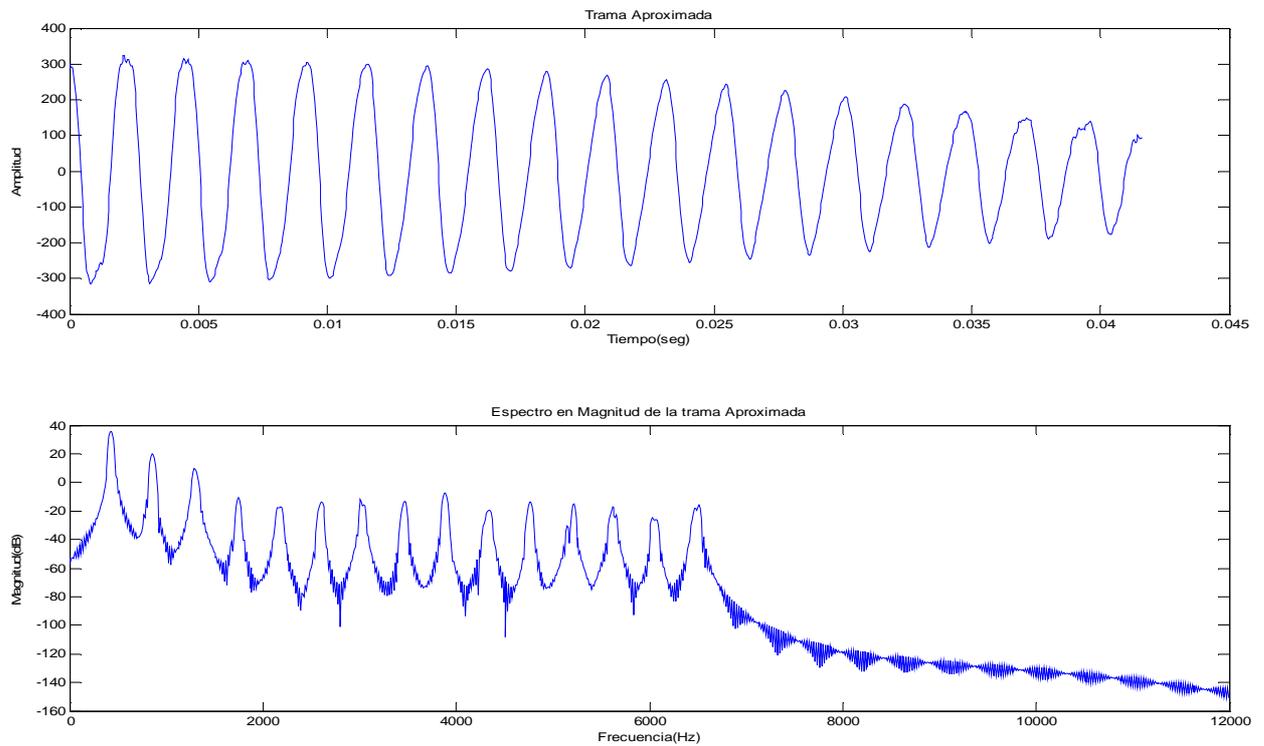
En este caso aproximábamos con 15 armónicos, vistos los resultados **se podría bajar el número de armónicos con los que se realiza el modelado, ya que apenas los 5 primeros son discernibles, pero esto produciría mucha distorsión** en la trama modelada respecto a la original.

Para los **3 primeros parciales** se aprecia como en la trama aproximada se consigue un margen entre el nivel del ruido y éstos de **más de 60 decibelios**, como en la trama anterior, en ésta **la mayor parte de la energía determinista está concentrada en estos parciales**, pero como se ha visto en el dominio temporal, **la mala aproximación del resto de los parciales si es determinante a la hora de modelar correctamente la trama, produciéndose Subfit**.

Veremos que sucede si bajamos el grado del polinomio de aproximación.

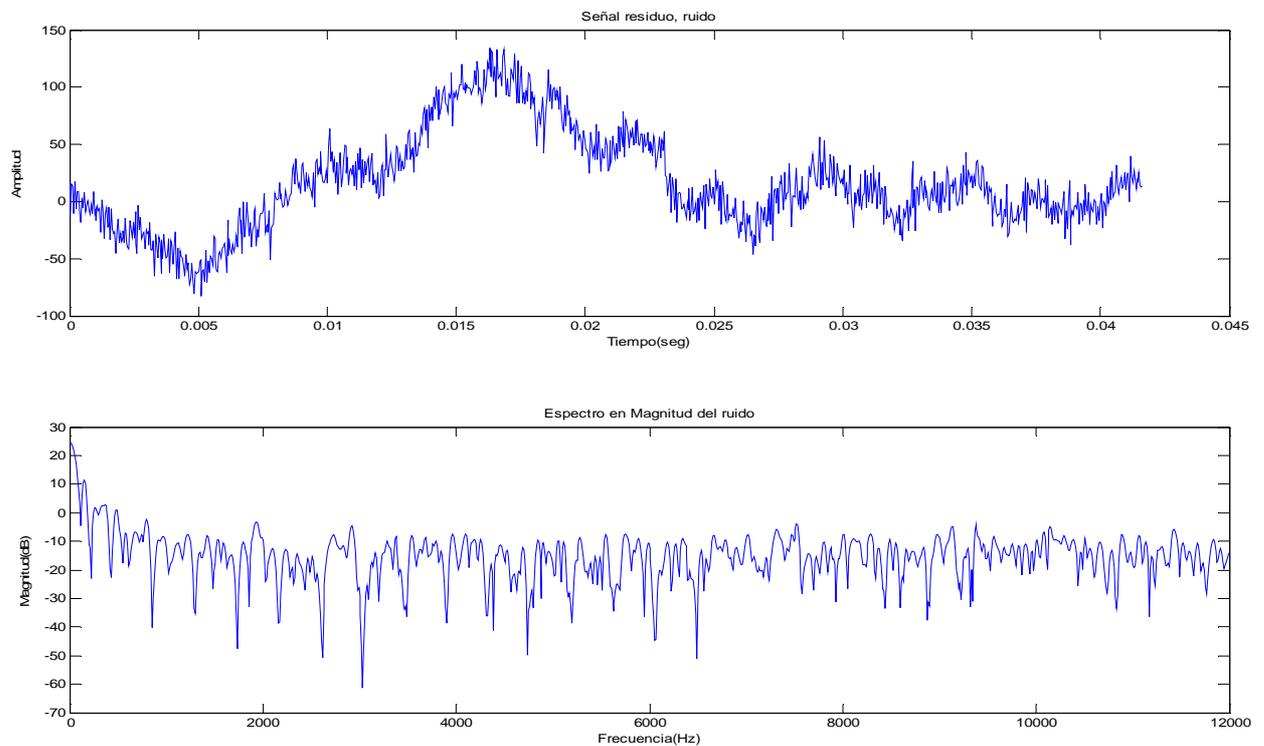
4.1.5.2 - Grado del polinomio= 3, número de armónicos=15.

Trama aproximada en el tiempo y su espectro:



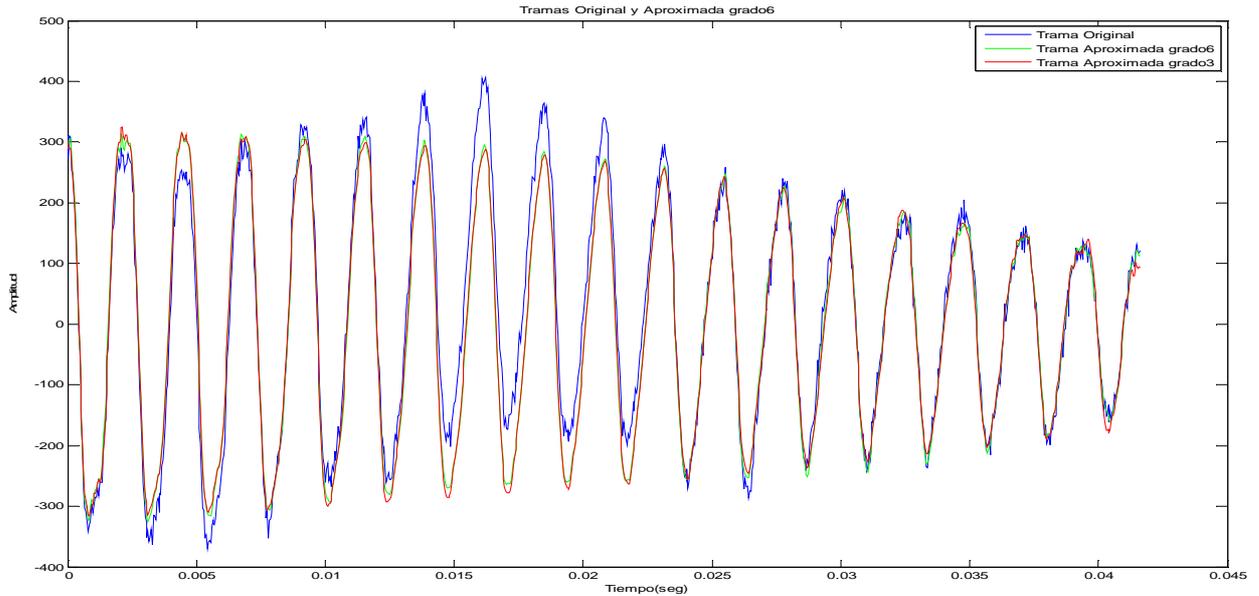
(Fig. 4.1.5.2.1 Trama 5 Modelada polinomialmente, $P=3$, $N_{harm}=15$)

Residuo:



(Fig. 4.1.5.2.2 Residuo o ruido de la Trama 5, $P=3$, $N_{harm}=15$)

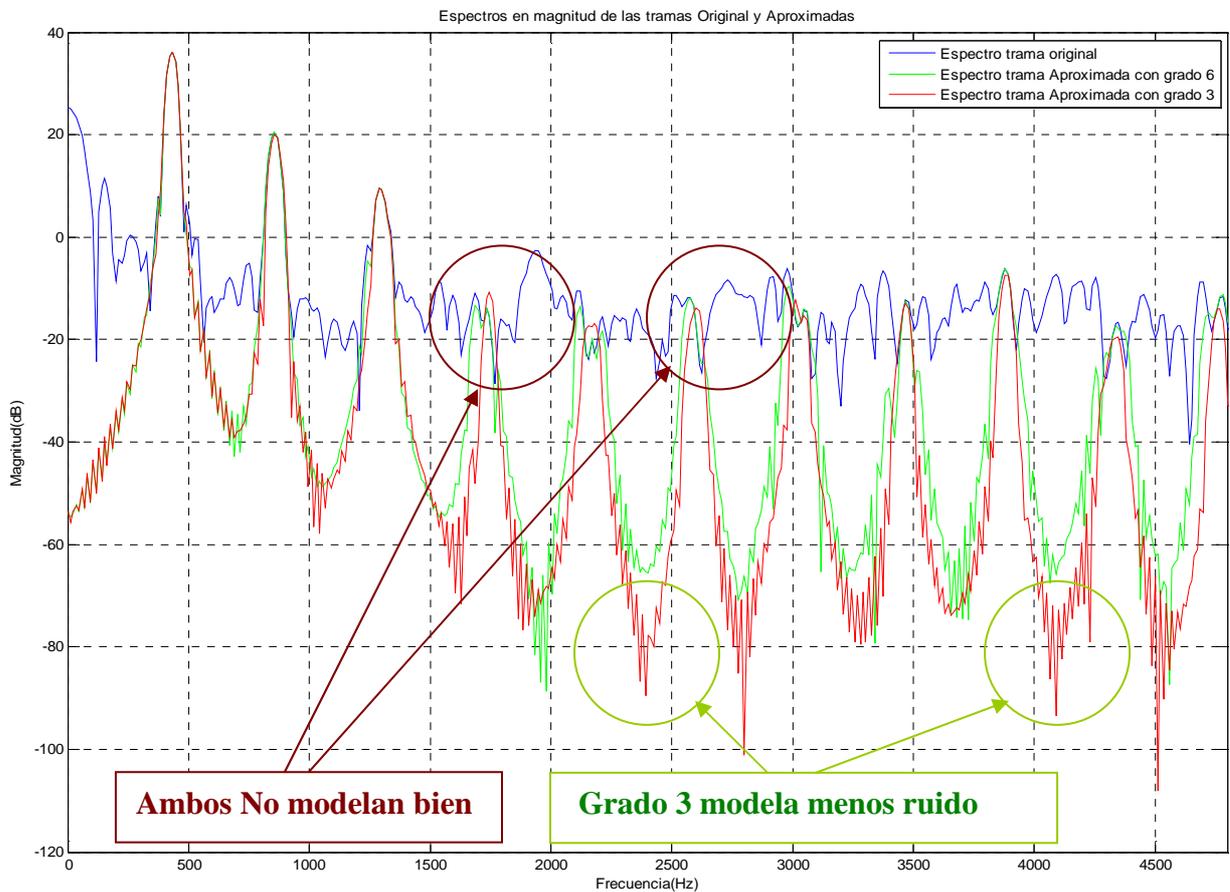
Graficamos juntas la trama original y las aproximadas con grado 3 y grado 6:



(Fig. 4.1.5.2.3 Trama 5 original y modeladas polinomialmente, $P=6$ y 3 , $N_{harm}=15$)

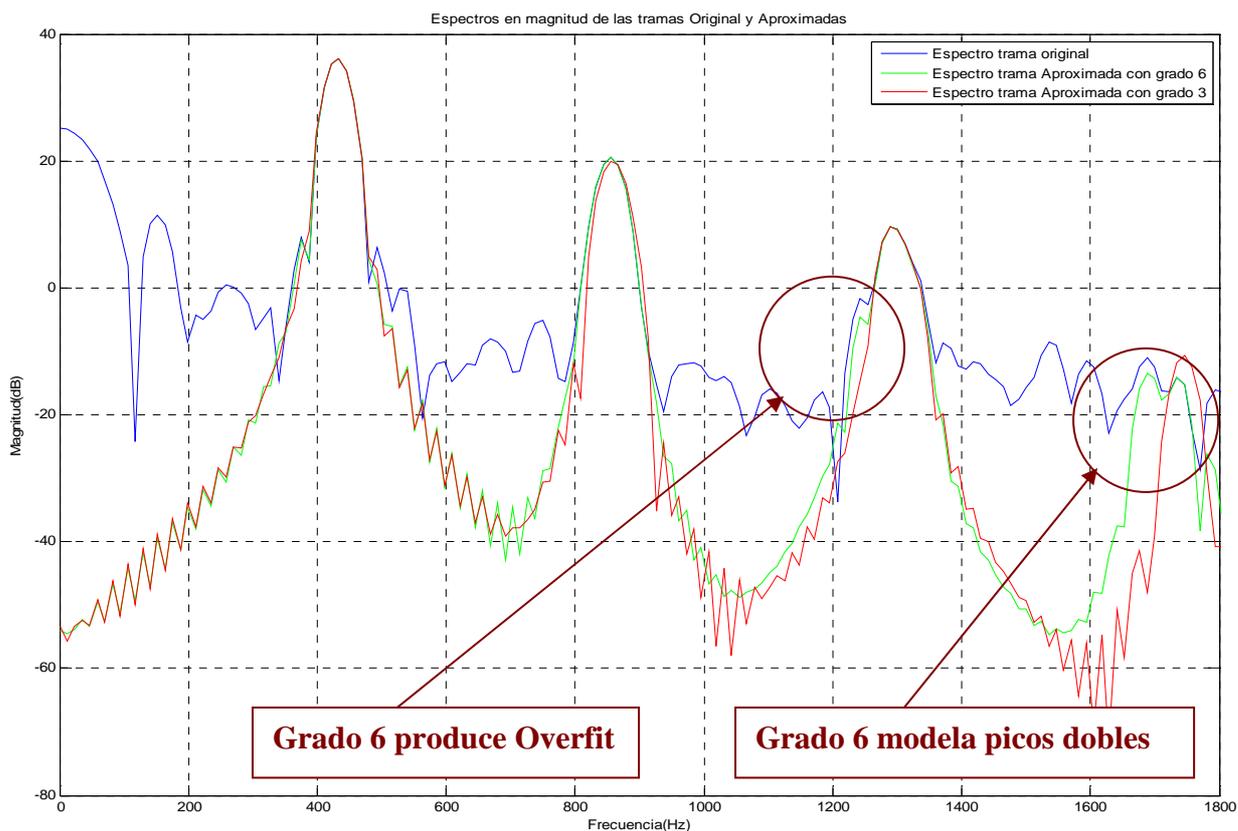
Se observa como a la aproximación de grado 3 en rojo le cuesta más seguir las variaciones de la trama original que a la aproximación de grado 6, aunque **la diferencia es ligera**.

Representamos juntos los primeros parciales de las aproximaciones de grado 3 y grado 6 para comparar:



(Fig. 4.1.5.2.4 Espectros de la Trama 5 original y aproximadas, $P=6$ y 3 , $N_{harm}=15$)

Ampliando los tres primeros parciales bien modelados:



(Fig. 4.1.5.2.5 Espectros ampliados de la Trama 5 original y aproximadas, $P=6$ y 3 , $N_{harm}=15$)

4.1.5.3 - Conclusiones al modelado polinomial de la Trama 5:

Ninguno de los polinomios consigue modelar bien el cuarto parcial y los siguientes. En **condiciones similares en cuanto al modelado obtenido en estos parciales y al Subfit producido**, el polinomio de **grado 3** es el que **más ruido elimina, sin producir Overfit, modelando con menos picos dobles y el que menor carga computacional generaría.**

Grado 2 en el polinomio de aproximación obtendría para esta trama resultados pésimos en cuanto a Subfit por la fuerte modulación que presenta.

Aproximando correctamente solo los 3 primeros parciales como es el caso **no se obtiene un buen modelado de la parte determinista de la trama original**, por esto llegados a este punto nos planteamos que **para este tipo de tramas, con mucha modulación AM-FM, muy ruidosas y en las que existe una fuerte inarmonicidad podríamos aplicar un preproceso de limpieza del ruido de la trama antes de modelarla.**

Este preproceso de limpieza del ruido estaría inspirado en la clasificación de picos espectrales (rama de estudio paralela al modelado) detectando los parciales con las técnicas y algoritmos ya desarrollados y atenuando los pulsos ruidosos.

4.2 PREPROCESADO DE LA TRAMA MEDIANTE CLASIFICACIÓN DE PICOS ESPECTRALES Y POSTERIOR MODELADO POLINOMIAL

Se plantea aplicar el modelado polinomial sobre la trama preprocesada con clasificación de picos y estudiar si se obtiene una señal más limpia y depurada de ruido que las obtenidas con ambos métodos por separado.

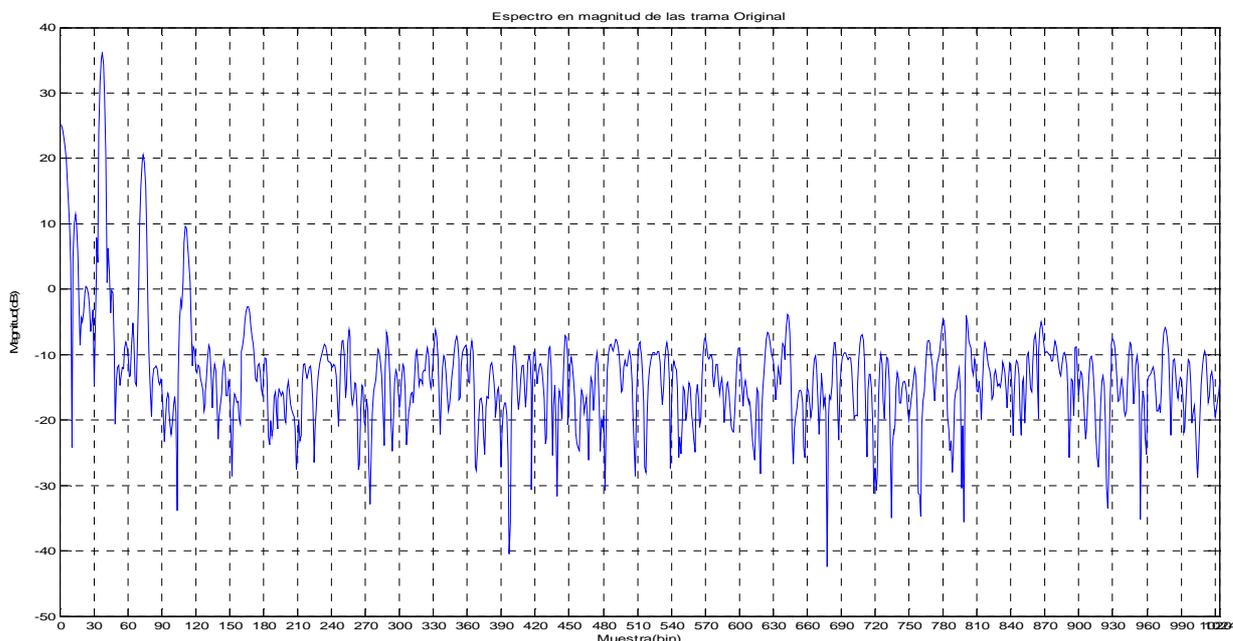
4.2.1 PREPROCESADO Y MODELADO POLINOMIAL TRAMA 5:

Se procesará la trama 5 simulando un proceso de detección/clasificación de picos y posteriormente se modelará el resultado con el método de aproximación polinomial estudiado anteriormente.

4.2.1.1 - Preproceso de limpiado:

Tomamos el espectro original de la trama y lo procesamos ‘manualmente’ para limpiar parte del ruido existente entre los parciales, imitando el efecto que produciría un proceso de clasificación de picos espectrales.

Al trabajar con señales reales, procesamos la mitad del espectro de la trama, tenemos 1024 bins:

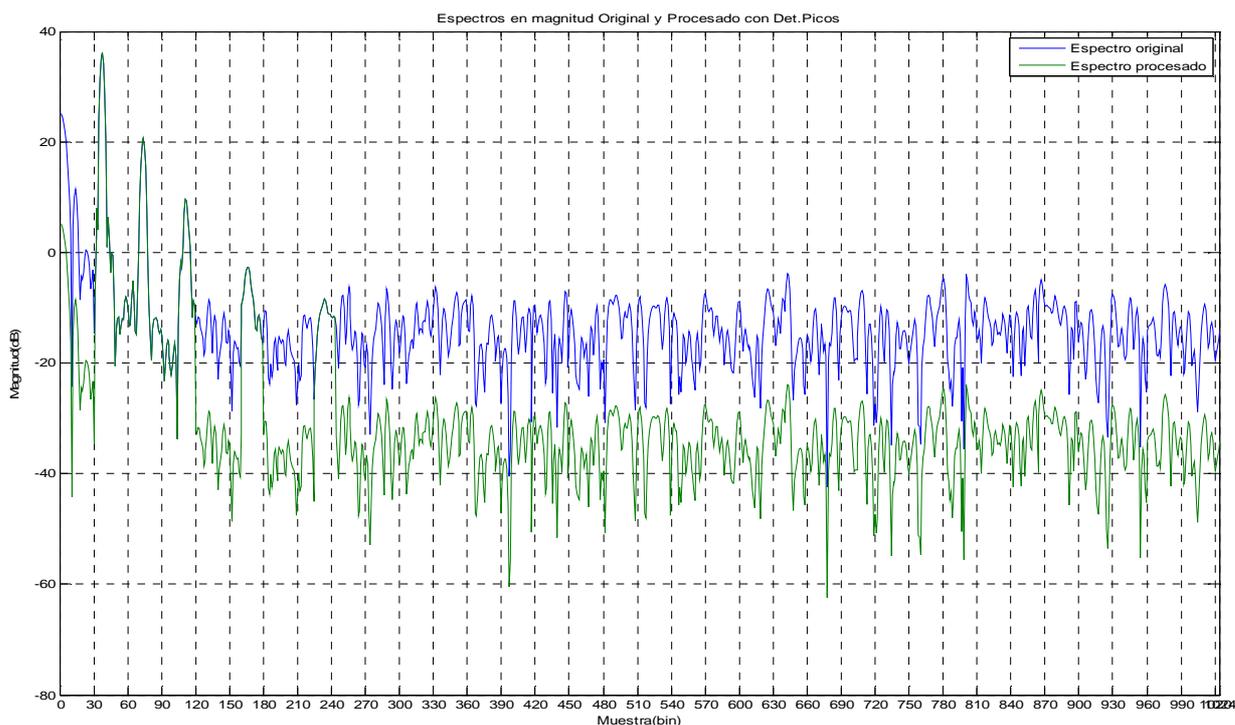


(Fig. 4.2.1.1.1 Espectro de la Trama 5 original, Magnitud vs muestra)

Atenuamos las zonas correspondientes a los **pulsos ruidosos más notables** por un factor L , en este caso $L=10$ (dividimos en lineal entre 10, equivale a restar 20 decibelios). Atenuamos los siguientes rangos muestrales:

[(1-30), (120-160), (180-224), (244-1024)]

Representamos el espectro original junto al procesado:



(Fig. 4.2.1.1.2 Espectros de la Trama 5 original y procesado con Detección de Picos)

Se observa en el espectro procesado como lo 5 primeros parciales han salido reforzados y definidos. El resto de parciales están tan inmersos en el ruido que no queda otro remedio que atenuarlos junto a éste.

La clasificación de picos puede separar los parciales del ruido aunque estos no se encuentren en sus posiciones armónicas, ya que atiende a otros criterios basados en descriptores que estudian entre otros aspectos la forma del pico y su densidad espectral de energía para decidir que picos son ruidosos y que picos son parciales.

El cuarto y quinto parcial son fáciles de extraer del ruido por su forma más propia de un pico determinista, la simulación realizada es válida.

Estamos trabajando sobre la mitad del espectro, para recuperar el espectro total deberemos reflejar respecto a la última muestra el mismo espectro pero conjugado, para tener así simetría hermítica en el espectro total y recuperar en el dominio temporal una señal real.

4.2.1.2 - Modelado de la trama procesada:

A continuación se realizará el modelado polinomial en el dominio temporal, de la trama procesada **aproximando la IFFT (Inverse Fast Fourier Transform) del espectro total** obtenido tras la clasificación de picos.

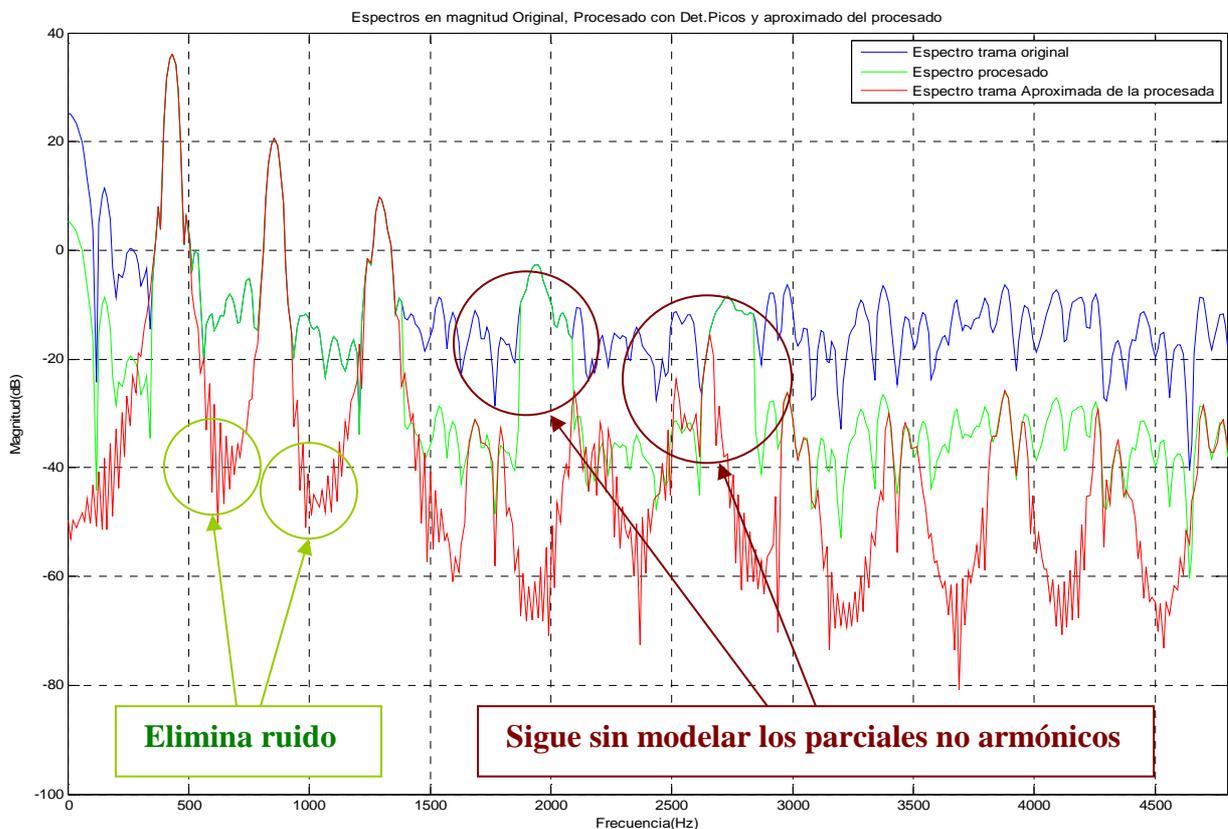
Hay que tener presente que **al devolver el espectro procesado al dominio temporal obtenemos una trama procesada ya enventanada**, ya que al calcular el espectro de la trama original se le aplicó a ésta el enventanado.

Anteriormente se aplicaba el modelado polinomial en el dominio temporal sobre las tramas sin enventanar. El enventanado no es fácil de deshacer. **Para modelar en este caso una trama ya enventanada deberemos modificar el algoritmo de modelado polinomial enventanando la base ortogonal de senos y cosenos para que el efecto de enventanado no afecte al modelado.**

De esta forma conseguimos que la aproximación sea coherente, de no hacerlo el propio enventanado de la trama supondría una fortísima componente de modulación AM, fuera de los límites de lo casi estacionario y no se modelaría correctamente.

Posteriormente **para calcular el espectro de la trama modelada de la procesada no será necesario aplicar nuevamente la ventana.**

Representamos los primeros parciales del espectro de la trama original, del espectro procesado y del espectro de la trama modelada tras ser procesada:



(Fig. 4.2.1.2.1 Espectros de la Trama 5 original, procesado con Detección de Picos y de la trama modelada tras ser procesada)

Se observa como **el modelado polinomial sigue sin aproximar bien los parciales no armónicos** pese a que éstos se encuentran tras el proceso de detección de picos muy definidos y separados del ruido. **El modelado elimina ruido en lo valles donde la clasificación de picos no había actuado**, por tanto en este sentido **ambos métodos se complementan**.

4.2.1.3 - Conclusiones al preprocesado y modelado polinomial de la Trama 5:

Se concluye que **el modelado no funciona bien en tramas que presentan fuerte inarmonicidad en sus parciales**, aún aplicando un preprocesado de clasificación de picos.

La inarmonicidad puede ocurrir para algunos instrumentos en algunas tramas pero es un fenómeno poco probable.

Para **tramas que presenten fuerte inarmonicidad en sus parciales** nos plantearíamos los siguientes caminos a seguir para separar su parte determinista del ruido:

- **Aplicar solamente Clasificación de picos. Ventaja:** extrae los parciales sin que influya que no estén armónicamente distribuidos y elimina bastante ruido. **Desventajas:** no funcionará del todo bien en tramas polifónicas. Mayor carga computacional que el modelado.
- **Utilizar la información obtenida en la clasificación de picos sobre la posición no armónica de los parciales para modelar mediante senos y cosenos en las posiciones obtenidas. Ventaja:** obtendría lo bueno de la clasificación de picos en cuanto a extracción de la posición de los parciales y lo bueno del modelado en cuanto a eliminación del ruido. **Desventaja:** seguiría sin funcionar bien en tramas polifónicas y generaría mucha carga computacional.

Como la inarmonicidad es un fenómeno poco probable, **no profundizaremos en su estudio y asumiremos cierto error en el modelado cuando se produzca**.

Para **tramas con mucha modulación AM-FM y mucho ruido** se plantea:

- **Modelar polinomialmente con grado de polinomio 6. Ventaja:** no produce Subfit y elimina ruido. **Desventaja:** Overfit.
- **Modelar polinomialmente tras clasificación de picos. Ventaja:** puede eliminar más ruido y complementa en este sentido la detección de picos. **Desventajas:** carga computacional mayor.
- **Aplicar modelado lineal por tramos.** Ventajas y desventajas se estudiarán posteriormente.

Antes de concluir que camino elegir para la separación de la parte determinista en tramas con mucha modulación y mucho ruido (obviaremos si tienen además fuerte inarmonicidad) **estudiaremos más profundamente las posibles ventajas de combinar modelado y clasificación de picos en otras tramas**.

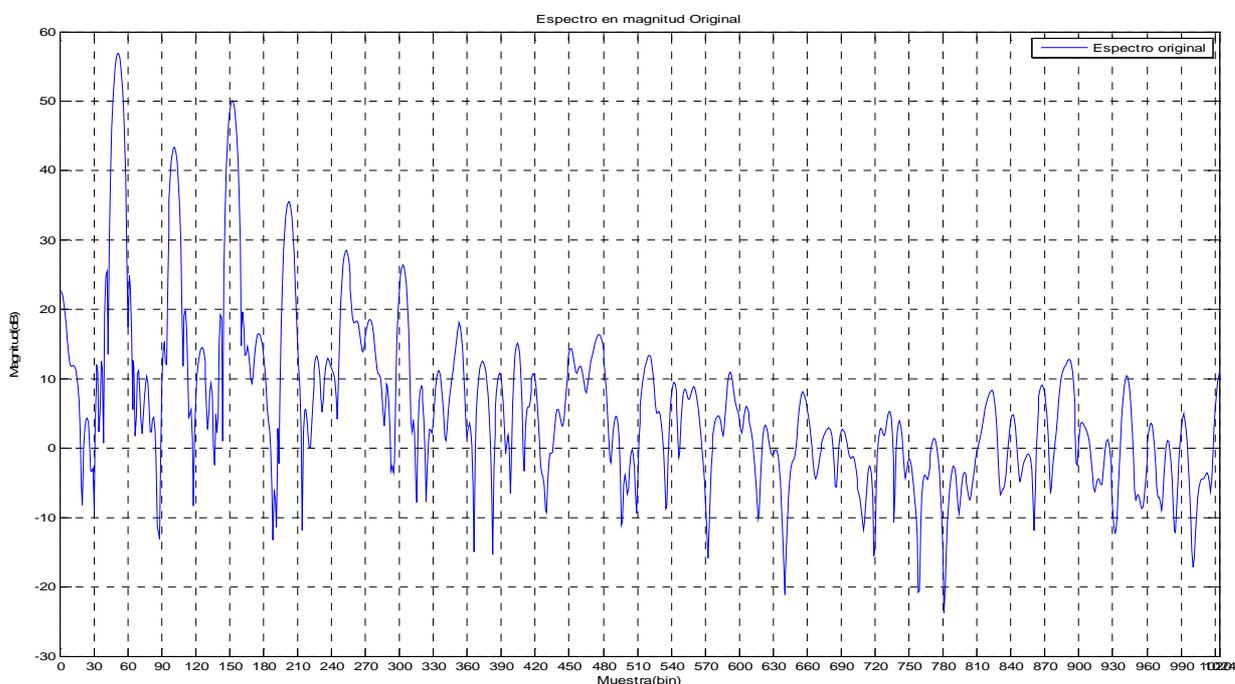
4.2.2 PREPROCESADO Y MODELADO POLINOMIAL TRAMA 4:

Estudiaremos a continuación el mismo planteamiento sobre una trama con los **parciales armónicamente distribuidos, poca modulación y bastante ruido** (trama 4).

Se procesará la trama simulando un proceso de clasificación de picos y posteriormente se modelará el resultado con el método de aproximación polinomial. Estudiaremos si ambos métodos aplicados conjuntamente obtienen mejores resultados que por separado en cuanto a eliminación del ruido.

4.2.2.1 - Preproceso de limpiado:

Tomamos la mitad del espectro de la trama, 1024 bins:

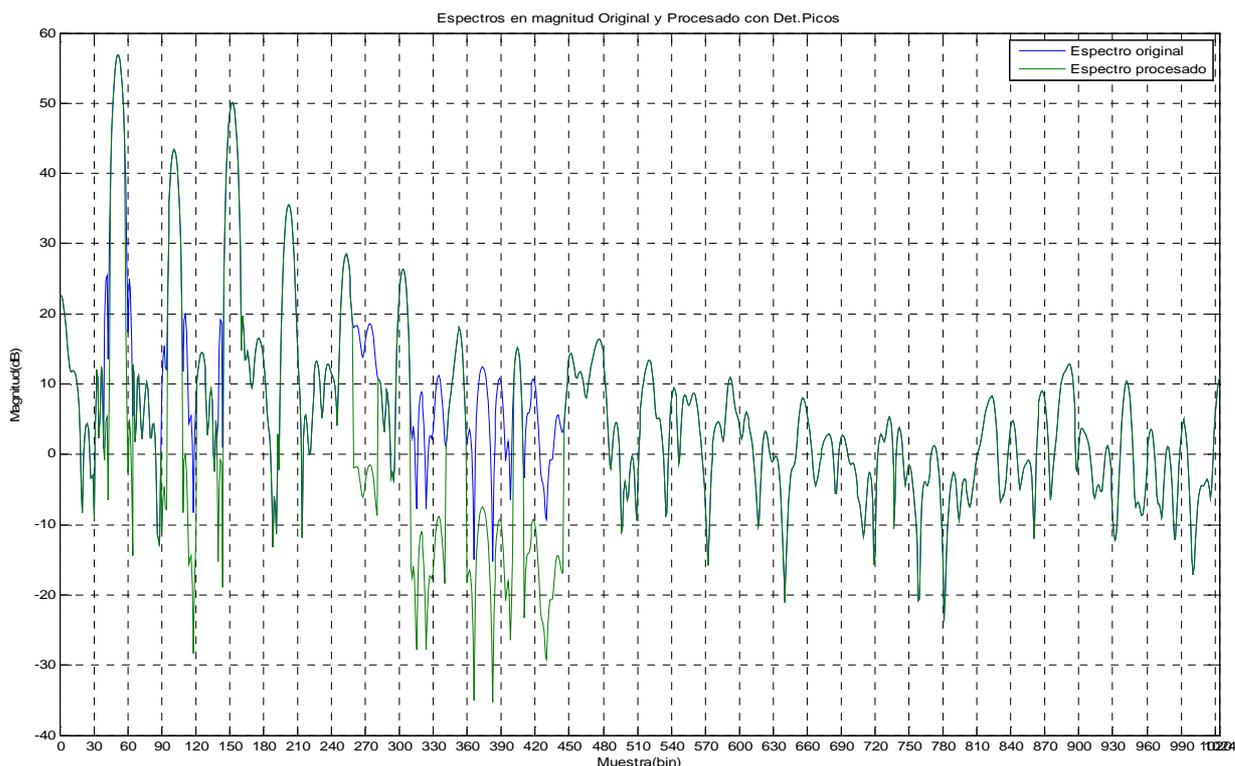


(Fig. 4.2.2.2.1 Espectro de la Trama 4 original, Magnitud vs muestra)

Atenuamos las zonas correspondientes a los **pulsos ruidosos más notables** por un factor L , en este caso $L=10$ (dividimos en lineal entre 10, equivale a restar 20 decibelios). Atenuamos los siguientes rangos muestrales:

[(40-44),(58-64),(90-94),(109-120),(140-144),(260-280),(310-340),(360-400),(410-444)]

Representamos el espectro original junto al procesado:



(Fig. 4.2.1.1.2 Espectros de la Trama 5 original y procesado con Detección de Picos)

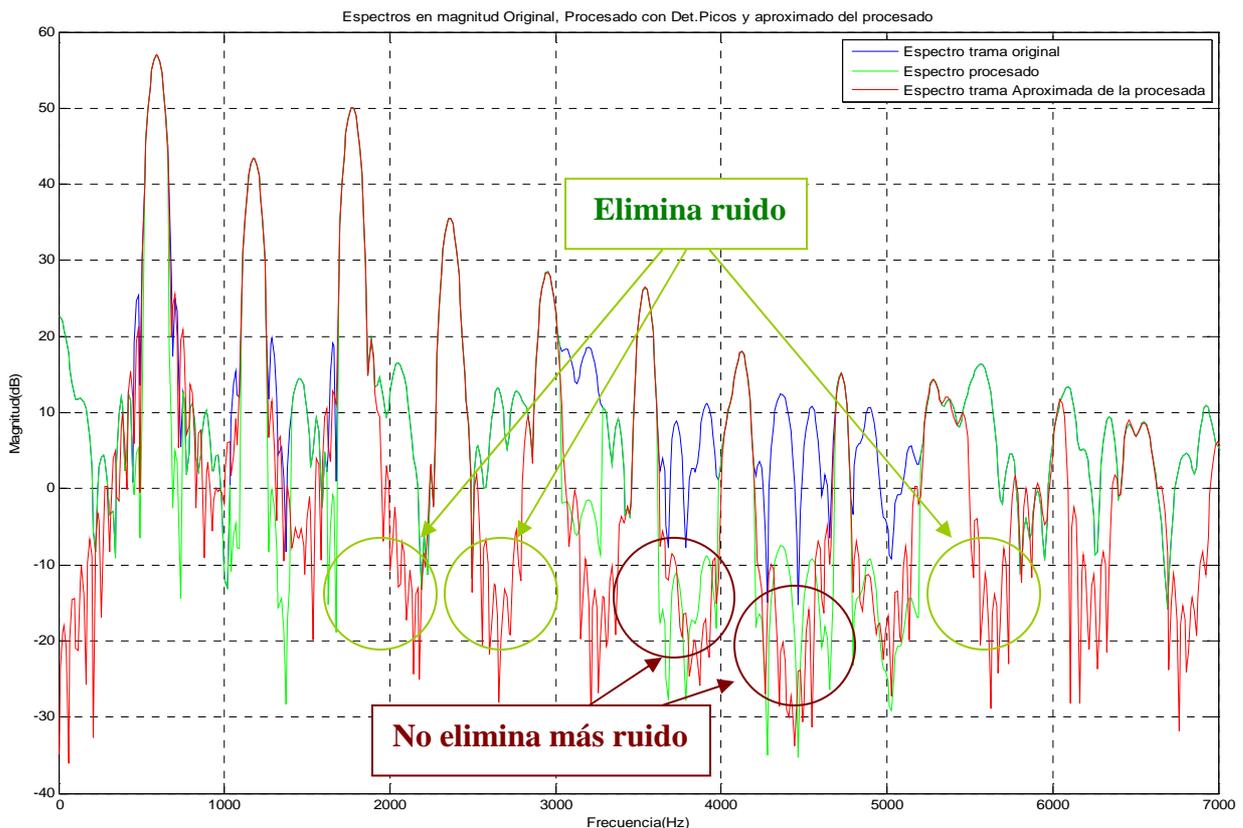
En este caso limpiamos únicamente los parciales donde se concentra la mayor parte de la energía sin atenuar la parte superior del espectro para no añadir más distorsión. **Los 8 primeros parciales han salido reforzados y definidos.**

Recuperamos el espectro total reflejando respecto a la última muestra el mismo espectro pero conjugado, obtenemos el espectro procesado total.

4.2.2.2 - Modelado de la trama procesada:

Realizamos el modelado polinomial **en el dominio temporal**, de la trama procesada **aproximando la IFFT del espectro total** obtenido tras la detección de picos, mediante el algoritmo modificado para poder modelar sin que afecte el enventanado de la trama procesada.

Representamos juntos el espectro de la trama original, el espectro procesado y el espectro de la trama modelada tras ser procesada:



(Fig. 4.2.2.2.1 Espectros de la Trama 4 original, procesado con Detección de Picos y de la trama modelada tras ser procesada)

Se observa como en las zonas donde no actuó la clasificación de picos y se dejaron los valles como en la trama original, el modelado ha eliminado ruido, modelando los valles más profundos. En cambio en las zonas donde ha actuado la clasificación de picos el modelado no consigue eliminar más ruido, se queda al mismo nivel que en la trama procesada.

4.2.2.3 - Conclusiones al preprocesado y modelado polinomial de la Trama 4:

Se intuye que el polinomio de aproximación no es capaz de eliminar más ruido en las zonas donde la clasificación ha actuado, en todo caso **ambos métodos se complementan**, ya que en las zonas donde no actúa la clasificación de picos, el **modelado termina limpiándolas**.

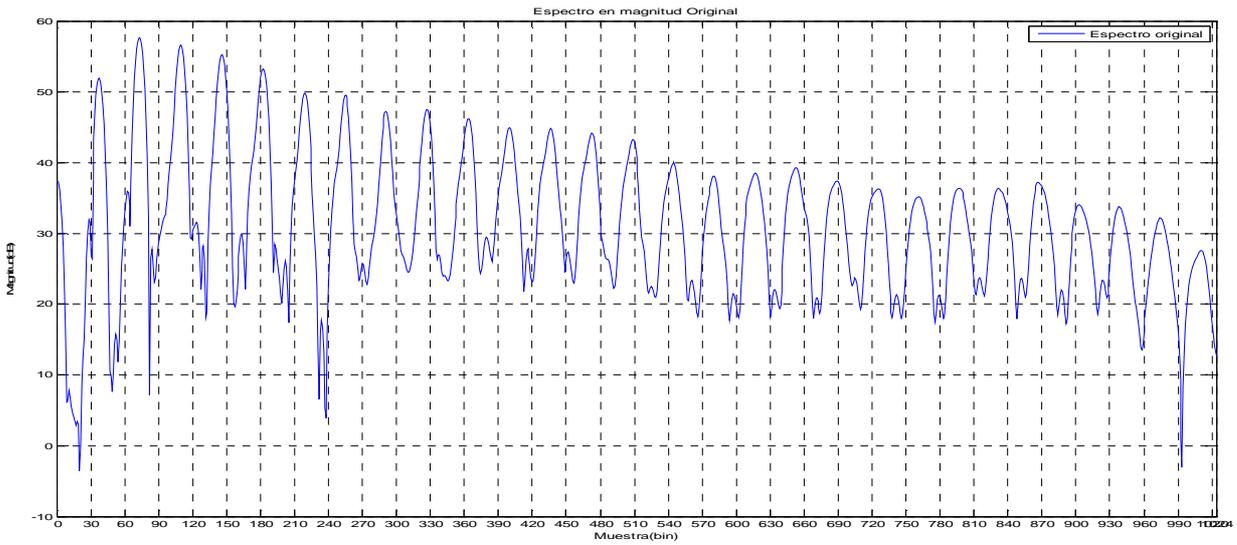
Nos preguntamos **que pasaría si el proceso de clasificación de picos funcionara eficazmente y eliminara la mayor parte del ruido presente en la trama, ¿el modelado polinomial posterior serviría de algo? ó ¿sería cargar computacionalmente el algoritmo sin obtener mejoras?** lo comprobaremos aplicando el mismo proceso sobre una trama con poco ruido y sus parciales armónicamente distribuidos (trama 2).

4.2.3 PREPROCESADO Y MODELADO POLINOMIAL TRAMA 2:

Aplicaremos la aproximación polinomial sobre la trama 2 (trama con poco ruido) preprocesada con clasificación de picos, siguiendo los mismos pasos que en las tramas anteriores.

4.2.3.1 - Preproceso de limpiado:

Tenemos el siguiente espectro:

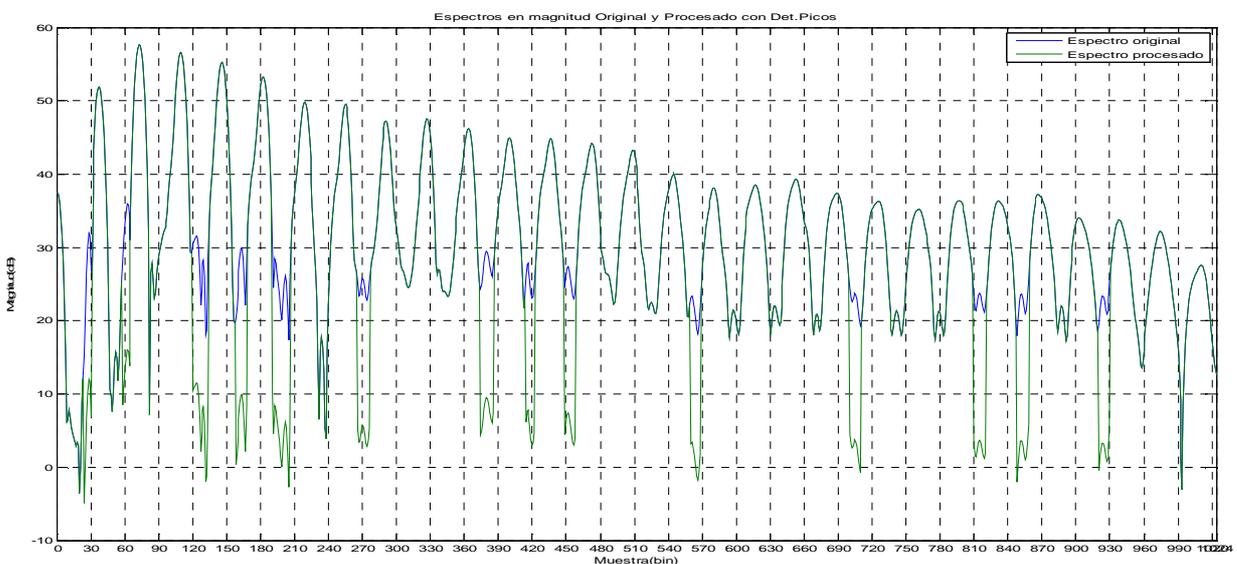


(Fig. 4.2.3.1.1 Espectro de la Trama 2 original, Magnitud vs muestra)

Atenuamos los siguientes rangos muestrales por un factor **L=10**:

[(24-30),(58-64),(120-134),(158,168),(190-206),(266-276),(374-386),(414-422),(448-458),(560-568),(700-710),(810-820),(848-858),(920-930)]

Representamos el espectro original junto al procesado:



(Fig. 4.2.3.1.2 Espectros de la Trama 2 original y procesado con Detección de Picos)

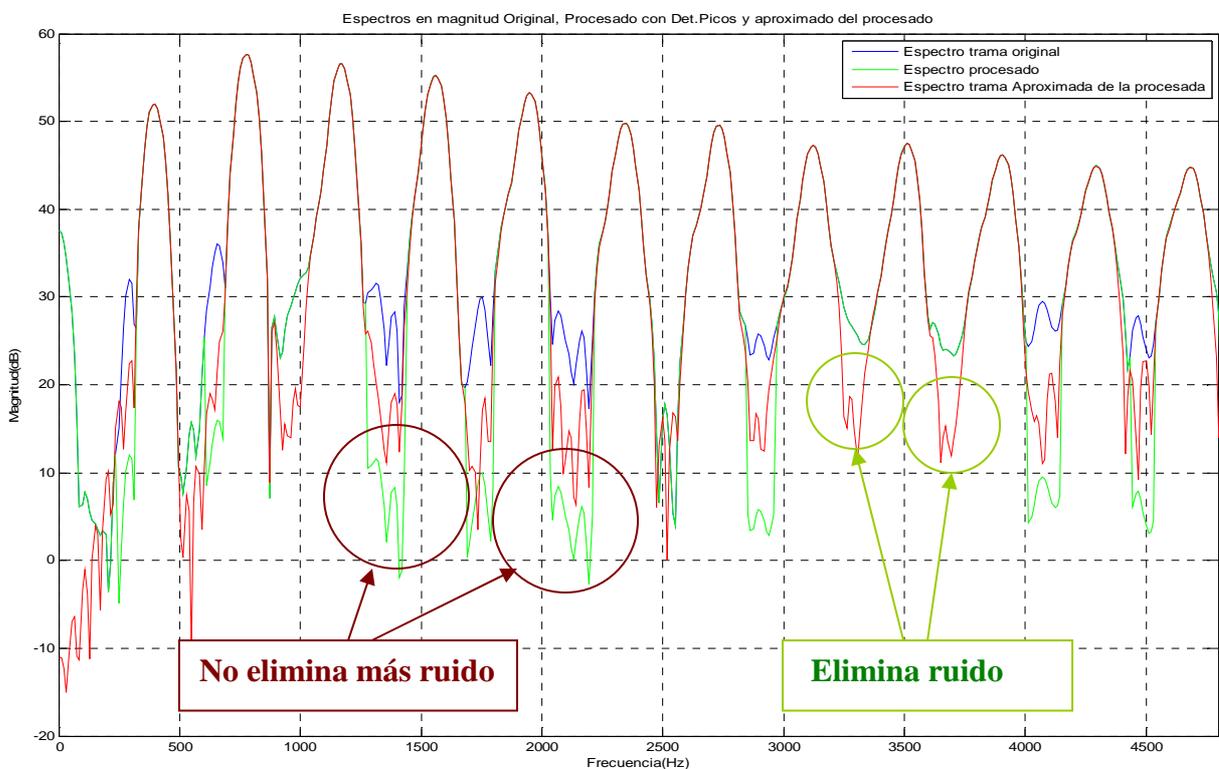
Dejamos algunas zonas intencionadamente sin limpiar; la trama no tenía apenas ruido, obtenemos una trama muy limpia tras procesarla.

Recuperamos el espectro total reflejando respecto a la última muestra el mismo espectro pero conjugado, obtenemos el espectro procesado total.

4.2.3.2 - Modelado de la trama procesada:

Realizamos el modelado polinomial de la trama procesada aproximando la IFFT del espectro total obtenido tras la clasificación de picos mediante el algoritmo modificado.

Representamos juntos el espectro de la trama original, el espectro procesado y el espectro de la trama modelada tras ser procesada:



(Fig. 4.2.3.2.1 Espectros de la Trama 2 original, procesado con Detección de Picos y de la trama modelada tras ser procesada)

Se observa nuevamente para esta trama como en las zonas donde intencionadamente no procesábamos con detección de picos el modelado ha eliminado ruido, pero **en las zonas donde la detección de picos actúa el modelado no consigue eliminar más ruido.**

4.2.3.3 - Conclusiones al preprocesado y modelado polinomial de la Trama 2:

Parece como si el polinomio de aproximación no pudiera bajar más abajo en los valles donde la clasificación de picos ha funcionado aunque **el modelado complementaría las zonas donde detección de picos no actuara.**

Nos preguntamos ahora, que pasaría si en el proceso de detección de picos en lugar de atenuar los valles por un factor $L=10$ directamente se quedara únicamente con los parciales. Es decir, **si se llevaran algunos de los valles a menos infinito ¿sería capaz el polinomio de aproximación de bajar tan abajo al modelar la trama así procesada?**

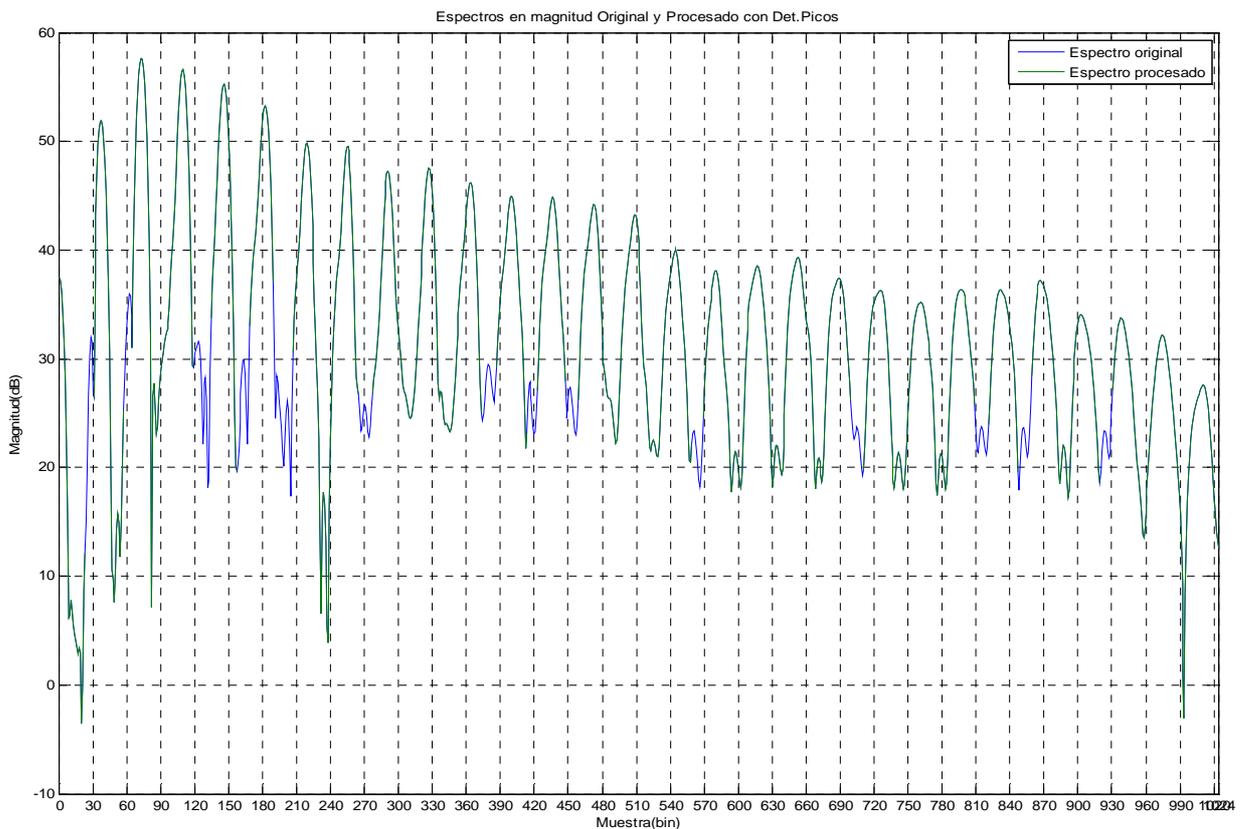
4.2.4 PREPROCESADO ANULANDO LOS PICOS RUIDOSOS Y MODELADO POLINOMIAL TRAMA 2:

En este experimento queremos estudiar que sucede si **en lugar de atenuar por un factor L los picos ruidosos los anulamos directamente**, llevándolos a menos infinito en el espectro en decibelios.

4.2.4.1 - Preproceso de limpiado:

Procesamos los mismos picos ruidosos que en el caso anterior para la trama 2, pero ahora con $L=0$.

Representamos el espectro original junto al procesado:

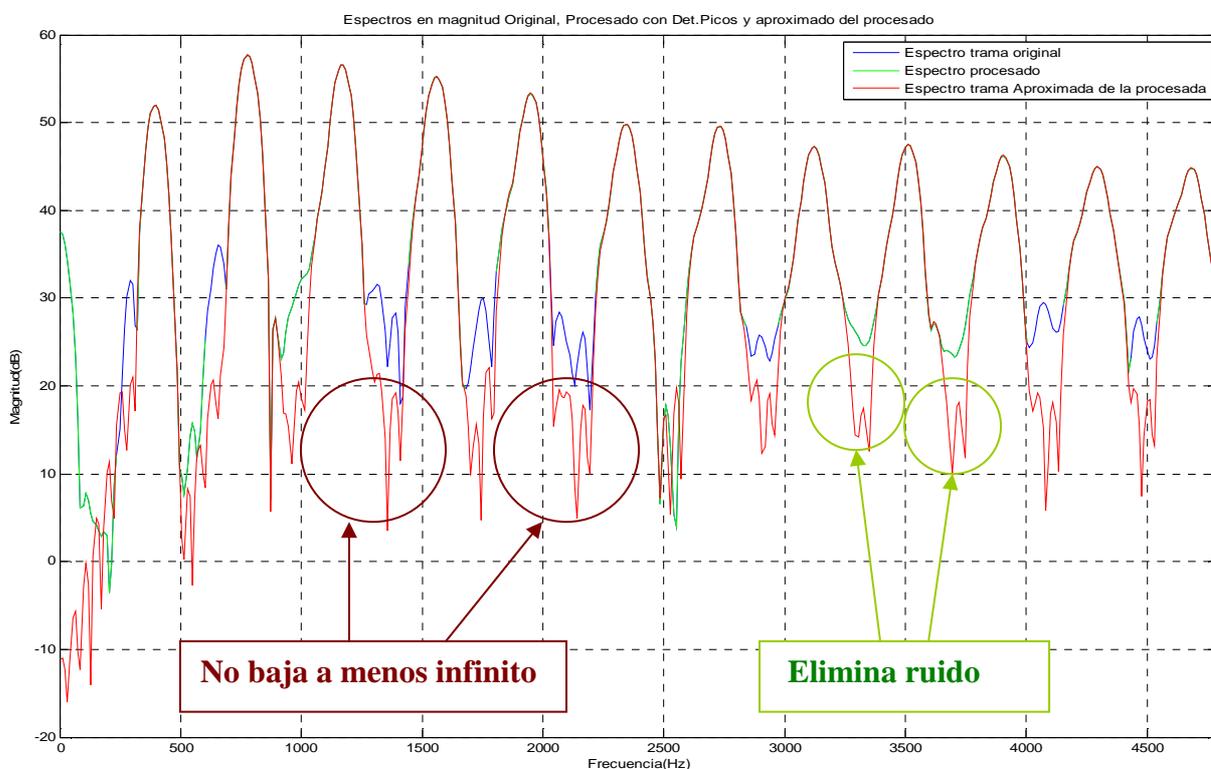


(Fig. 4.2.4.1.1 Espectros de la Trama 2 original y procesado con Detección de Picos, $L=0$)

Se observa como en este caso **los picos ruidosos son llevados a menos infinito**.

4.2.4.2 - Modelado de la trama procesada:

Realizamos el modelado polinomial de la trama procesada. Representamos juntos el espectro de la trama original, el espectro procesado y el espectro de la trama modelada tras ser procesada:



(Fig. 4.2.4.2.1 Espectros de la Trama 2 original, procesado con Detección de Picos y de la trama modelada tras ser procesada, $L=0$)

Se observa como el polinomio de aproximación es incapaz de bajar más abajo pese a que la trama que está modelando (la procesada) tiene los valles en menos infinito. Al modelar no consigue eliminar más ruido, al contrario. Para los picos ruidosos que no son procesados con clasificación de picos el modelado si elimina más ruido.

4.2.5 CONCLUSIONES AL PREPROCESADO MEDIANTE CLASIFICACIÓN DE PICOS Y POSTERIOR MODELADO:

El polinomio de aproximación polinomial (y el lineal por tramos) modula una envolvente sinusoidal con una tendencia continua, que no puede seguir la variación brusca de la detección de picos y por tanto no puede mejorar sus resultados. **El modelado no eliminará una mayor cantidad de ruido en las zonas procesadas con clasificación de picos.** (Se ha comprobado con los dos tipos de modelado, por simplicidad se omite el estudio similar de preprocesado y modelado lineal por tramos)

Actualmente los algoritmos desarrollados para la clasificación de picos **en tramas monofónicas** funcionan eficazmente, detectando y limpiando los parciales en todo su espectro, por tanto **aplicar modelado tras clasificación de picos no obtendría mejora y sería cargar en vano computacionalmente.** Para **tramas polifónicas** donde se producen solapamientos entre parciales y la clasificación de picos no funciona del todo bien, **el modelado podría mejorar sus resultados, no profundizaremos más el estudio.**

4.3 ESTUDIO DEL OVERFIT EN EL MODELADO LINEAL POR TRAMOS

A continuación estudiaremos el segundo algoritmo planteado para el modelado no estacionario de la parte determinista, el modelado lineal por tramos, en el que la base de senos y cosenos es ponderada por funciones triangulares (segmentos rectos) con lo que se consigue describir variaciones rápidas en las modulaciones AM-FM dentro de la trama.

Este es el algoritmo adecuado para modelar tramas transitorias y con mucha modulación, por ello **lo aplicaremos a las dos tramas con mayor modulación** que hemos estudiado hasta ahora, **tramas 3 y 5**.⁴ Trataremos de **buscar el valor óptimo de la variable breaks**, que define el número de segmentos con el que modelar la variación en la modulación AM-FM.

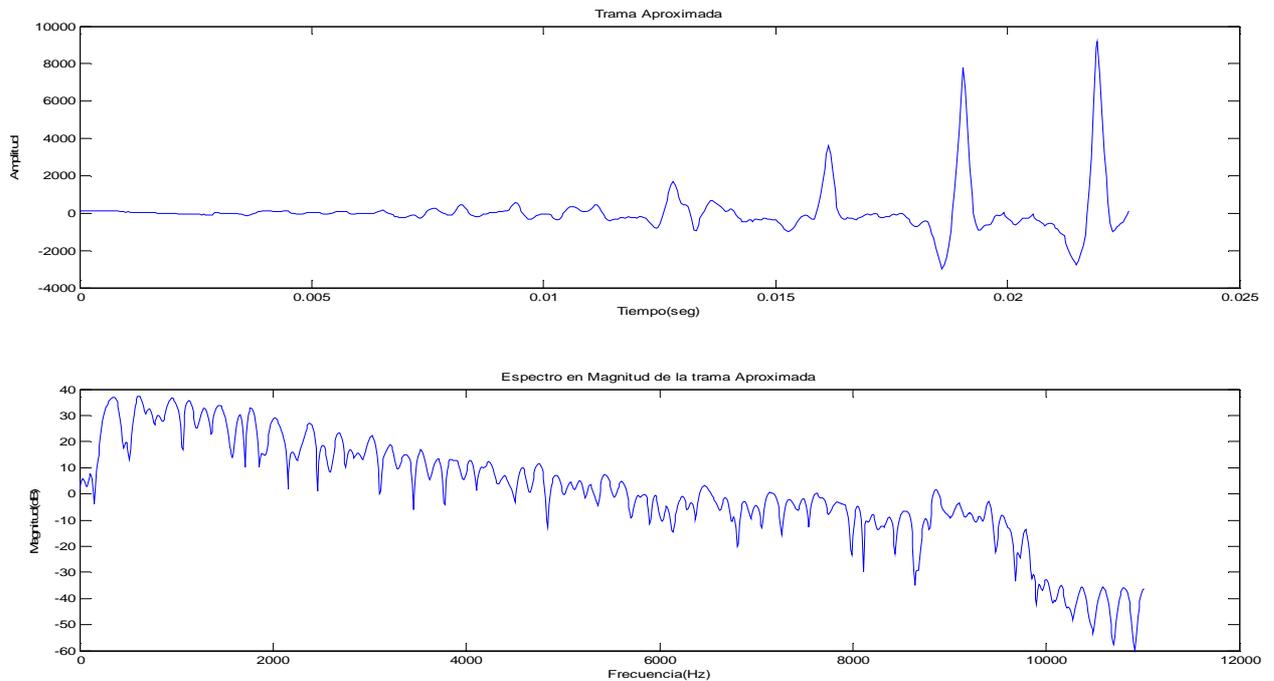
⁴ El modelado lineal por tramos está implementado en la función “**pwl_ls.m**” y se aplica sobre las tramas sin inventanar. Tendremos: “**x**” trama original, “**xapp_pwl_valorBreaks**” trama aproximada con valor Breaks y “**z**” residuo o ruido ($z=x-xapp_pwl_valorBreaks$). Aplicamos la FFT sobre dichas señales inventanadas y obtenemos: “**X**” espectro de la trama original, “**Xapp_pwl_valorBreaks**” espectro de la trama aproximada y “**Z**” espectro del residuo. Todas las variables y resultados obtenidos son guardados en archivos de variables Matlab, uno para cada trama.

4.3.1 MODELADO LINEAL POR TRAMOS TRAMA 3:

Trama transitoria poco ruidosa. Trumpet

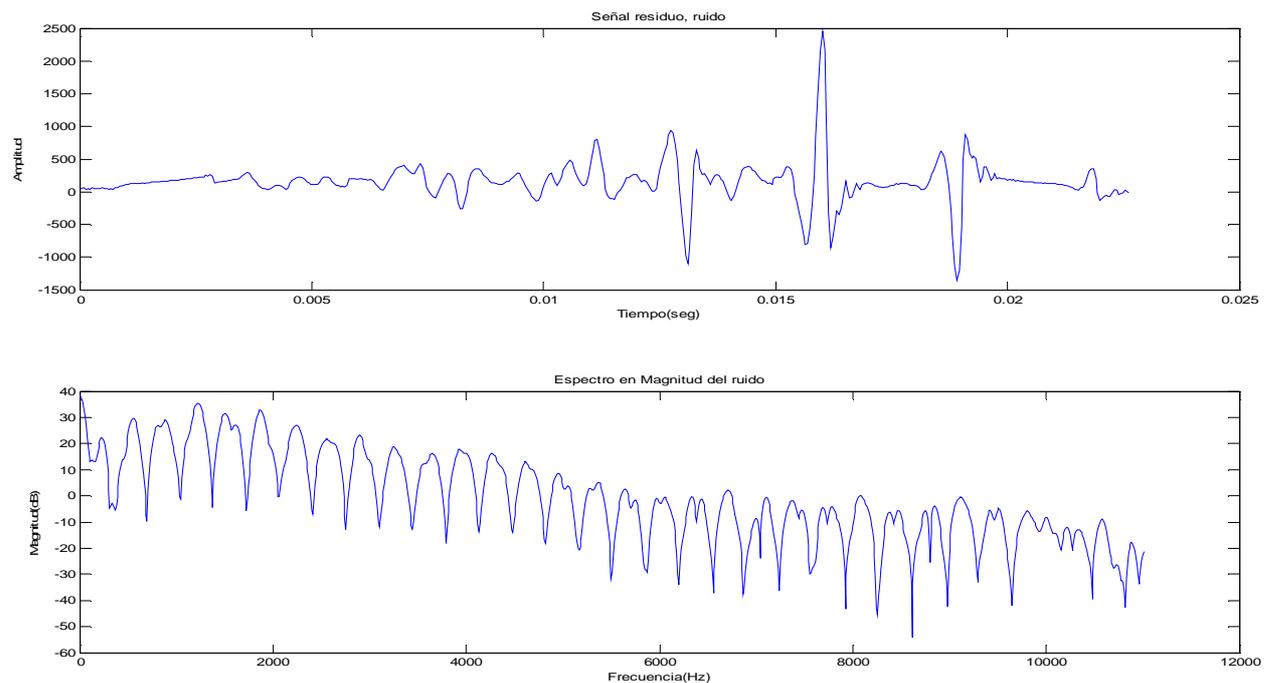
4.3.1.1 - Número de armónicos=28, Breaks=5.

Trama aproximada en el tiempo y su espectro:



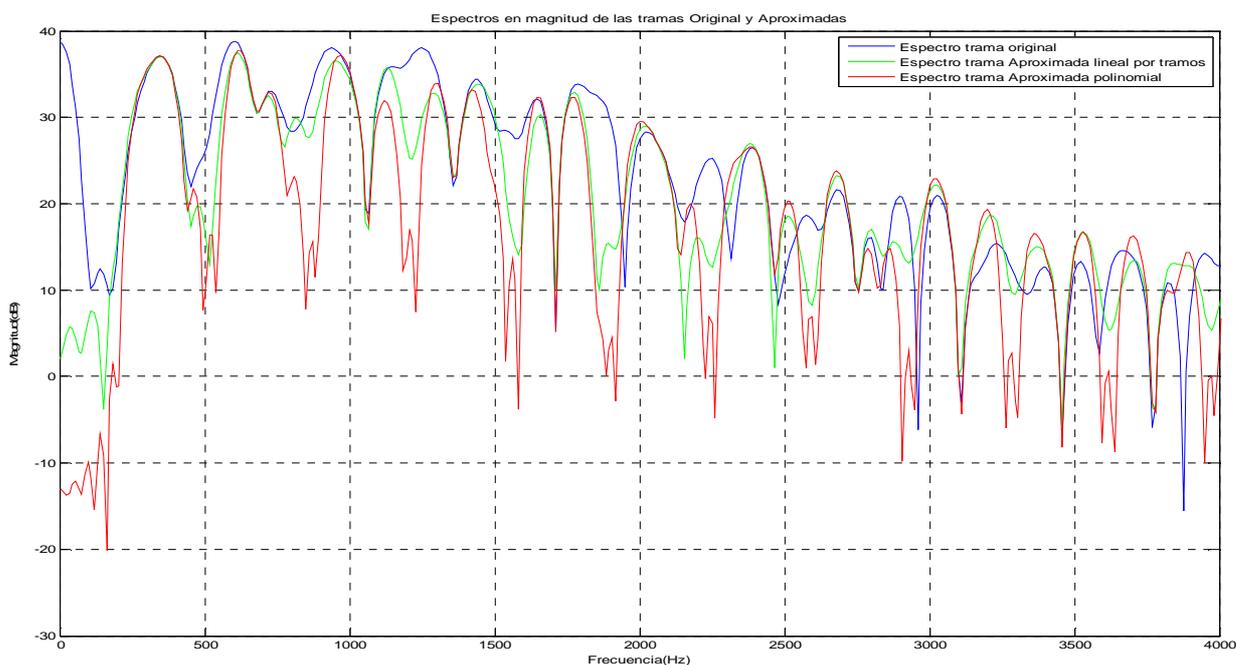
(Fig. 4.3.1.1.1 Trama 3 Modelada lineal por tramos, Breaks=5, Nharm=28)

Residuo:



(Fig. 4.3.1.1.2 Residuo o ruido de la Trama 3, Breaks=5, Nharm=28)

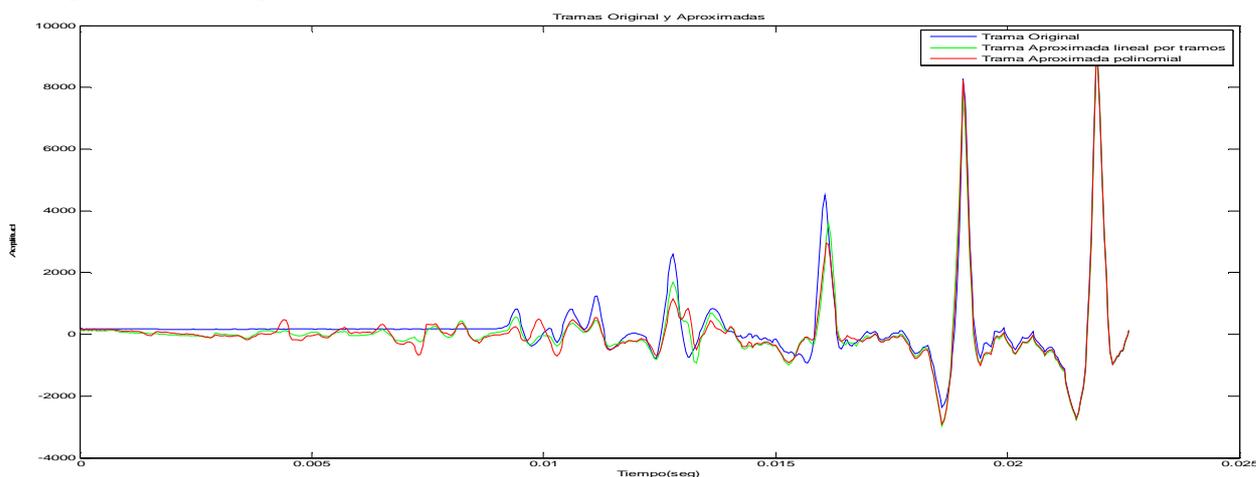
Representamos juntos los espectros de la trama original, el de la trama aproximada lineal por tramos y el de la aproximada polinomialmente de grado 6 (mayor grado):



(Fig. 4.3.1.1.3 Espectros de la Trama 3 original, aproximada lineal por tramos y aproximada polinomial, Breaks=5, P=6, Nharm=28)

Se observa como **la aproximación lineal por tramos sigue mejor los pulsos un tanto erráticos del espectro de la trama transitoria**, pese a modelar doubles picos en los solapamientos entre parciales de la trama original como sucede en el cuarto/quinto parcial.

Representamos la trama original junto a las aproximadas lineal por tramos y aproximada polinomial de grado 6:



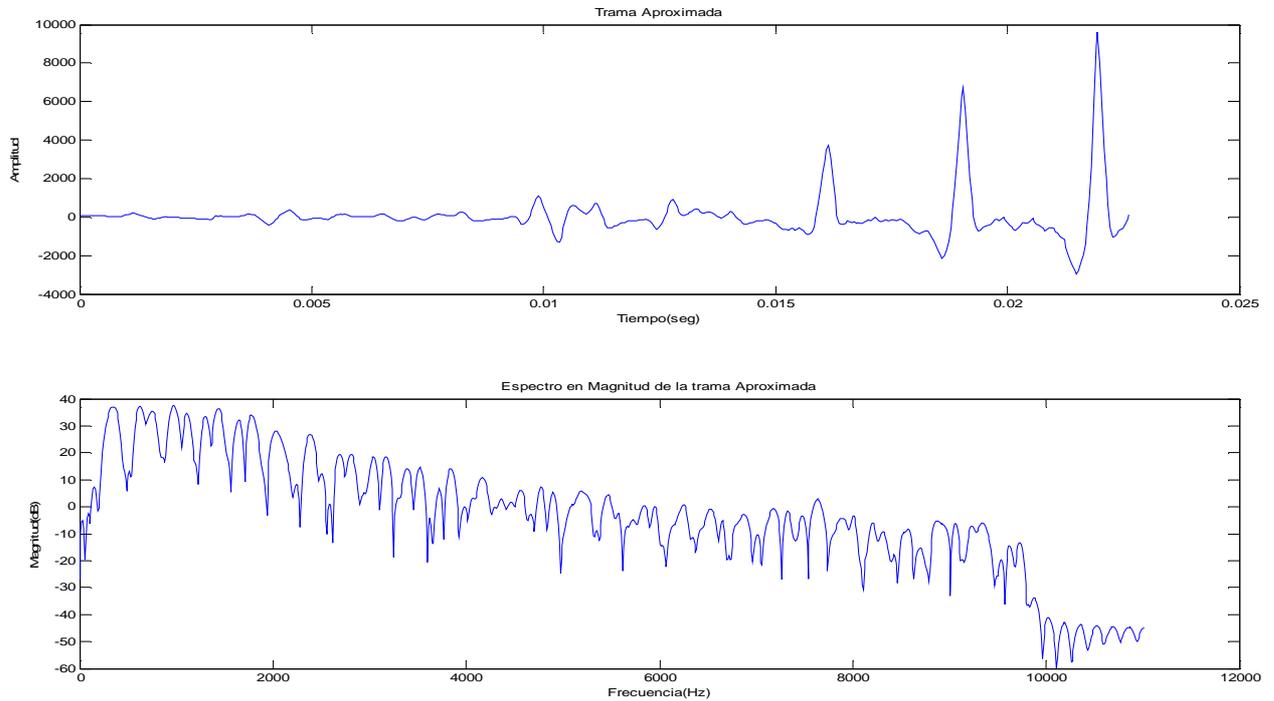
(Fig. 4.3.1.1.4 Trama 3 original, aproximada lineal por tramos y aproximada polinomial, Breaks=5, P=6, Nharm=28)

Se observa como **la aproximación lineal por tramos (en verde), sigue mejor la variación AM de la trama original asociada al ataque inicial**.

Veremos que sucede si bajamos el valor de Breaks.

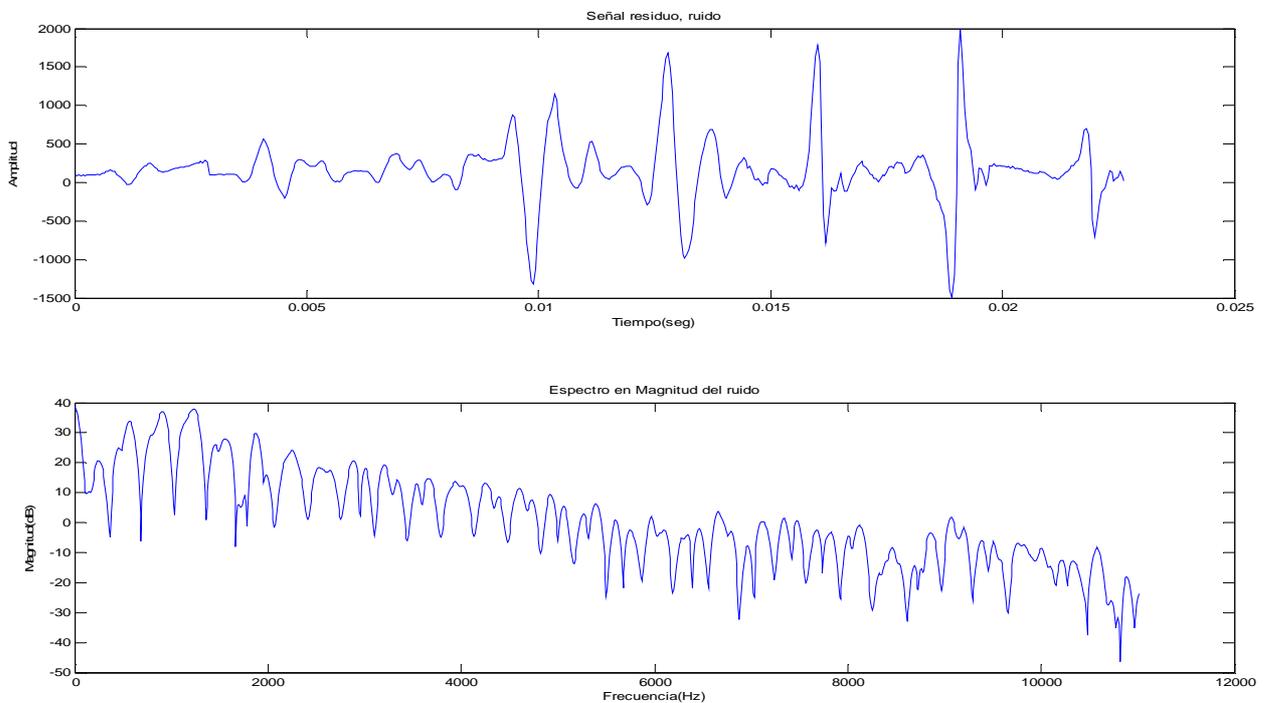
4.3.1.2 - Número de armónicos=28, Breaks=4.

Trama aproximada en el tiempo y su espectro:



(Fig. 4.3.1.2.1 Trama 3 Modelada lineal por tramos, Breaks=4, Nharm=28)

Residuo:

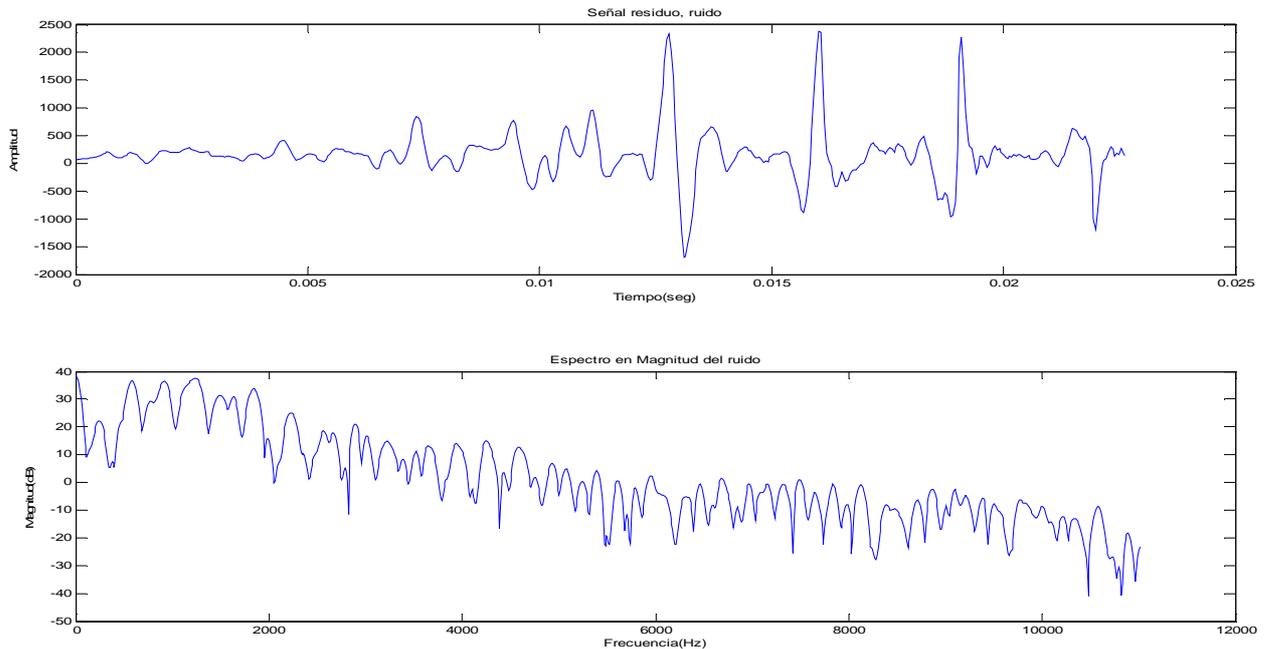


(Fig. 4.3.1.2.2 Residuo o ruido de la Trama 3, Breaks=4, Nharm=28)

En este caso se observa como en el residuo hay más pulsos transitorios, **se modela peor la parte determinista.**

4.3.1.3 - Número de armónicos=28, Breaks=3.

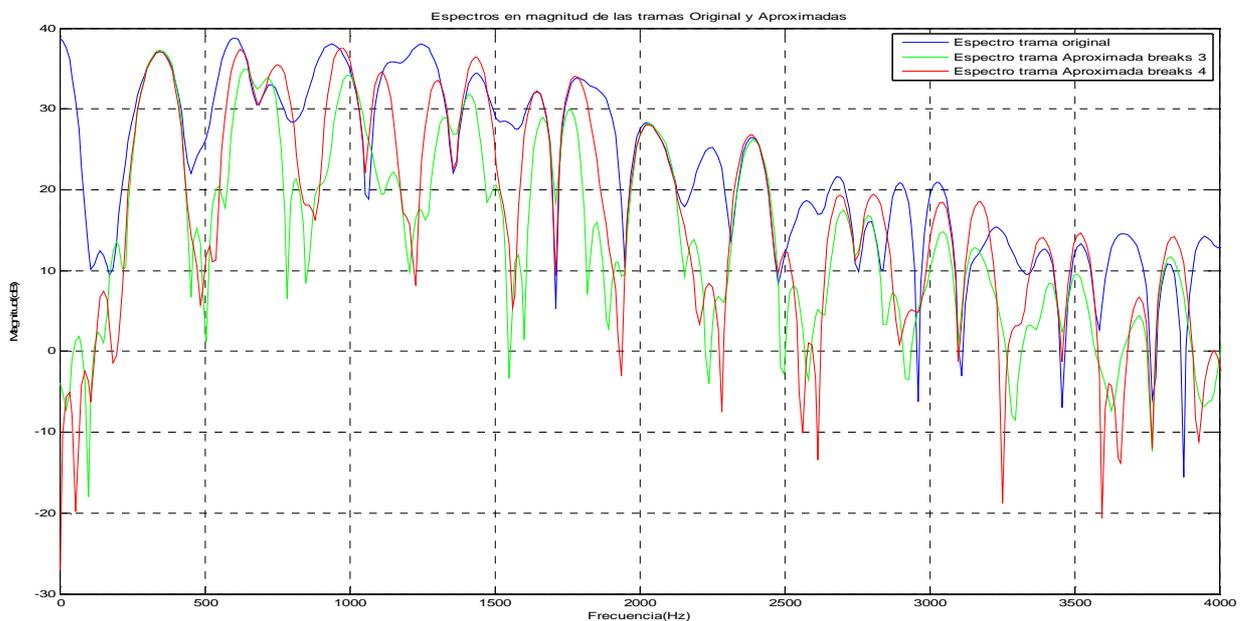
Representaremos únicamente el residuo de la aproximación:



(Fig. 4.3.1.2.3 Residuo o ruido de la Trama 3, Breaks=3, Nharm=28)

Se tiene **mucho componente determinista en el residuo, no se modela bien.**

Representamos juntos los espectros de la trama original junto a las aproximaciones lineales por tramos con Breaks 4 y 3:

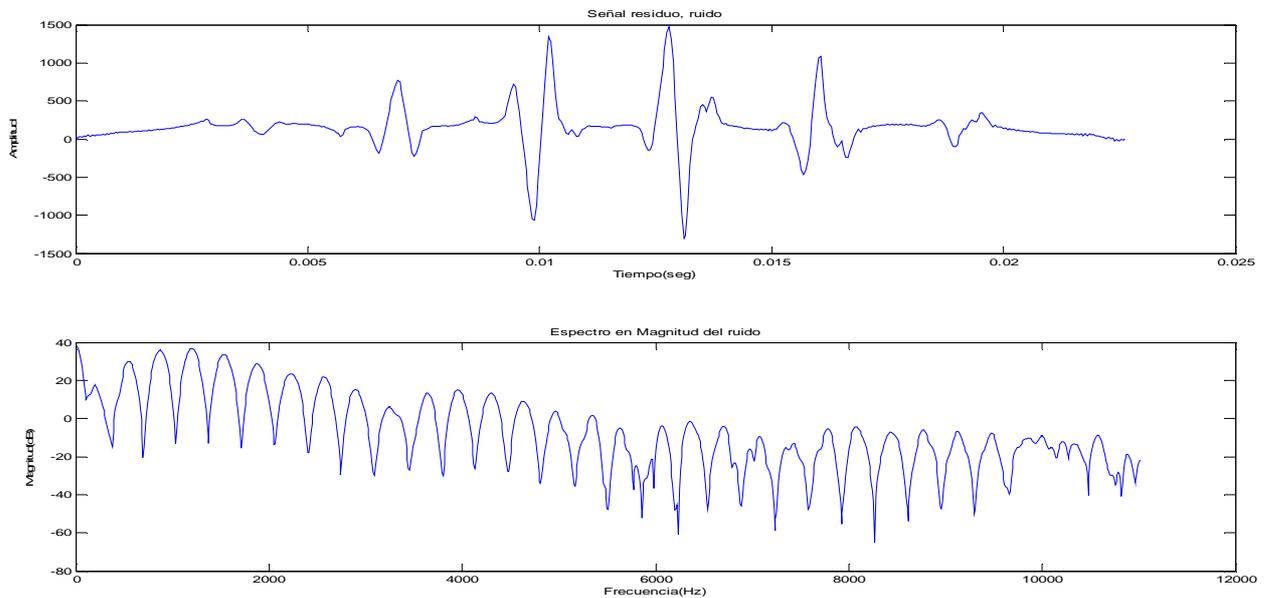


(Fig. 4.3.1.2.4 Espectros de la Trama 3 original y aproximadas lineal por tramos Breaks=3 y 4, Nharm=28)

Al bajar el valor de Breaks se produce un fenómeno de **Subfit** similar al que teníamos en el modelado polinomial. Con **Breaks=3** se observa como **los parciales modelados ya no tienen mucho que ver con los originales salvo el primero.** En el dominio temporal **al disminuir Breaks aumenta la parte determinista en el residuo.**

4.3.1.4 - Número de armónicos=28, Breaks=6.

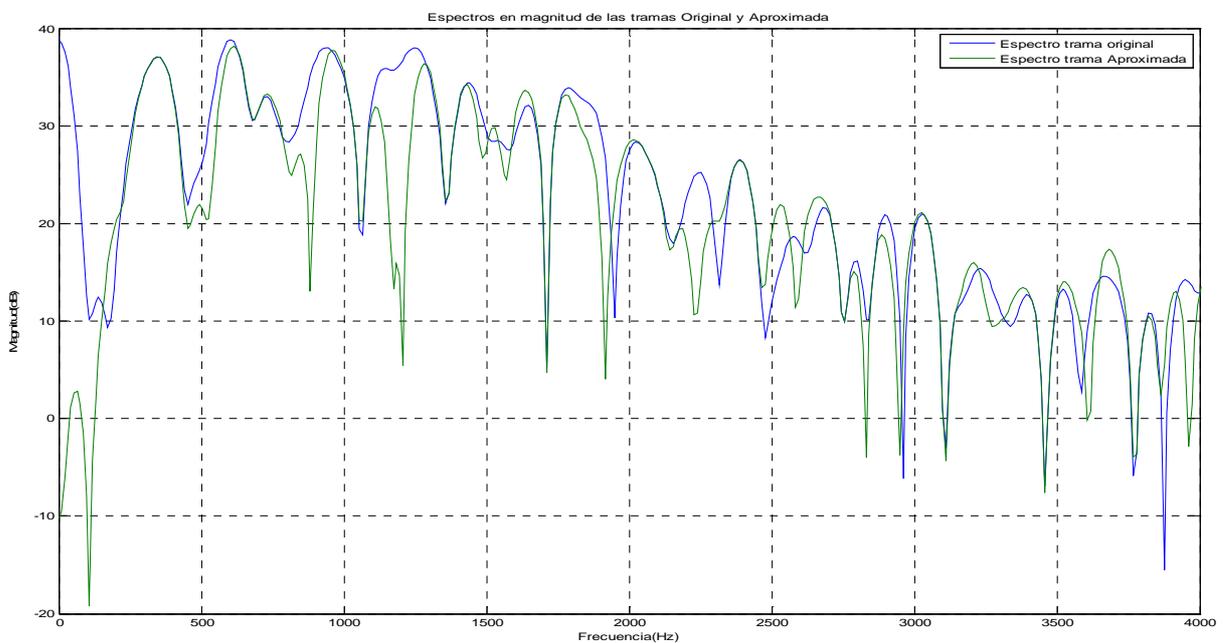
Para el residuo de la aproximación tenemos:



(Fig. 4.3.1.4.1 Residuo o ruido de la Trama 3, Breaks=6, Nharm=28)

En el dominio temporal el nivel del residuo ha bajado pero **ya no tenemos tanta componente determinista.**

En el espectro tenemos:

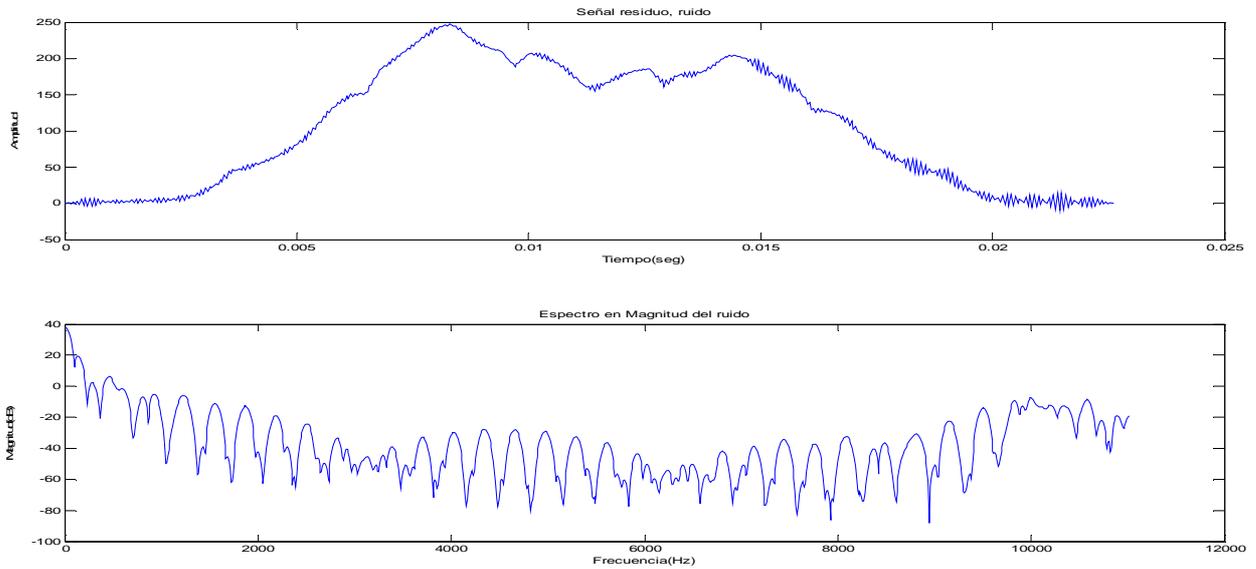


(Fig. 4.3.1.4.2 Espectros de la Trama 3 original y aproximada lineal por tramos, Breaks=6, Nharm=28)

Observamos lo mismo, ya no se elimina tanto ruido entre los parciales.

4.3.1.5 - Número de armónicos=28, Breaks=7.

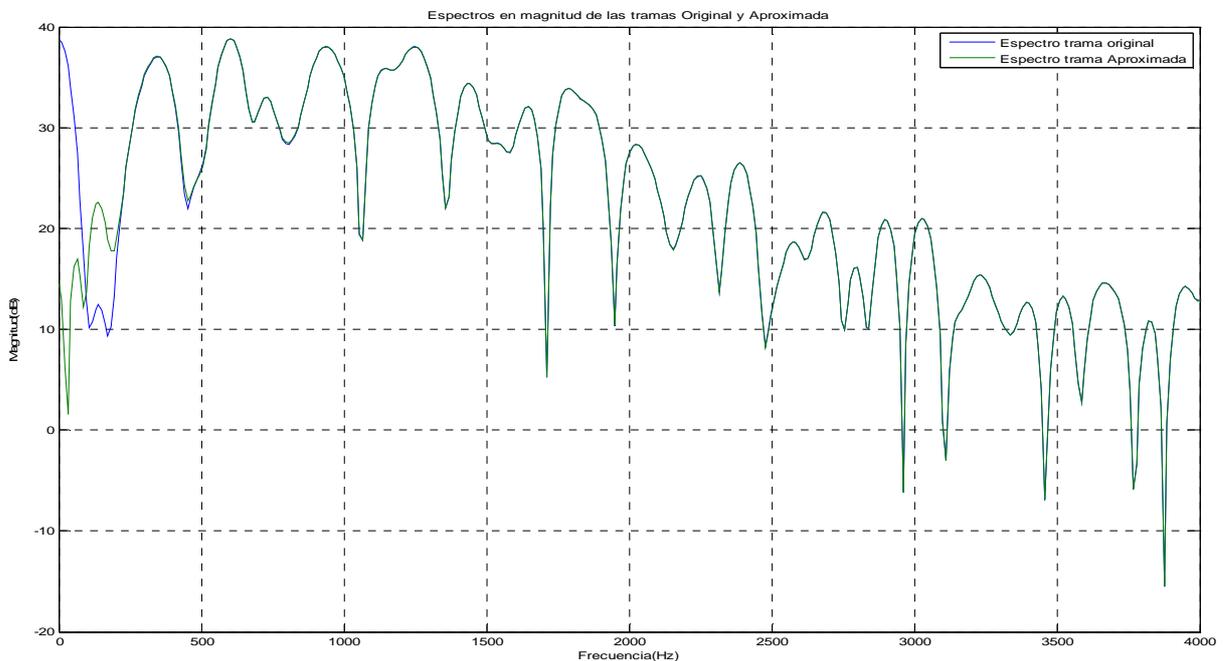
Para el residuo tenemos:



(Fig. 4.3.1.5.1 Residuo o ruido de la Trama 3, Breaks=7, Nharm=28)

Se obtiene un **residuo con un nivel bajísimo, no se separa apenas ruido.**

En el espectro tenemos:



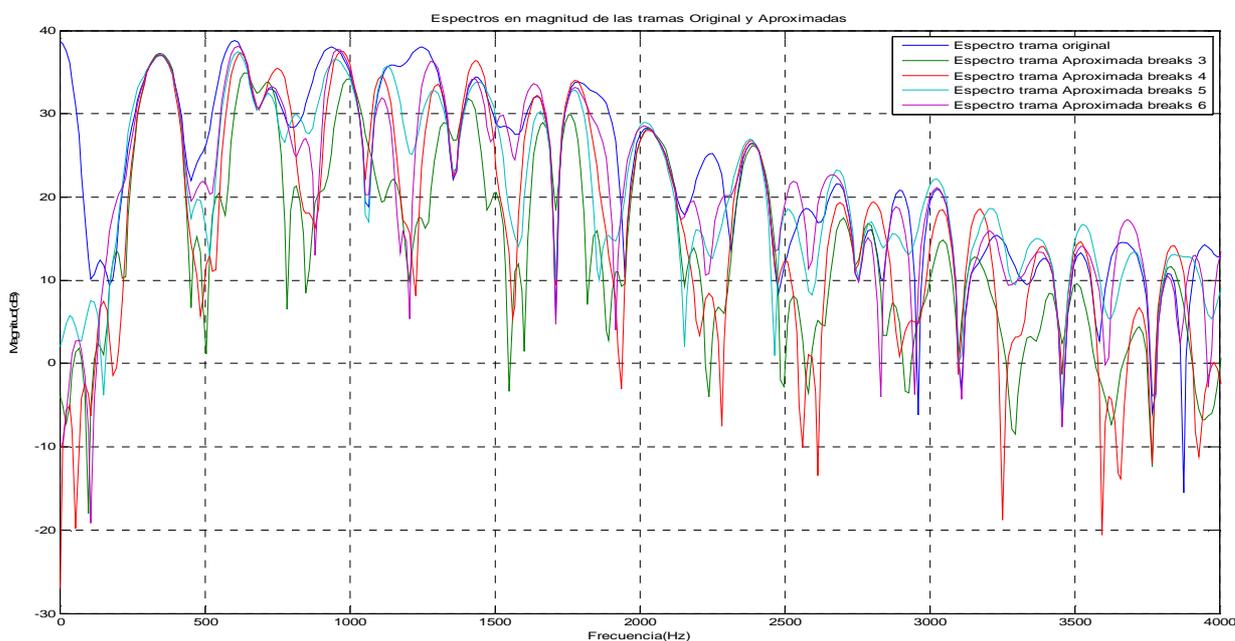
(Fig. 4.3.1.5.2 Espectros de la Trama 3 original y aproximada lineal por tramos, Breaks=7, Nharm=28)

No se elimina apenas ruido, se produce mucho Overfit.

A medida que subimos el valor de Breaks se va produciendo un **mayor Overfit hasta el caso límite de Breaks=8**, en el que se produce **Overfit Absoluto** y **no se elimina nada de ruido** al modelar. (No se representa el caso extremo por simplicidad)

4.3.1.6 – Conclusiones al modelado lineal por tramos de la Trama 3:

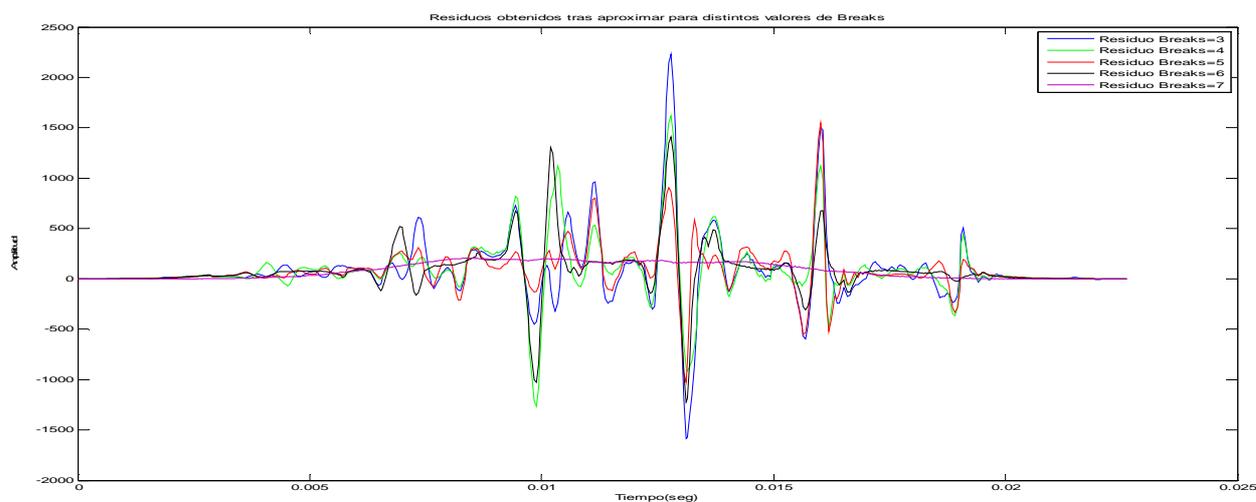
Representamos los espectros obtenidos para la trama modelada lineal por tramos con los distintos valores de Breaks junto al espectro de la trama original:



(Fig. 4.3.1.6.1 Espectros de la Trama 3 original y aproximadas lineal por tramos Breaks=3, 4, 5 y 6, Nharm=28)

Con valores bajos de Breaks ya hemos visto como se produce **mucho Subfit** y no se modela bien, este efecto es más claro observando el residuo en el dominio temporal. Para valores altos de Breaks se produce **Overfit** hasta el caso extremo de Breaks=8 en el que se modela la parte determinista como toda la señal y no se separa nada de ruido.

Representamos los residuos obtenidos en el dominio temporal:



(Fig. 4.3.1.6.2 Residuos de las aproximaciones, Breaks=3, 4, 5, 6 y 7, Nharm=28)

El mejor residuo obtenido es el que elimina algo de ruido sin llevarse consigo pulsos deterministas. Por tanto para **tramas transitorias** nos quedamos con el valor intermedio **Breaks=5**.

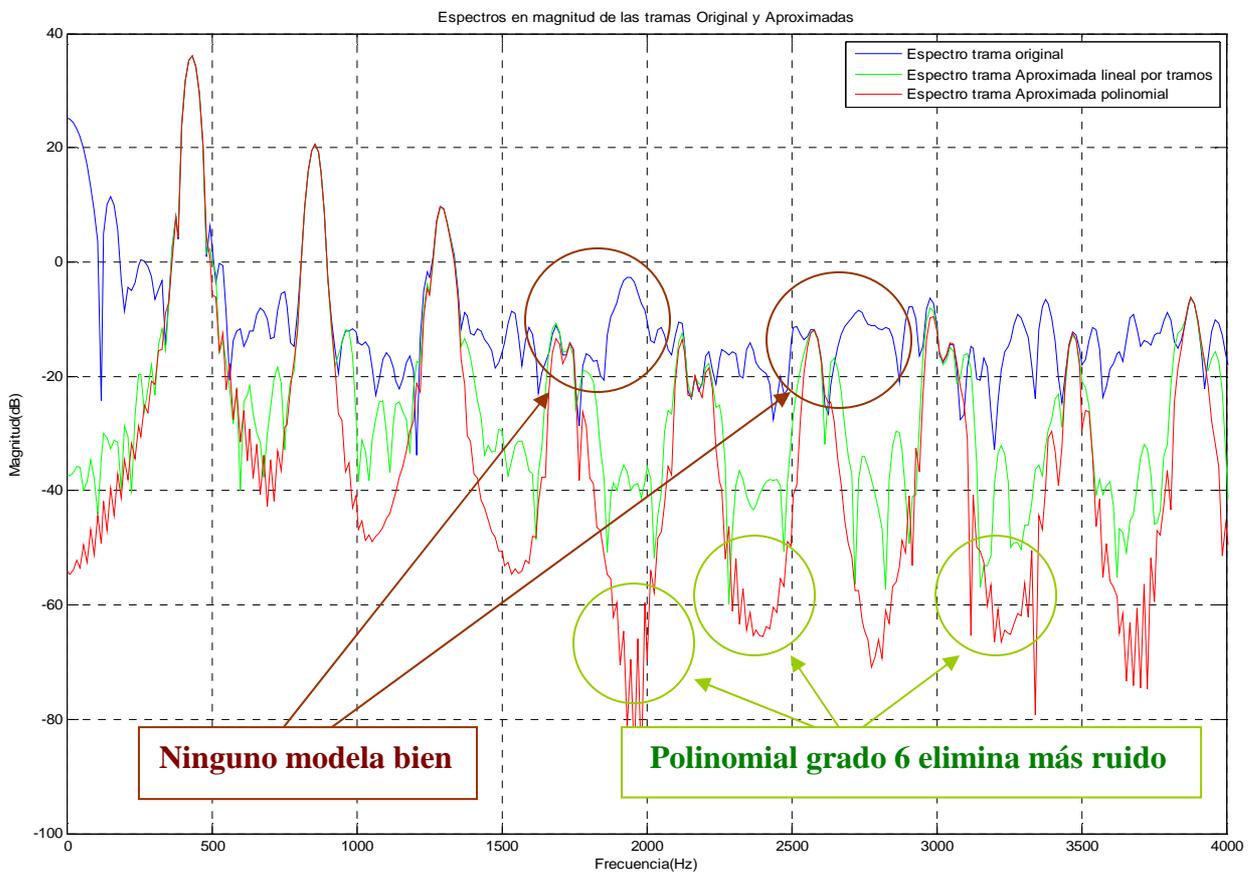
4.3.2 MODELADO LINEAL POR TRAMOS TRAMA 5:

Trama cuasi estacionaria ruidosa y con mucha modulación. Shakuhachi

Volvemos a estudiar la trama más complicada que podemos encontrarnos, la que tiene mucha modulación en el límite de lo transitorio y mucho ruido. Veremos que sucede si la modelamos mediante el algoritmo lineal por tramos y con todo lo estudiado concluiremos cual es el mejor modelado para este tipo de trama.

4.3.2.1 - Número de armónicos=15, Breaks=5.

Representamos los primeros parciales del espectro de la trama original junto al obtenido para la trama aproximada lineal por tramos y la aproximada polinomial de grado 6:



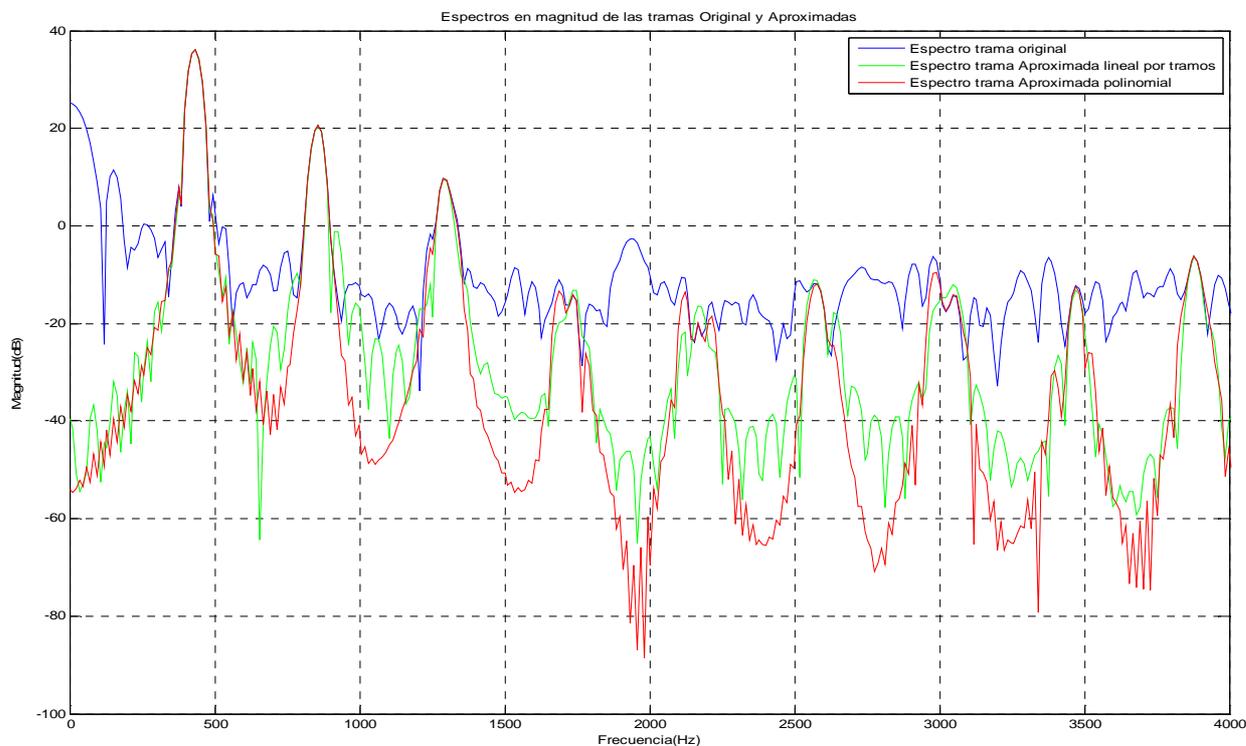
(Fig. 4.3.2.1.1 Espectros ampliados de la Trama 5 original, aproximada lineal por tramos y aproximada polinomial, Breaks=5, P=6, Nharm=15)

Se observa como la aproximación lineal por tramos sigue sin modelar correctamente los parciales que se encuentran en posiciones no armónicas. Además la aproximación lineal por tramos no mejora la aproximación polinomial de grado 6 en cuanto a eliminación de ruido. (Dentro de las aproximaciones polinomiales, la de grado 6 era la que menos ruido eliminaba)

Veremos que sucede si bajamos el valor de Breaks a 3, que eliminaría más ruido pero con riesgo de producir Subfit.

4.3.2.2 - Número de armónicos=15, Breaks=3.

Representamos los primeros parciales del espectro de la trama original junto al obtenido para la trama aproximada lineal por tramos y la aproximada polinomial de grado 6:



(Fig. 4.3.2.2.1 Espectros ampliados de la Trama 5 original, aproximada lineal por tramos y aproximada polinomial, Breaks=3, P=6, Nharm=15)

En este caso la **aproximación polinomial de grado 6 sigue modelando con menos ruido que la lineal por tramos y similar comportamiento en los parciales no armónicos.**

4.3.2.3 – Conclusiones al modelado lineal por tramos de la Trama 5:

Si obviamos la presencia de inarmonicidad en esta trama, concluimos que **para tramas con mucha modulación AM-FM y mucho ruido, lo adecuado será modelar con modelado polinomial de grado 6, ya que elimina más ruido que el modelado lineal por tramos y es el que menos Subfit produce dentro del modelado polinomial.**

TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3
Mucha modulación y poco ruido	Modelado no estacionario polinomial grado 6
Poca modulación y mucho ruido	Modelado no estacionario polinomial grado 2
Mucha modulación y mucho ruido	Modelado no estacionario polinomial grado 6

4.4 CONCLUSIONES AL ESTUDIO DEL OVERFIT EN EL MODELADO NO ESTACIONARIO

Recogemos en el siguiente cuadro las conclusiones respecto al valor óptimo de los parámetros y el modelo adecuado a aplicar para la extracción de la parte determinista del ruido a los distintos tipos de trama que nos podemos encontrar en secuencias audio reales:

TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3
Mucha modulación y poco ruido	Modelado no estacionario polinomial grado 6
Poca modulación y mucho ruido	Modelado no estacionario polinomial grado 2
Mucha modulación y mucho ruido	Modelado no estacionario polinomial grado 6
Transitoria	Modelado lineal por tramos Breaks= 5
Con fuerte inarmonicidad en sus parciales	Monofónica: Clasificación de picos
	Polifónica: Posible combinación entre Clasificación de picos y Modelado no estacionario polinomial grado 6

Se deja planteado el camino a seguir en el caso de encontrarnos con tramas en las que exista una fuerte inarmonicidad. **En la optimización de nuestro algoritmo de modelado no se profundizará en el estudio o separación de las tramas según su fuerte inarmonicidad ya que la probabilidad de encontrar este efecto es muy pequeña.**

5. Estudio y búsqueda de un método de baremo y determinación del grado de modulación de una trama

Este es el segundo bloque principal del presente proyecto, en el que nos centramos en investigar y crear un método o baremo de medida que nos indique el grado de modulación presente en una trama, con el fin de en función de dicho grado de modulación aplicar unos valores de los parámetros y tipo de modelado óptimos.

Estudiaremos **3 métodos diferentes, los implementaremos y los probaremos sobre las distintas tramas**, buscamos el método que nos determine el grado de modulación de una trama de la forma más fiable y eficaz. En todos ellos habrá que **definir o estimar unos umbrales** a partir de los cuales se establezca que grado se asigna a cada trama a través del método de medida, clasificándolas según 3 grados de modulación:

- Trama con poca modulación
- Trama con mucha modulación
- Trama transitoria

Los 3 métodos son muy novedosos, apenas se han desarrollado y únicamente el tercero ha sido publicado en un artículo de investigación. Son los siguientes:

- **Método 1:** Basado en el **Tiempo medio y la duración temporal**.
- **Método 2:** Basado en el **descriptor NBD** (Normalized Bandwidth Descriptor) usado en detección de picos.
- **Método 3:** Basado en el **Sparsity** (esparcimiento) **de la DWT** (Discrete Wavelet Transform) y **la MDCT** (Modified Discrete Cosine Transform)

5.1 MÉTODO 1: TIEMPO MEDIO Y DURACIÓN TEMPORAL

5.1.1 DESCRIPCIÓN

Este primer método está basado en dos parámetros utilizados en el estudio tiempo/frecuencia de señales moduladas, conocidos como **tiempo medio** y **duración temporal** que son indicadores de cómo se encuentra distribuida la energía en el dominio temporal [5].

El **tiempo medio** de una señal o trama se define como:

$$\langle t \rangle = \int t |s(t)|^2 dt \quad (\text{Ecuación 5.1.1.1})$$

Si trabajamos con señales muestreadas como es el caso (tramas) tendremos:

$$\langle n \rangle = \sum_{n=0}^{N-1} n |s(n)|^2 \quad (\text{Ecuación 5.1.1.2})$$

Es un **indicador de como se concentra la densidad de energía de la señal en el dominio temporal**.

Al trabajar con distintas tramas de diferentes secuencias/archivos musicales, es **conveniente normalizar** estos tiempos medios dividiendo entre la energía total de la señal, por tanto trabajaremos con:

$$\langle n \rangle = \frac{\sum_{n=0}^{N-1} n s(n)^2}{\sum_{n=0}^{N-1} s(n)^2} \quad (\text{Ecuación 5.1.1.3})$$

La **duración temporal** se define como:

$$T = \sqrt{\int (t - \langle t \rangle)^2 |s(t)|^2 dt} \quad (\text{Ecuación 5.1.1.4})$$

Es un parámetro que **se define como una desviación típica y nos indica como de concentrada en torno al tiempo medio se encuentra la energía** de la señal/trama en el dominio temporal.

Es así mismo un indicador de la duración de la señal, si la desviación típica o duración temporal es pequeña la mayoría de la señal estará concentrada en la media y la señal durará poco temporalmente, si la desviación o duración temporal es grande la energía estará dispersa en el tiempo y durará más temporalmente.

Trabajaremos con la versión normalizada para señales muestreadas:

$$T = \sqrt{\frac{\sum_{n=0}^{N-1} (n - \langle n \rangle)^2 |s(n)|^2}{\sum_{n=0}^{N-1} s(n)^2}}$$

(Ecuación 5.1.1.5)

5.1.2 GRADO DE MODULACIÓN SEGÚN LOS VALORES DEL TIEMPO MEDIO Y LA DURACIÓN TEMPORAL

1- En caso de que el valor obtenido para el **tiempo medio** $\langle n \rangle$ sea cercano al **valor medio de la trama** (el centro de la trama) podremos interpretarlo de dos maneras, según el valor obtenido para la **duración temporal T**:

- **Si T es grande**, se deduce que **la energía de la trama en el tiempo está muy distribuida desde el centro**, por tanto existe poca modulación AM o concentración de energía, se **tratará de una trama cuasi estacionaria**.
- **Si T es pequeño**, se deduce que **la energía de la trama en el tiempo está concentrada**, por tanto existe una modulación AM centrada en la trama con un ataque y un desvanecimiento, se **tratará de una trama transitoria** si supera un umbral a establecer.

2- En el caso de que el valor obtenido para el **tiempo medio** $\langle n \rangle$ esté alejado del **centro de la trama** se puede interpretar como:

- **Si el tiempo medio está hacia el final de la trama**, tenemos un **ataque**, la energía se concentra al final, estaríamos ante una **trama transitoria**.
- **Si el tiempo medio está al comienzo de la trama**, tenemos un **desvanecimiento**, la energía se concentra al comienzo y va decayendo, **trama transitoria**.

Los valores del tiempo medio y la duración temporal solo nos indican el **grado de modulación AM** de la trama. En señales reales la modulación AM y la FM van acopladas por lo que **determinando el grado de modulación AM podemos estimar el grado de modulación total**, además **la transitoriedad** en las señales reales siempre se debe a ataques, desvanecimientos y chasquidos asociados a una **fuerte AM**.

El problema que se plantea ahora es por tanto **establecer unos umbrales** a partir de los cuales, con los valores obtenidos para cada trama para el tiempo medio y la duración temporal, **poder asignar un grado de modulación a cada trama**, con el fin de clasificarla dentro de uno de los 3 tipos descritos anteriormente.

5.1.3 ESTIMACIÓN DE UNOS UMBRALES DE DECISIÓN

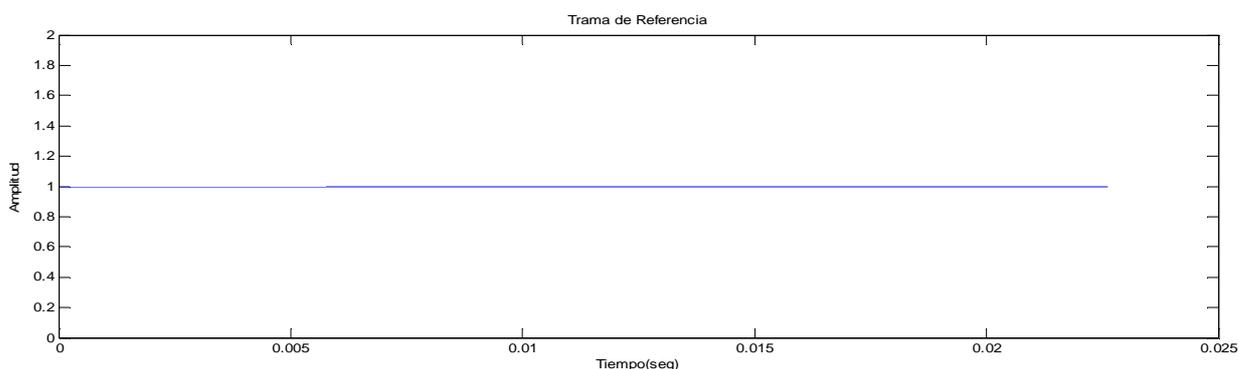
A continuación realizaremos un **estudio empírico sobre diferentes tramas** de los archivos ya estudiados anteriormente, calcularemos el tiempo medio y la duración temporal para cada una de ellas y asociaremos los valores obtenidos a las características de la trama.

Trataremos de **estimar unos umbrales con los que asignar un grado de modulación** a cada trama, clasificándolas en 3 tipos:

- **Trama con poca modulación**
- **Trama con mucha modulación**
- **Trama transitoria**

5.1.3.1 – Referencia 1. Ventana rectangular.

La trama con la que establecer una referencia sobre los valores del tiempo medio y la duración temporal será la ventana rectangular ya que por definición será la más centrada y la que tenga su energía en el dominio temporal mas distribuida. (**500 muestras, SR=22050**)



(Fig. 5.1.3.1 Ventana rectangular, 500 muestras)

Calculamos los valores del tiempo medio y duración temporal para esta trama:

- **Tiempo medio = 249.50**
- **Duración temporal= 144.33**

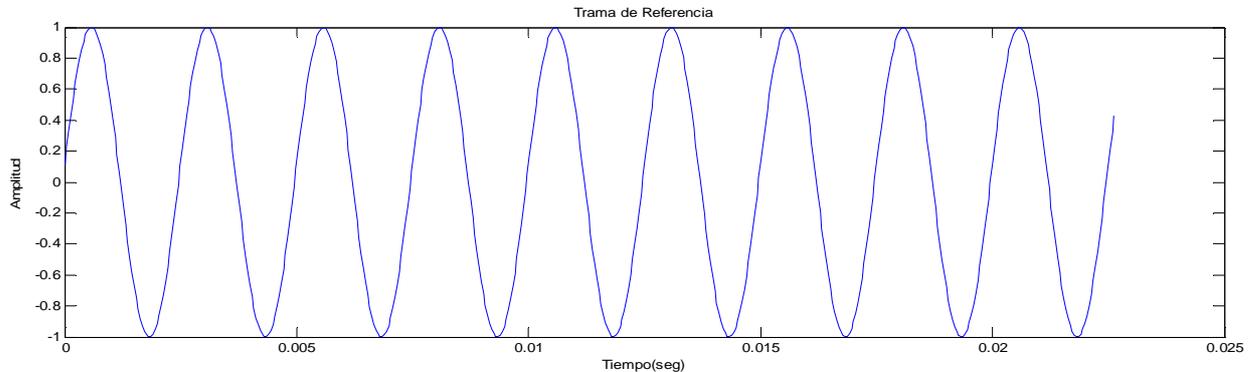
El centro de la trama es 250 por ser de duración 500 muestras, **el tiempo medio no es exactamente 250 por tener un número par de muestras**. Calculamos los valores relativos a la duración de la trama para el tiempo medio y la duración temporal, para que se puedan comparar con tramas de otros tamaños:

- **Tiempo medio relativo = 49.90%**
- **Duración temporal relativa= 28.86%**

La duración temporal de esta trama de referencia es grande, por tanto podemos hacer una **primera estimación de la duración temporal para una trama cuasi estacionaria en torno al 28% de la duración total de la trama**.

5.1.3.2 – Referencia 2. Tono puro.

Veremos que sucede si tomamos como segunda referencia un tono puro, que por definición tendrá así mismo su energía centrada y distribuida por toda la trama. ($f_0=400\text{Hz}$, $SR=22050$, 500 muestras)



(Fig. 5.1.3.2 Tono puro, 500 muestras, $f_0=400\text{Hz}$, $SR=22050$)

Calculamos los valores del tiempo medio y duración temporal para esta trama:

- **Tiempo medio = 247.41**
- **Duración temporal= 143.28**

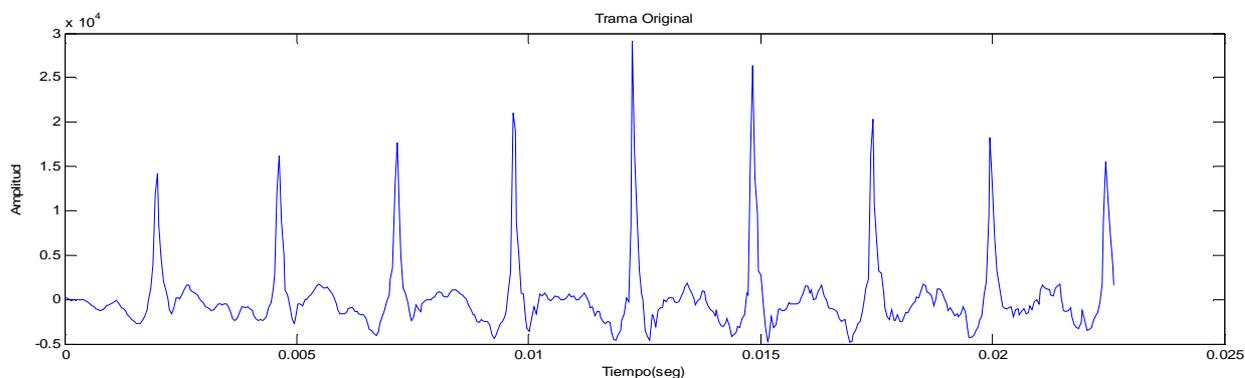
- **Tiempo medio relativo = 49.48%**
- **Duración temporal relativa= 28.65%**

Con los valores obtenidos para ambas tramas de referencia podemos hacer una **estimación inicial** de los valores que nos aseguren que una **trama es cuasi estacionaria**, si el **tiempo medio está dentro del 49% - 51%** y la **duración temporal en torno al 28%**.

A continuación estudiaremos tramas de secuencias musicales reales.

5.1.3.3 – TRAMA 2: cuasi estacionaria poco ruidosa con fuerte modulación. Trumpet

Trama con una modulación AM en el límite de lo cuasi estacionario. Calcularemos su tiempo medio y su duración temporal con el fin de asociarlos al límite de lo cuasi estacionario.



(Fig. 5.1.3.3 Trama 2, 500 muestras)

Calculamos los valores del tiempo medio y duración temporal para esta trama:

- **Tiempo medio = 261.70**
- **Duración temporal= 62.81**
- **Tiempo medio relativo = 52.34%**
- **Duración temporal relativa= 12.56%**

Como se puede deducir a simple vista **el tiempo medio de la trama está muy centrado, la energía está centrada.** Es por tanto el valor de la duración temporal el que nos indicará si la energía está concentrada o no, o lo que es lo mismo si existe una fuerte modulación AM o no.

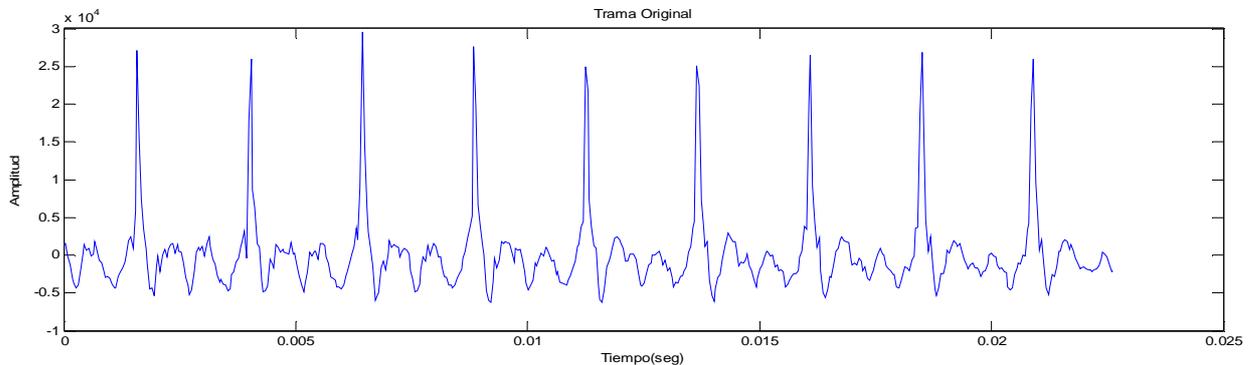
En este caso el valor obtenido para la **duración temporal es un 12.56%** de la duración de la trama, este valor es muy pequeño, **la energía está muy concentrada en el tiempo.**

Podemos establecer como **límite inferior de lo cuasi estacionario** el valor de **T=12.56%** pero hay que tener en cuenta que los parámetros tiempo medio y duración temporal **solo indican el nivel de modulación AM.** En esta trama concreta tenemos un **caso especial** en el que **predomina la modulación AM**, en el caso de que a esta modulación AM **se le sumara una fuerte modulación FM**, la trama pasaría de estar en el límite de lo cuasi estacionario a ser **transitoria.**

Los parámetros **tiempo medio y duración temporal no detectarían la modulación FM** y con el umbral establecido **daría como estacionaria una trama transitoria**, por tanto esta trama concreta la consideramos cuasi estacionaria en el límite por no tener FM, pero en caso de tenerla no podríamos detectarla con este sistema y **no podemos establecer su duración temporal como una referencia para el umbral.**

5.1.3.4 – TRAMA 1: cuasi estacionaria poco ruidosa con poca modulación. Trumpet

Veremos ahora que valores obtenemos en el caso de la primera trama estudiada en los apartados anteriores.



(Fig. 5.1.3.4 Trama 1, 500 muestras)

Calculamos los valores del tiempo medio y duración temporal para esta trama:

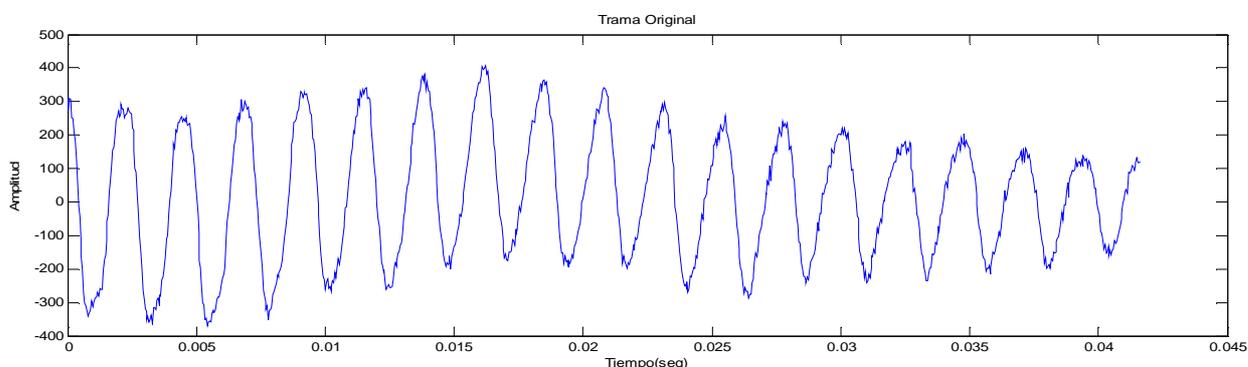
- **Tiempo medio = 249.38**
- **Duración temporal = 71.86**

- **Tiempo medio relativo = 49.87%**
- **Duración temporal relativa = 14.37%**

Si el valor del **tiempo medio** está **muy centrado (49% - 51%)** y la dispersión o **duración temporal T** está en torno al **14.37%** podríamos decir que **la trama es cuasi estacionaria con poca modulación**.

5.1.3.5 – TRAMA 5: cuasi estacionaria ruidosa y con mucha modulación. Shakuachi

Veremos que valores obtenemos para la trama 5, en la que teníamos una modulación conjunta AM-FM en el límite de lo cuasi estacionario y con mucho ruido.



(Fig. 5.1.3.5 Trama 5, 1000 muestras)

Calculamos los valores del tiempo medio y duración temporal para esta trama:

- **Tiempo medio = 470.44**
- **Duración temporal= 139.62**

- **Tiempo medio relativo = 47.04%**
- **Duración temporal relativa= 13.96%**

El valor del tiempo medio es del **47.04%** de la duración de la trama, por tanto **no podemos decir que se trate de una trama muy estacionaria** (dentro del 49% - 51%) pero tampoco podemos decir que se trate de una trama transitoria.

En esta trama **tenemos tanto modulación AM como FM**, por tanto podemos asignar el valor obtenido para la duración temporal en esta trama como **umbral de lo cuasi estacionario** por tener ambas modulaciones y ser el peor caso.

Para la duración temporal **asociamos el valor T=13.96% como umbral a partir del cual una trama se considere cuasi estacionaria.**

Establecemos así mismo un **límite inferior para el tiempo medio del 47%**, por **debajo de este tiempo medio** la energía está más al comienzo de la trama y se asocia a una modulación importante, **no estacionaria.**

Para el **límite superior** nos basamos en el valor obtenido en la trama 2, que tenía un **52.34% de tiempo medio**, establecemos que valores para el **tiempo medio por encima del 53 %** son considerados como correspondientes a **tramas transitorias.**

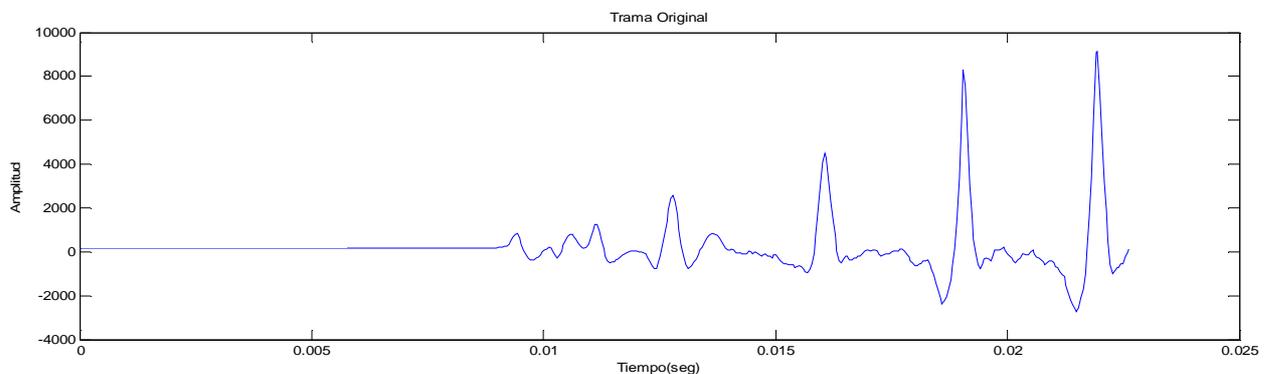
Por tanto establecemos las primeras estimaciones para los umbrales:

- Si el **tiempo medio** está fuera del 47% - 53 % la trama se considerará **transitoria**.
- Si el valor de la **duración temporal T** es inferior al valor umbral 13.96% la trama se considerará **transitoria**.
- Si el **tiempo medio** está en torno al 47% - 48% ó 52% - 53% y **T** es mayor que el 13.96%, se considerará que la trama tiene **mucha modulación**.
- Si el **tiempo medio** está dentro del 47% - 53 % y el valor de la duración temporal **T** esta entre el 13.96% - 15% consideraremos que la trama tiene **mucha modulación**.
- Si el **tiempo medio** se encuentra dentro del 49% - 51% y la duración temporal **T** es superior al 13.96%, tendremos que la trama tiene **poca modulación**.

Continuamos el estudio de más tramas para contrastar las primeras estimaciones.

5.1.3.6 – TRAMA 3: transitoria poco ruidosa. Trumpet

Establecidas estas estimaciones iniciales para los valores umbrales, veremos que sucede con la trama claramente transitoria que ya estudiamos anteriormente.



(Fig. 5.1.3.6 Trama 3, 500 muestras)

Calculamos los valores relativos del tiempo medio y la duración temporal para esta trama:

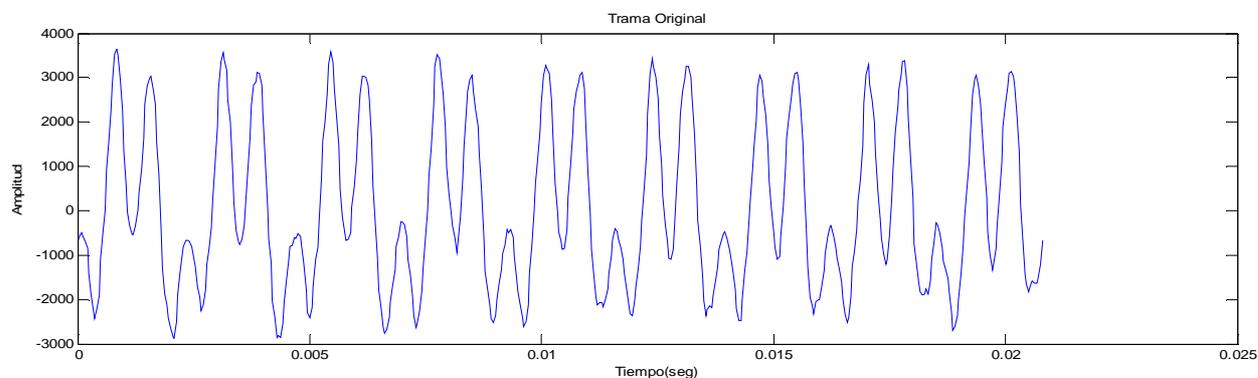
- **Tiempo medio relativo = 85.76%**
- **Duración temporal relativa = 13.57%**

Obtenemos un **tiempo medio alejadísimo del centro** de la trama y de los umbrales (47% - 53%) planteados anteriormente, es claramente **un ataque**.

Para la duración temporal tenemos un valor inferior al umbral estimado del 13.96%, como se deducía a simple vista es una **trama transitoria**.

5.1.3.7 – TRAMA 6: elegida al azar. Shakuhachi

Tomamos ahora una trama nueva, correspondiente al archivo/secuencia audio ‘Shakuhachi.aiff’. Concretamente la trama 2000 con ventana de **500 muestras**, muestras 1000001:1000500 (tras diezmar) correspondiente al minuto 0:416 de la pista original.



(Fig. 5.1.3.7 Trama 6, 500 muestras)

Observamos que se trata de una trama bastante estacionaria, veremos que valores obtenemos para el tiempo medio y la duración temporal y si asignamos correctamente el grado de modulación según los umbrales estimados anteriormente.

- **Tiempo medio relativo = 48.77%**
- **Duración temporal relativa= 28.69%**

Obtenemos un valor para el tiempo medio dentro del umbral **47% - 53%** y para la duración temporal tenemos un valor muy superior al umbral del **13.96%**.

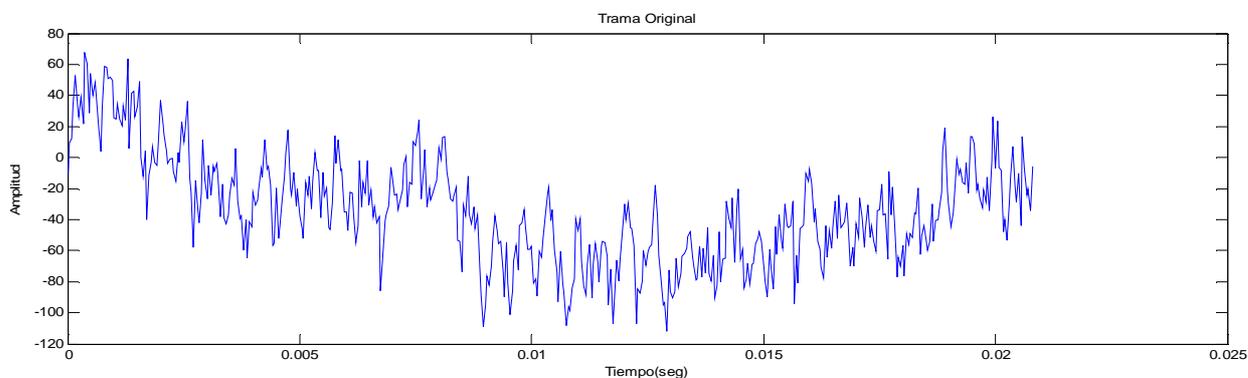
Aunque el valor obtenido para el tiempo medio no está dentro de los límites estimados anteriormente del **49%-51%** vemos que en este caso está muy cerca (**48.77%**) lo cual asociado a la gran duración temporal (**T del 28.69%**) hace que se pueda considerar la trama con **poca modulación.**

Por tanto **replanteamos los umbrales** iniciales añadiendo:

- Si el **tiempo medio** está dentro del **47% - 53%** y la duración temporal **T** es mayor que el **28%**, consideraremos que la trama tiene **poca modulación.**

5.1.3.8 – TRAMA 7: elegida al azar. Shakuhachi

Tomamos otra trama nueva, correspondiente al archivo 'Shakuhachi.aiff'. Trama 1008 con ventana de **500 muestras**, muestras 504001:504500 (tras diezmar) correspondiente al minuto 0:21 de la pista original.



(Fig. 5.1.3.8 Trama 7, 500 muestras)

Se trata de una trama totalmente ruidosa, veremos que nos dice nuestro método con los umbrales estimados:

- **Tiempo medio relativo = 56.15%**
- **Duración temporal relativa= 21.43%**

El tiempo medio está fuera de los umbrales estimados de cuasi estacionariedad (**47% - 53%**) por tanto se deduce que es una **trama transitoria**.

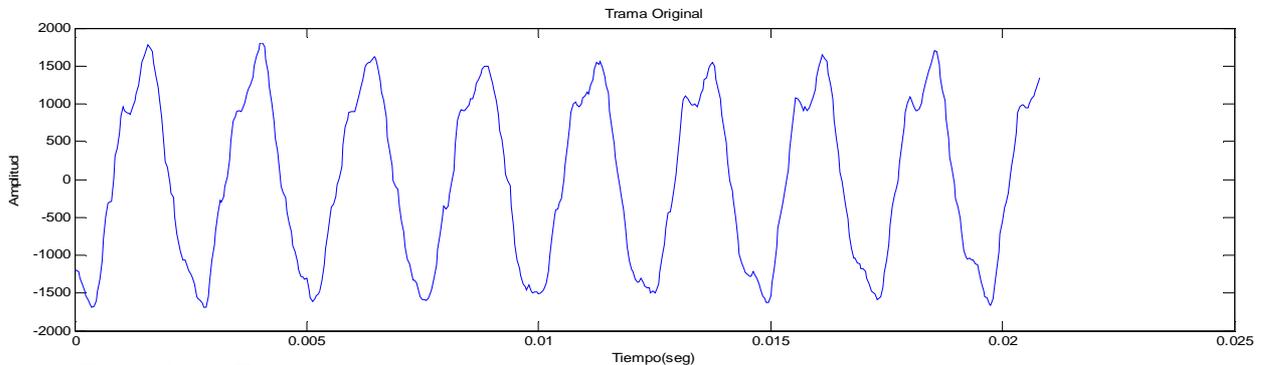
Pero esta trama realmente **no es una trama transitoria**, ya que **no tiene componente determinista** apreciable, se tiene solamente ruido, es una **trama ruidosa pura**.

Este tipo de **trama en el que solo hay ruido no se modelan**, al no tener componente determinista no merece la pena procesarlas, toda la trama tal cual iría a formar parte del residuo de la secuencia procesada. Existen métodos empleados en **procesado de voz** que estudian la energía de la trama para saber cuando una trama es exclusivamente ruidosa ó si existe componente determinista, también conocidas como **tramas sordas** (solo ruido) y **tramas sonoras** (con parte determinista).

Antes de aplicar el modelado a la secuencia total, aplicaremos estos criterios para separar las tramas que sean exclusivamente ruidosas (sordas) y llevarlas directamente al residuo sin necesidad de procesarlas.

5.1.3.9 – TRAMA 8: elegida al azar. Shakuhachi

Tomamos ahora otra trama del archivo ‘Shakuhachi.aiff’. Concretamente la **trama 2400 con ventana de 500 muestras, muestras 1200001:1200500 (tras diezmar)** correspondiente al minuto 0:25 de la pista original.



(Fig. 5.1.3.9 Trama 8, 500 muestras)

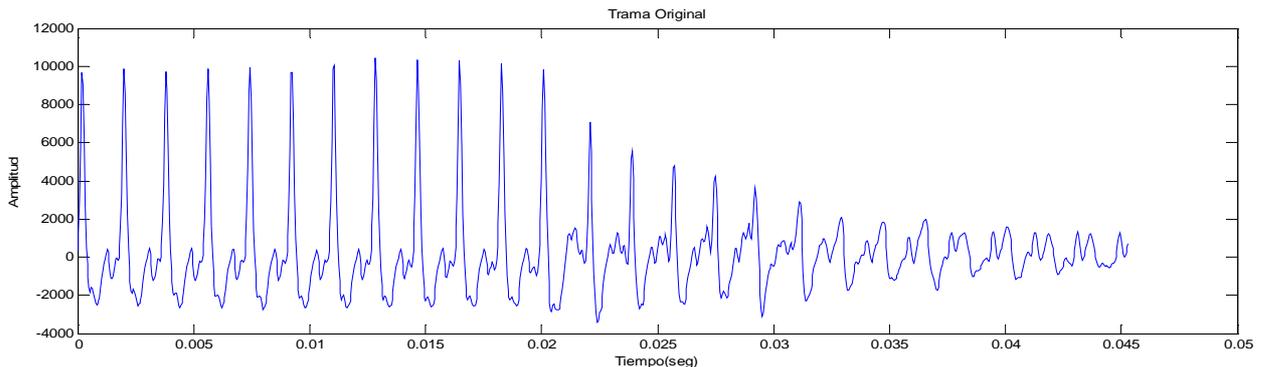
Observamos que se trata de una trama ligeramente modulada, veremos que nos dice nuestro sistema de medida:

- **Tiempo medio relativo = 48.66%**
- **Duración temporal relativa= 29.14%**

El tiempo medio está dentro del umbral del **47% - 53%** y la duración temporal por encima del **28%**, por tanto se trata de una trama con **poca modulación**.

5.1.3.10 – TRAMA 9: desvanecimiento. Trumpet

Tomamos ahora una nueva trama del archivo audio ‘trumpet4.aiff’. Concretamente la **trama 66 con ventana de 1000 muestras, muestras 66001:67000 (tras diezmar)** correspondiente a un desvanecimiento.



(Fig. 5.1.3.10 Trama 9, 1000 muestras)

Es una trama con mucha modulación. Veremos que nos dice nuestro método:

- **Tiempo medio relativo = 30.40%**
- **Duración temporal relativa= 20.46%**.

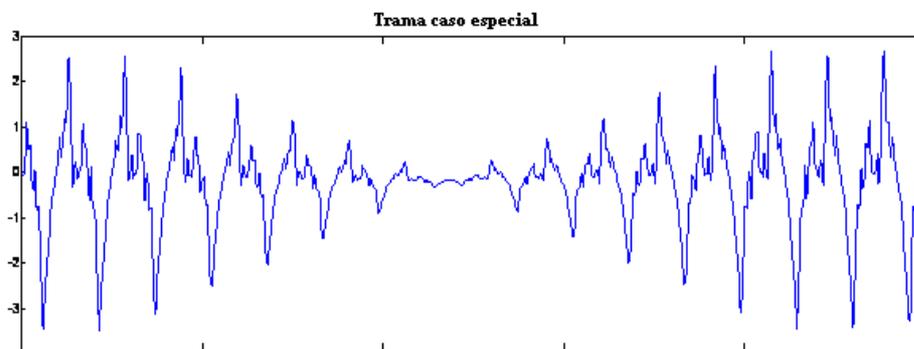
El tiempo medio está alejadísimo del centro de la trama y de los umbrales estimados, tenemos un **desvanecimiento** (la energía está concentrada al comienzo de la trama), estamos ante una **trama transitoria**.

5.1.4 CONCLUSIONES. UMBRALES DE DECISIÓN ESTIMADOS

Concluimos dando por válidas las estimaciones planteadas en el estudio:

- Si el **tiempo medio** está fuera del 47% - 53 % se considerará la **trama transitoria**.
- Si el valor de la **duración temporal T** es inferior al valor umbral 13.96% se considerará la **trama transitoria**.
- Si el **tiempo medio** está en torno al 47% - 48% ó 52% - 53% y **T** es mayor que el 13.96%, se considerará la **trama con mucha modulación**.
- Si el **tiempo medio** está dentro del 47% - 53 % y el valor de la duración temporal **T** esta entre el 13.96% - 15% se considerará la **trama con mucha modulación**.
- Si el **tiempo medio** se encuentra dentro del 49% - 51% y la duración temporal **T** es superior al 13.96%, se considerará la **trama con poca modulación**.
- Si el **tiempo medio** está dentro del 47% - 53% y la duración temporal **T** es mayor que el 28%, se considerará la **trama con poca modulación**.

Hay que tener en cuenta que este sistema de determinación del grado de modulación en una trama basado en los valores que toman el tiempo medio y la duración temporal, no funciona bien para el caso especial en el que dentro de una misma trama se produzca un desvanecimiento seguido de un ataque. En este caso el tiempo medio estaría muy centrado, la duración temporal sería grande y determinaríamos que la trama tiene poca modulación, cuando en realidad podría ser una trama transitoria. **La probabilidad de encontrar una trama de este tipo con ventanas de 500 o 1000 muestras en señales reales es mínima.**



(Fig. 5.1.4 Trama especial en la que el método no funcionaría, poco probable con ventanas pequeñas)

5.2 MÉTODO 2: DESCRIPTOR NBD USADO EN CLASIFICACIÓN DE PICOS

En este segundo método investigamos la posibilidad de utilizar la información proporcionada por un descriptor empleado en los algoritmos de clasificación de picos espectrales para determinar el grado de modulación de una trama. Es un **método nuevo, no estudiado anteriormente**.

5.2.1 DESCRIPCIÓN

Emplearemos el descriptor **NBD** (Normalized Bandwidth Descriptor), definido como:

$$NBD = \frac{BW_{rms}}{L} = \frac{1}{L} \sqrt{\frac{\sum_k (k - \bar{k})^2 |X(k)|^2}{\sum_k |X(k)|^2}} \quad (\text{Ecuación 5.2.1.1})$$

$$\bar{k} = \frac{\sum_k k |X(k)|^2}{\sum_k |X(k)|^2} \quad (\text{Ecuación 5.2.1.2})$$

donde L es el número de muestras del pico detectado en el espectro, X(k) es la DFT de la trama enventanada, BW_{rms} es el ancho de banda cuadrático medio del pico y \bar{k} la frecuencia media. El descriptor **NBD es un cociente entre el ancho de banda del pico y su anchura en bins** [4].

Se compara la distribución de energía del pico con la distribución de energía en todo el espectro para decidir si tiene mayor o menor probabilidad de ser ruidoso ó determinista.

Aplicaremos a cada trama el algoritmo ya implementado de clasificación de picos que emplea el descriptor NBD para asignar a cada pico un grado ó valor asociado de probabilidad de que dicho pico sea o no determinista⁵.

Según los valores que tome el descriptor NBD para los picos deterministas en cada trama a estudiar, se buscarán relaciones o patrones similares entre tramas de iguales características en cuanto a modulación con el fin de si existieran, poder establecer una relación entre los valores que va tomando el descriptor NBD y el grado de modulación de la trama.

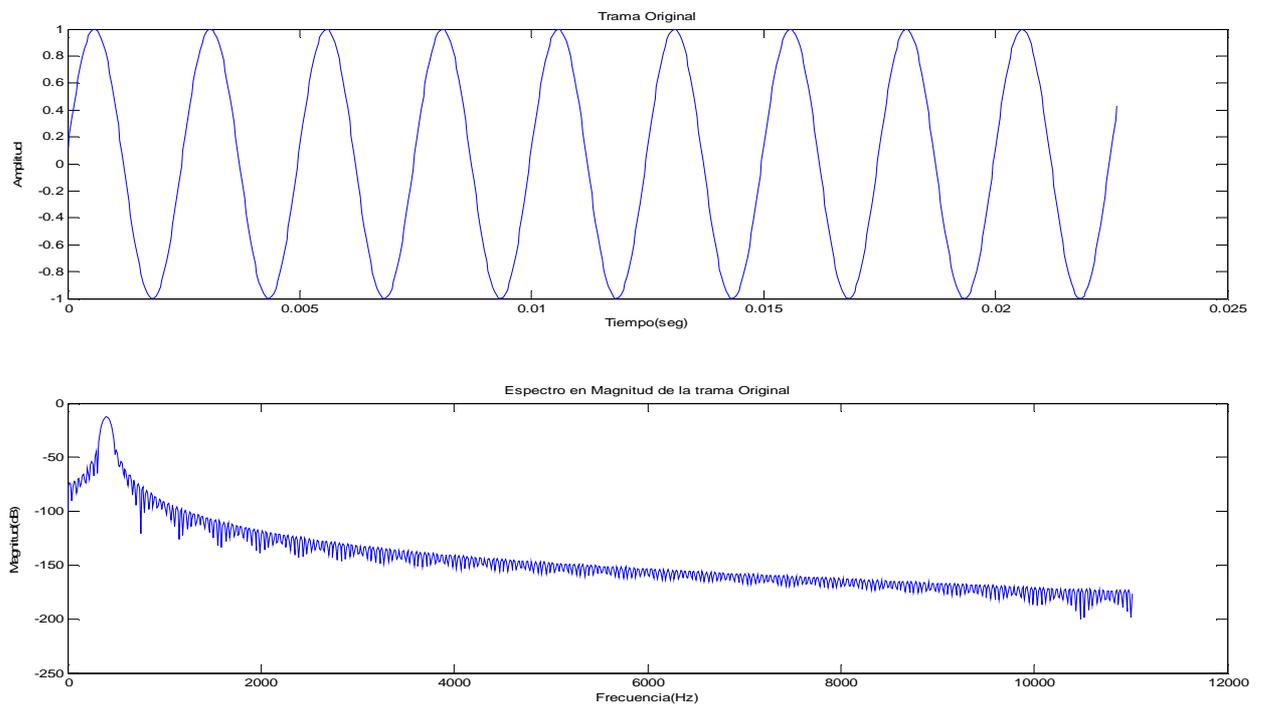
⁵ El algoritmo se encuentra implementado en la función “**descriptors.m**” que devuelve un **vector B** con los valores del descriptor NBD para todos los picos detectados.

5.2.2 ESTUDIO DEL GRADO DE MODULACIÓN SEGÚN LOS VALORES NBD

Procedemos al estudio de la información proporcionada por el descriptor NBD aplicando clasificación de picos a algunas de las tramas estudiadas anteriormente.

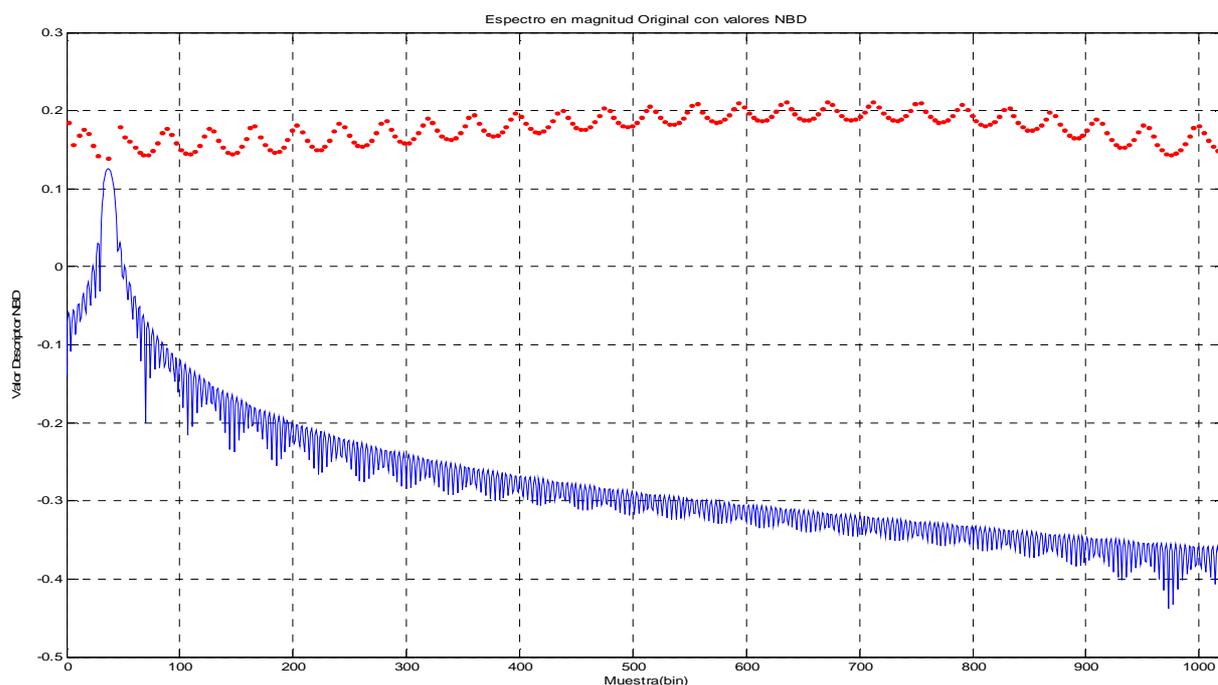
5.2.2.1 – Referencia. Tono puro.

Veremos que sucede si tomamos como referencia un tono puro, que por definición tendrá **un único parcial o pico determinista**, veremos que valor se obtiene para el descriptor. ($f_0=400\text{Hz}$, $\text{SR}=22050$ y 500 muestras)



(Fig. 5.2.2.1.1 Tono puro, 500 muestras, $f_0=400\text{Hz}$, $\text{SR}=22050$)

Aplicamos el algoritmo de clasificación de picos y asignación del valor del descriptor NBD para cada pico detectado, obtenemos la siguiente representación:



(Fig. 5.2.2.1.2 Valores del descriptor NBD sobre el espectro de la Trama Tono puro)

Obtenemos un vector con los valores del descriptor NBD para 245 picos detectados. Como ya sabemos **solo uno de ellos es determinista, el resto son lóbulos laterales por efecto del enventanado** visibles por no existir ruido.

El vector obtenido tiene las siguientes características estadísticas:

- Número de picos detectados: 245
- Valor máximo: 0.2112
- Valor mínimo: 0.1387
- Valor medio: 0.1779
- Varianza: 3.1501e-004

El valor mínimo del descriptor se obtiene para el lóbulo principal, es decir, para el parcial a 400 Hz, a menor valor del descriptor más determinista es el pico.

Nos interesará estudiar las características del vector con los valores obtenidos para el descriptor en los picos que sabemos que son deterministas, en este caso tendríamos un único valor, el correspondiente al pico en 400Hz, con NBD=0.138.

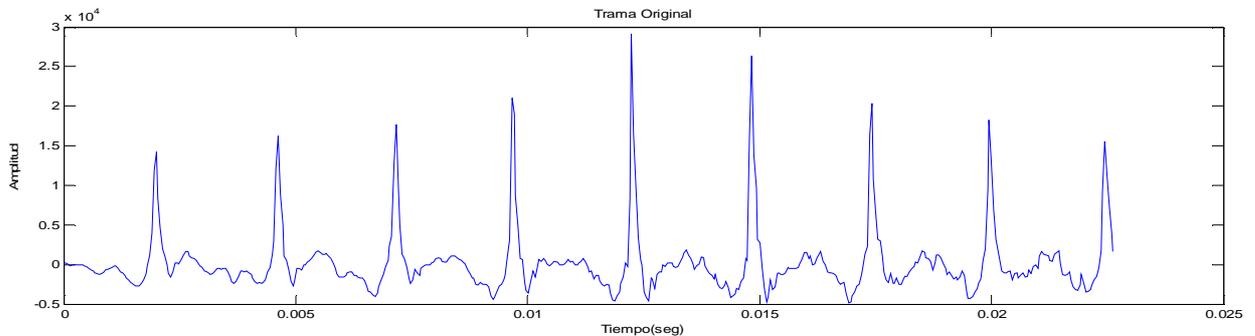
A continuación estudiaremos tramas de señales musicales reales.

“Solo podremos contrastar los resultados que se obtengan para los distintos tipos de trama entre las tramas de un mismo instrumento, ya que cada instrumento posee una envolvente espectral diferente según su timbre y sería una variable adicional al estudio”

Estudiaremos primeramente tramas del archivo en el que suena una Trompeta.

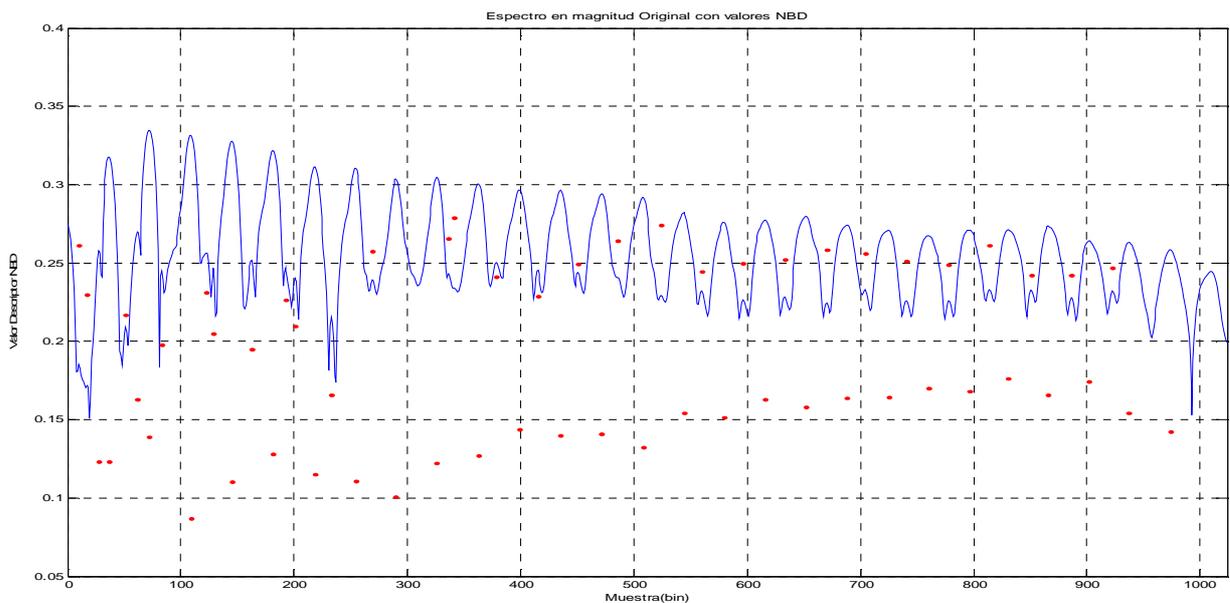
5.2.2.2 – TRAMA 2: cuasi estacionaria poco ruidosa con fuerte modulación. Trumpet

Trama con una modulación AM en el límite de lo cuasi estacionario.



(Fig. 5.2.2.2.1 Trama 2, 500 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor del descriptor NBD para cada pico detectado:



(Fig. 5.2.2.2.2 Valores del descriptor NBD sobre el espectro de la Trama 2)

Obtenemos un vector “**B**” con los valores del descriptor NBD para 58 picos detectados.

El vector obtenido tiene las siguientes características estadísticas:

- Número de picos detectados: 58
- Valor máximo: 0.2791
- Valor mínimo: 0.0870
- Valor medio: 0.1908
- Varianza: 0.0030

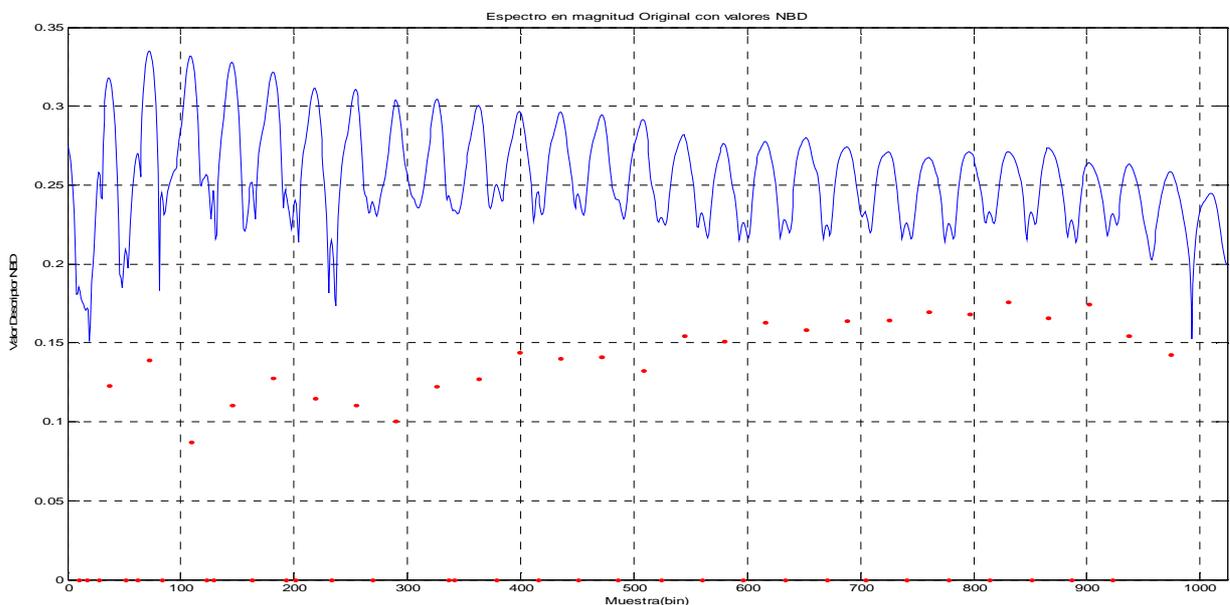
Los picos detectados toman valores muy diferentes para el descriptor, existe una gran varianza. Esto se debe a que los parciales están bastante diferenciados de los picos ruidosos ya que éstos últimos son pequeños, y los valores asignados para el descriptor NBD son muy diferentes. A los **picos deterministas** se les asignan los **valores más bajos** del descriptor.

Nos interesa estudiar **como se comportan los valores obtenidos para el descriptor NBD en los picos deterministas**, para ello debemos quedarnos en el vector obtenido con los valores que corresponden a parciales.

Mediante observación directa de la gráfica, nos quedaremos con los siguientes valores del vector:

B(4,7,9,12,14,17,19,21,22,25,27,29,31,33,35,37,39,41,43,45,47,49,51,53,55,57,58)

Obtenemos un nuevo vector con los valores del descriptor para los picos deterministas y con ceros en los picos ruidosos.



(Fig. 5.2.2.2.3 Valores del NBD para los parciales sobre el espectro de la Trama 2)

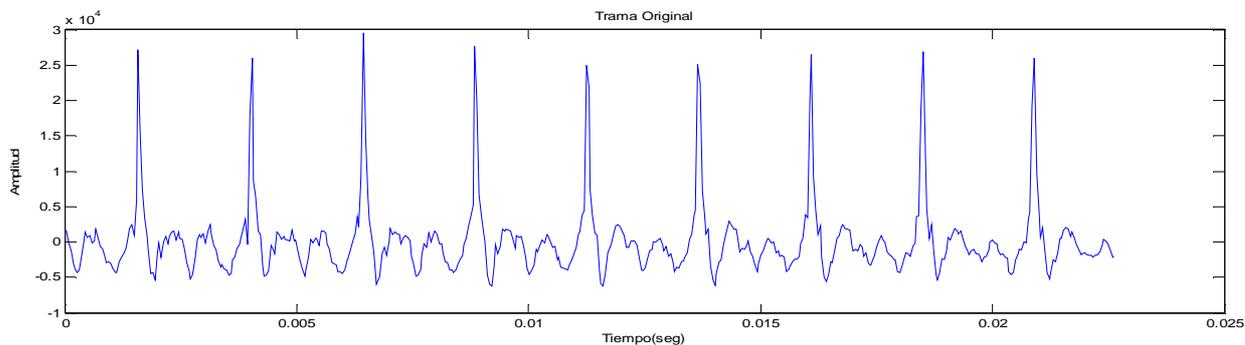
Calculamos los estadísticos del vector formado por los valores del descriptor NBD para los picos deterministas:

- Número de parciales detectados: **27**
- Valor máximo: **0.1761**
- Valor mínimo: **0.0870**
- Valor medio: **0.1418**
- Varianza: **5.7810e-004**

Para esta **trama con mucha modulación** y con poco ruido se observa como **los valores del descriptor para los parciales suben a medida que avanzamos en el espectro**, se observa una **tendencia creciente**, que se **podría asociar al alto grado de modulación**. La modulación FM es más pronunciada a altas frecuencias y de ahí se deriva que los parciales se vayan ensanchando al subir en frecuencia y se obtengan valores del descriptor mayores (los interpreta como menos deterministas).

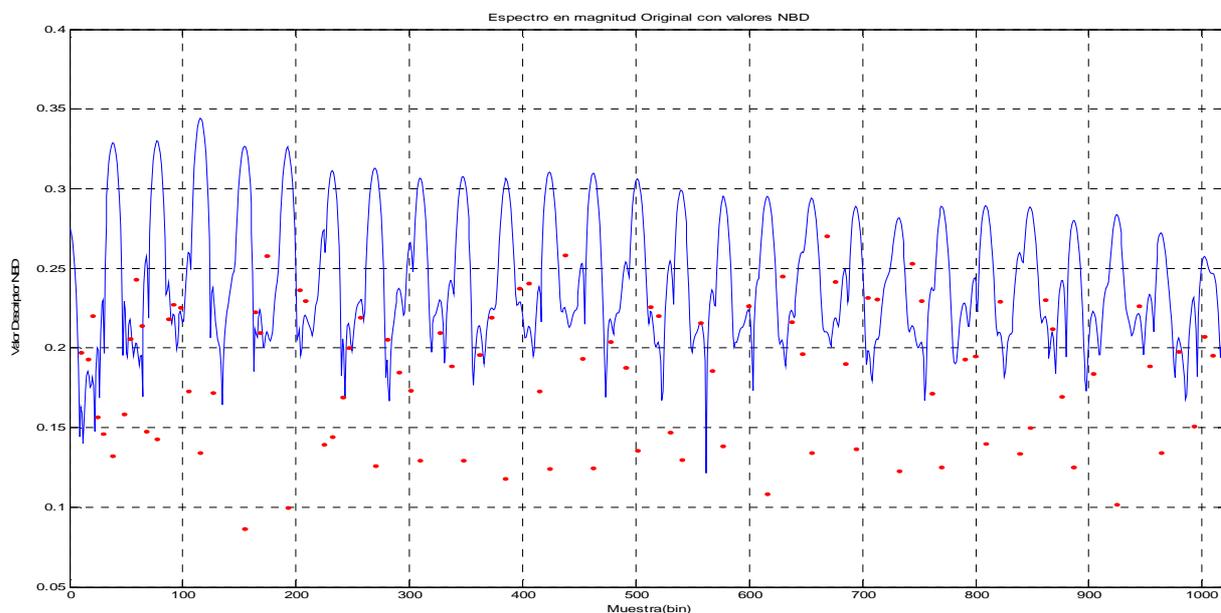
5.2.2.3 – TRAMA 1: cuasi estacionaria poco ruidosa con poca modulación. Trumpet

Trama cuasi estacionaria poco ruidosa y con poca modulación.



(Fig. 5.2.2.3.1 Trama 1, 500 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:

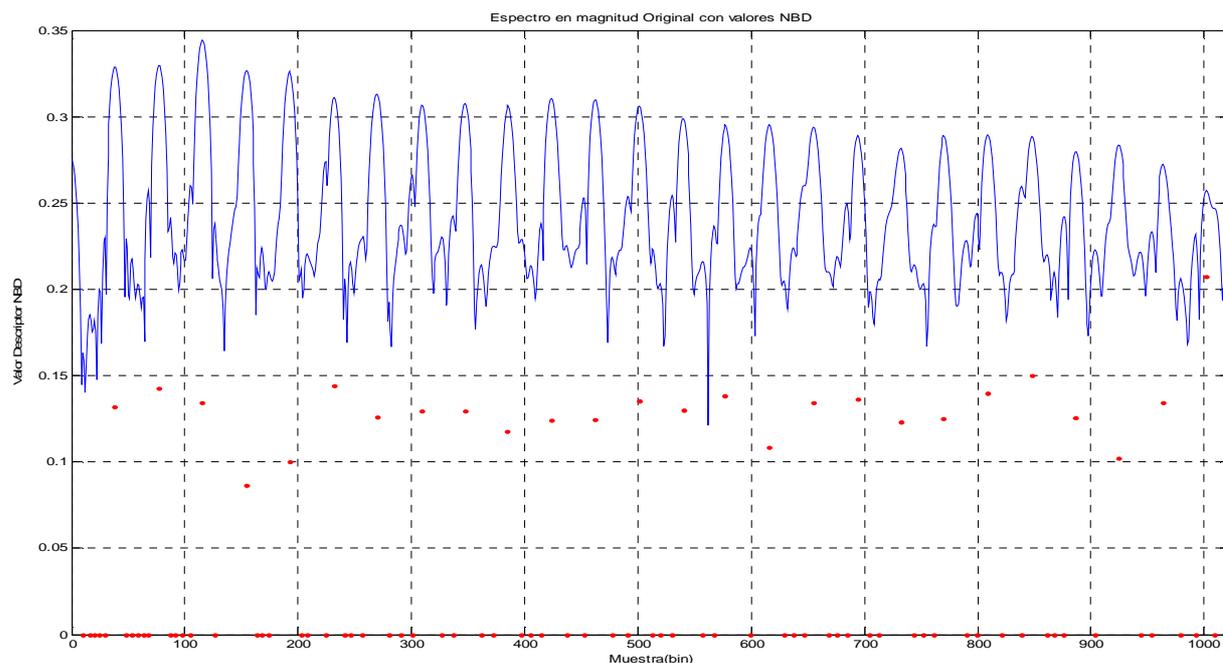


(Fig. 5.2.2.3.2 Valores del descriptor NBD sobre el espectro de la Trama 1)

Se detectan 94 picos. Realizamos el filtrado de manera manual, quedándonos con los valores correspondientes a picos que estimamos deterministas:

B(6,12,17,19,23,27,31,35,38,41,45,48,51,55,58,60,64,68,71,75,78,81,85,87,90,93)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.3.3 Valores del NBD para los parciales sobre el espectro de la Trama 1)

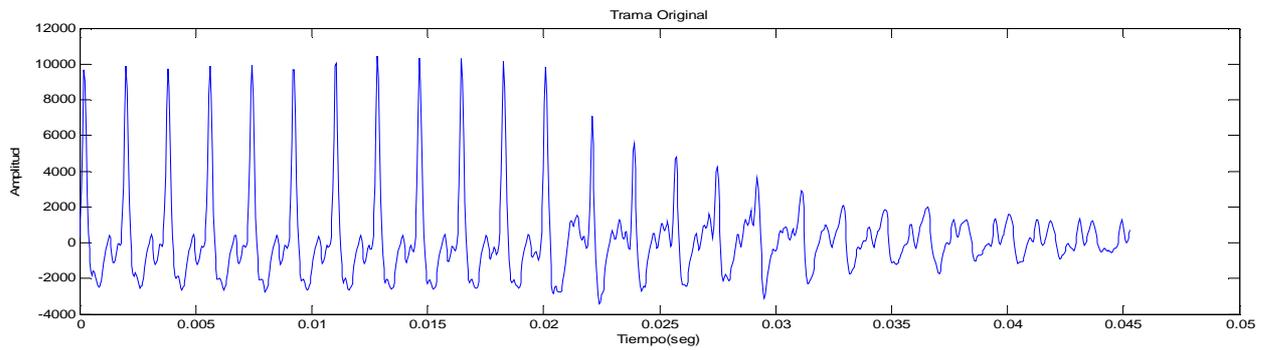
Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

- Número de parciales detectados: **26**
- Valor máximo: **0.2074**
- Valor mínimo: **0.0863**
- Valor medio: **0.1301**
- Varianza: **4.5732e-004**

En este caso para una **trama con poca modulación** y poco ruido, se observa como el vector de **los valores del descriptor NBD para los picos deterministas tiene una tendencia plana**, ya no se observa una tendencia creciente al subir en frecuencia como en la trama con mucha modulación perteneciente al mismo instrumento (misma envolvente). Se tiene que **todos los parciales están igualmente ensanchados**, ya que al no existir apenas modulación ésta afecta por igual a todas las frecuencias y no se pronuncia más en los picos altos como en el caso anterior.

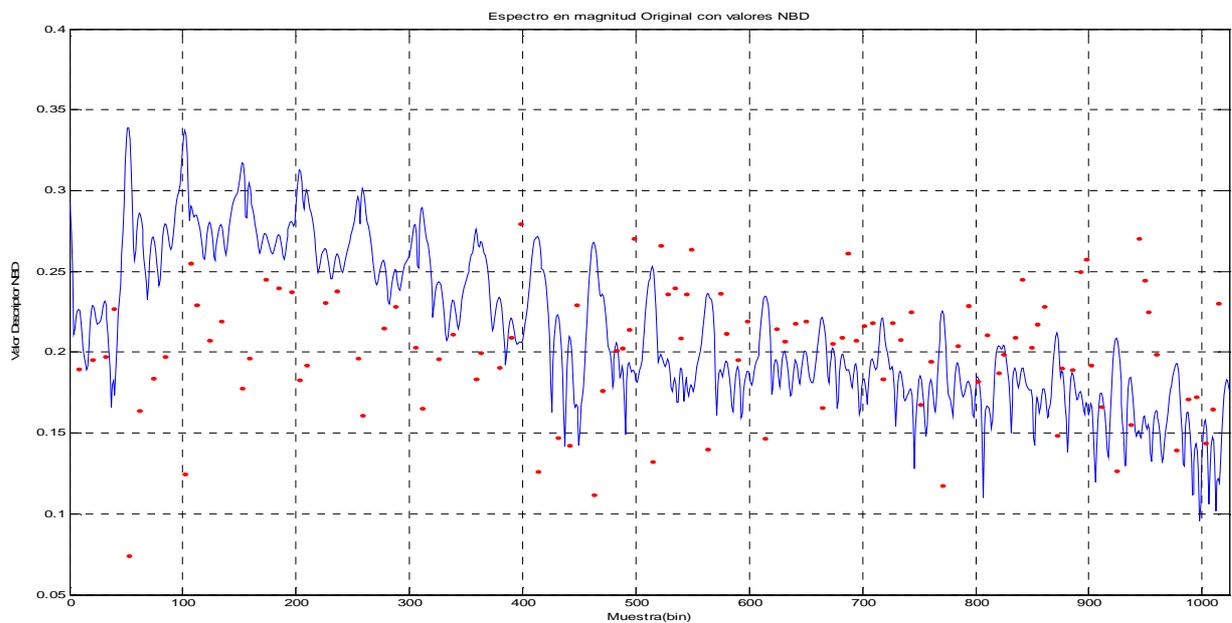
5.2.2.4 – TRAMA 9: desvanecimiento. Trumpet

Trama con mucha modulación y ruido.



(Fig. 5.2.2.4.1 Trama 9, 1000 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:

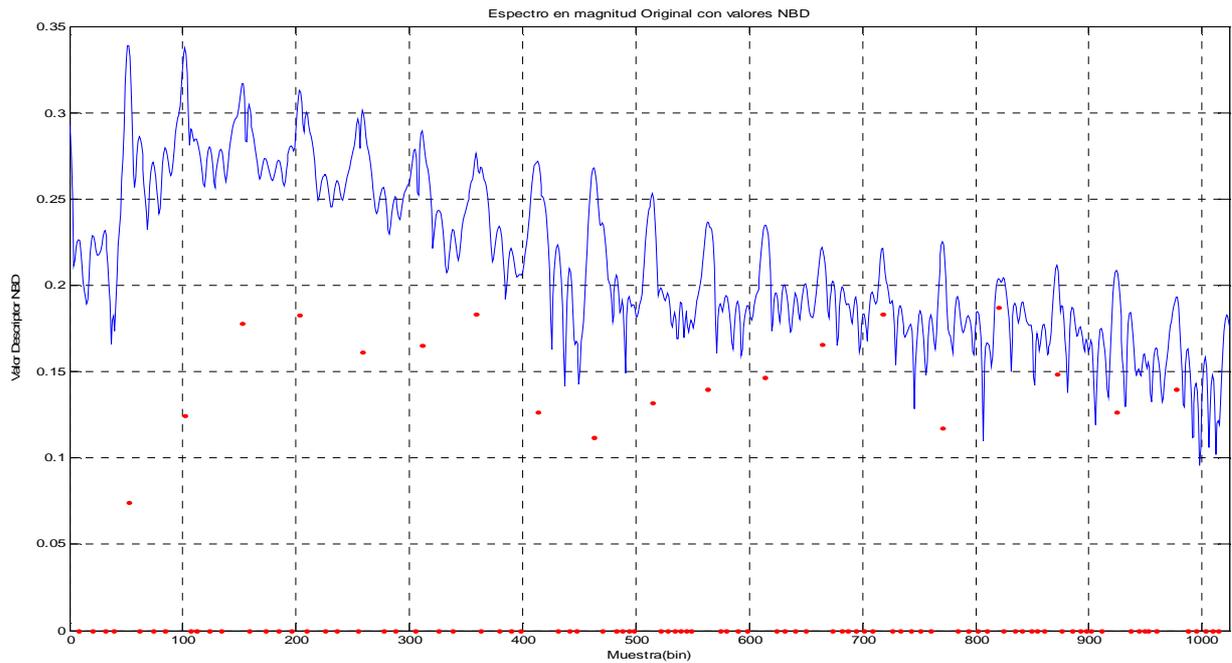


(Fig. 5.2.2.4.2 Valores del descriptor NBD sobre el espectro de la Trama 9)

Se detectan 106 picos. Realizamos el filtrado, quedándonos con los valores correspondientes a los picos que estimamos deterministas:

B(5,9,14,19,24,28,31,36,40,46,53,58,63,70,76,81,88,95,101)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.4.3 Valores del NBD para los parciales sobre el espectro de la Trama 9)

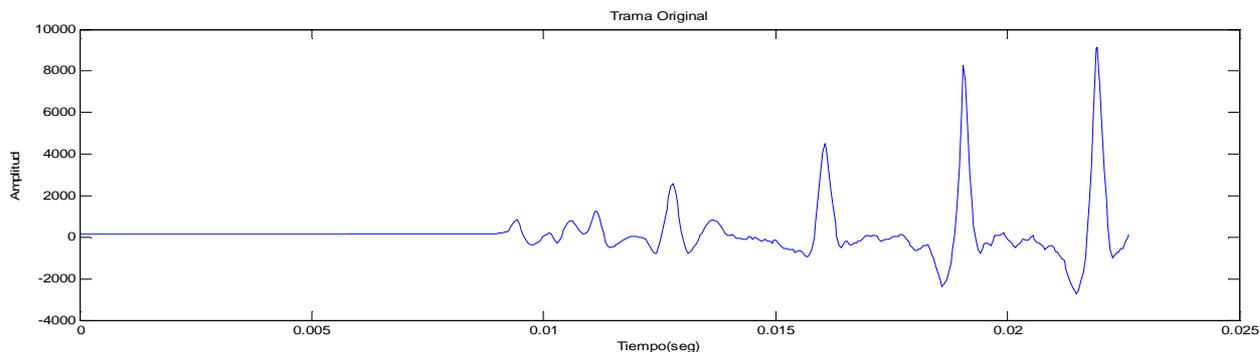
Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

- Número de parciales detectados: **19**
- Valor máximo: **0.1874**
- Valor mínimo: **0.0742**
- Valor medio: **0.1471**
- Varianza: **9.1385e-004**

Para esta **trama con mucha modulación** y poco ruido, se observa en el vector formado por los **valores del descriptor para los picos deterministas una tendencia creciente alta**, que se puede asociar al **alto grado de modulación** presente en la trama.

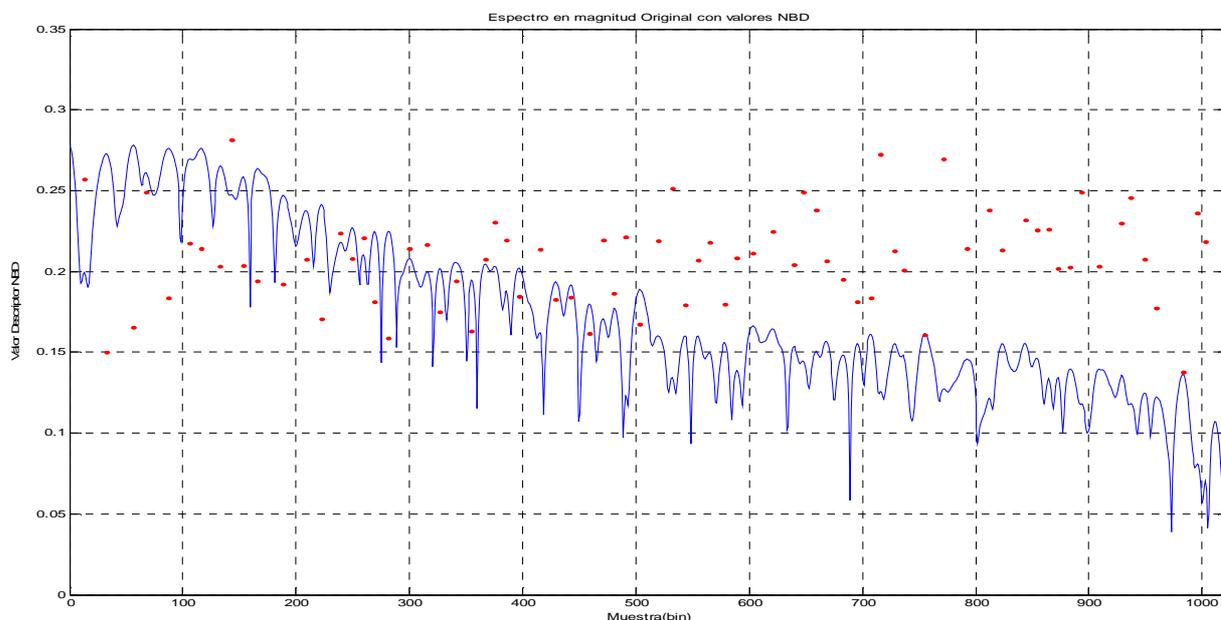
5.2.2.5 – TRAMA 3: transitoria poco ruidosa. Trumpet

Veremos que sucede con la trama transitoria.



(Fig. 5.2.2.5.1 Trama 3, 500 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:



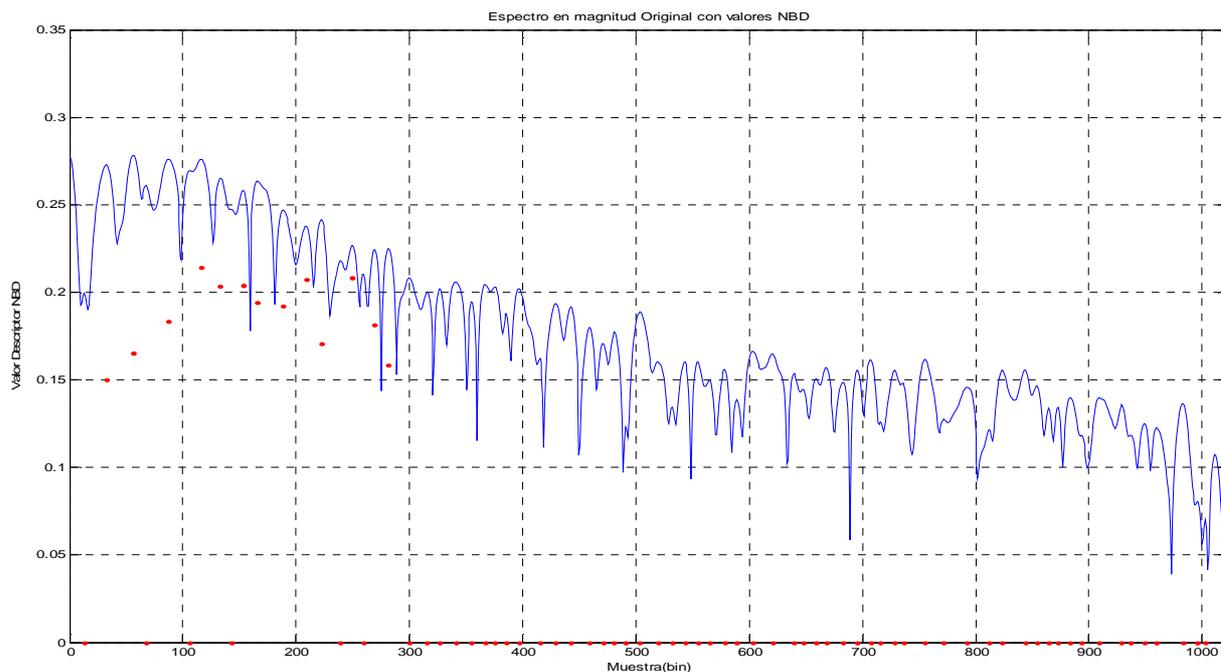
(Fig. 5.2.2.5.2 Valores del descriptor NBD sobre el espectro de la Trama 3)

Se detectan 74 picos, se observa como el valor del descriptor para los parciales y el ruido es muy similar, apenas se distinguen parciales. Solo los primeros parciales son interpretados como tales, a partir del séptimo la gran cantidad de ruido y la transitoriedad de la trama hacen difícil la detección.

Realizamos el filtrado, quedándonos con los valores correspondientes a los picos que estimamos deterministas:

B(2,3,5,7,8,10,11,12,13,14,16,18,19)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.5.3 Valores del NBD para los parciales sobre el espectro de la Trama 3)

Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

- Número de parciales detectados: **13**
- Valor máximo: **0.2144**
- Valor mínimo: **0.1502**
- Valor medio: **0.1872**
- Varianza: **4.3171e-004**

En esta **trama transitoria** se observa **en el vector con los valores del descriptor para los picos deterministas una tendencia creciente alta**, apreciable a pesar de ser únicamente los primeros parciales discernibles.

Esta **tendencia creciente alta** se puede asociar al alto grado de modulación y a una **transitoriedad en la trama**, ya que a los picos más altos en frecuencia les afecta más la modulación, ensanchándolos más y obteniendo mayores valores del descriptor.

5.2.2.6 – CONCLUSIONES A LAS TRAMAS DE LA TROMPETA

“Los datos obtenidos solo podemos contrastarlos entre tramas de un mismo instrumento, para que no afecte la variable diferencial de la envolvente espectral concreta del instrumento”

Comparamos los valores obtenidos para el vector NBD en las tramas del archivo en el que suena la **trompeta**:

Trama 1: Poca modulación

- Número de parciales detectados: **26**
- Valor máximo: **0.2074**
- Valor mínimo: **0.0863**
- **Valor medio: 0.1301**
- Varianza: **4.5732e-004**

Trama 2: Mucha modulación

- Número de parciales detectados: **27**
- Valor máximo: **0.1761**
- Valor mínimo: **0.0870**
- **Valor medio: 0.1418**
- Varianza: **5.7810e-004**

Trama 9: Desvanecimiento

- Número de parciales detectados: **19**
- Valor máximo: **0.1874**
- Valor mínimo: **0.0742**
- **Valor medio: 0.1471**
- Varianza: **9.1385e-004**

Trama 3: Transitoria

- Número de parciales detectados: **13**
- Valor máximo: **0.2144**
- Valor mínimo: **0.1502**
- **Valor medio: 0.1872**
- Varianza: **4.3171e-004**

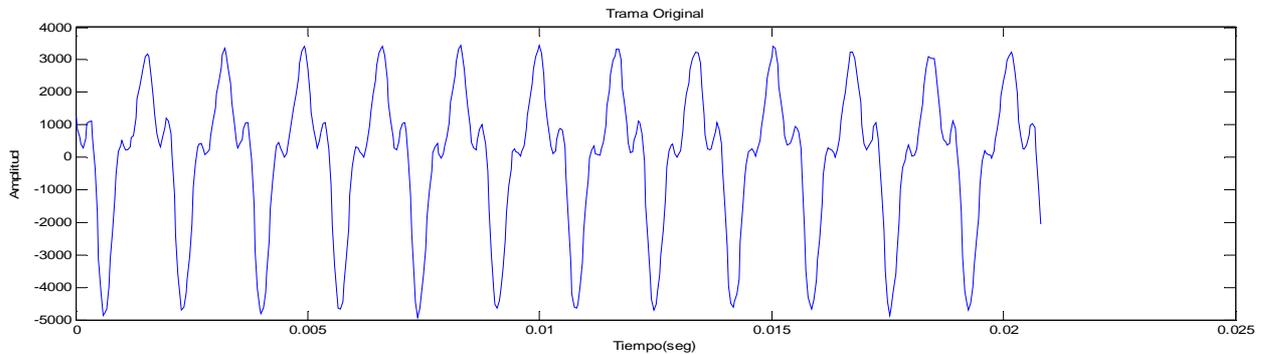
Se observa como **el valor medio del vector con los valores del descriptor para los parciales es mayor a medida que aumenta la modulación de la trama**, ya que la tendencia ascendente de los valores hace que crezca el valor medio.

Por tanto podríamos obtener un **estimador del grado de modulación de una trama a través del valor medio que toma el vector con los valores del descriptor NBD para los picos deterministas.**

Estudiaremos a continuación tramas correspondientes al archivo en el que suena un Shakuachi.

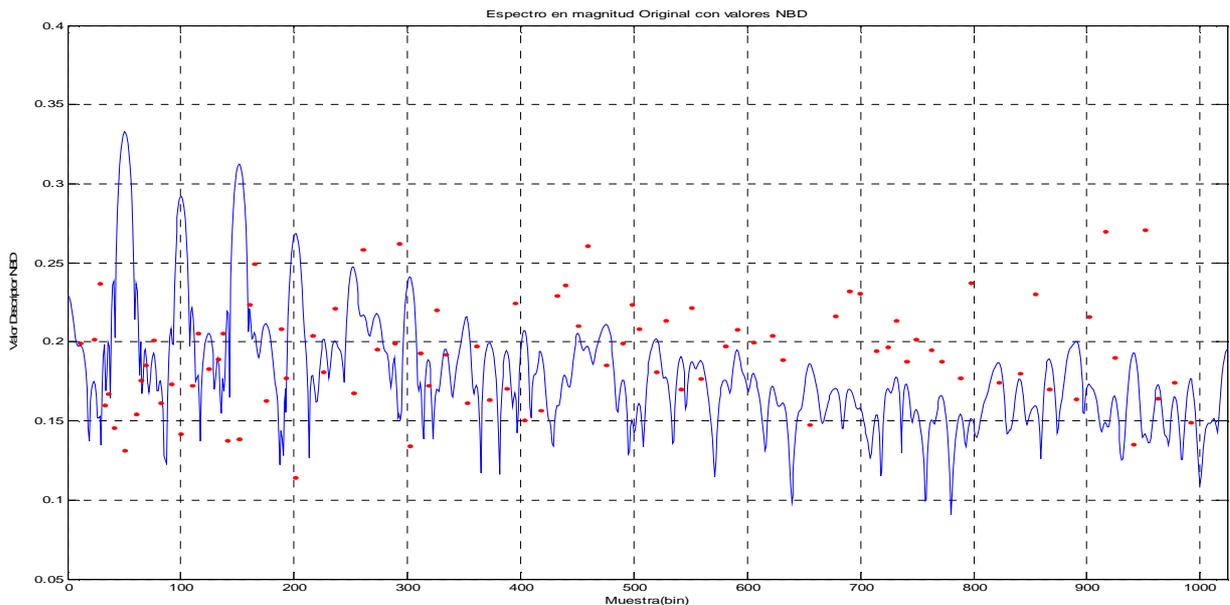
5.2.2.7 – TRAMA 4: cuasi estacionaria ruidosa y con poca modulación. Shakuhachi

Trama muy ruidosa con poca modulación.



(Fig. 5.2.2.7.1 Trama 4, 500 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:



(Fig. 5.2.2.7.2 Valores del descriptor NBD sobre el espectro de la Trama 4)

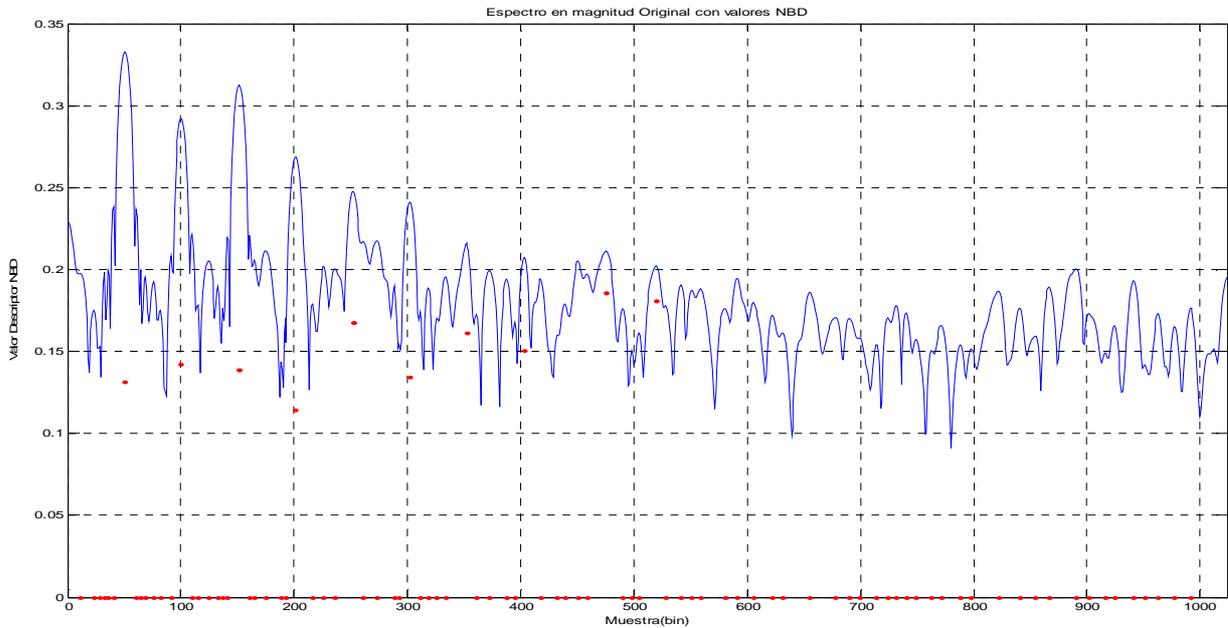
Se detectan 91 picos. El **número de picos detectados** puede depender de las **características concretas del instrumento** (como se distribuye la energía en el espectro según la envolvente) y **del nivel de ruido** presente en la trama, pero no del grado de modulación.

En esta trama y para este instrumento, los valores obtenidos para el descriptor no son tan diferentes para los picos deterministas y los ruidosos como en las tramas de la trompeta, ya que aquí **el alto nivel de ruido hace difícil una clara diferenciación entre lo determinista y lo estocástico.**

Realizamos el filtrado, quedándonos con los valores correspondientes a los picos que estimamos deterministas:

B(7,14,21,27,31,36,41,46,52,56)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.7.3 Valores del NBD para los parciales sobre el espectro de la Trama 4)

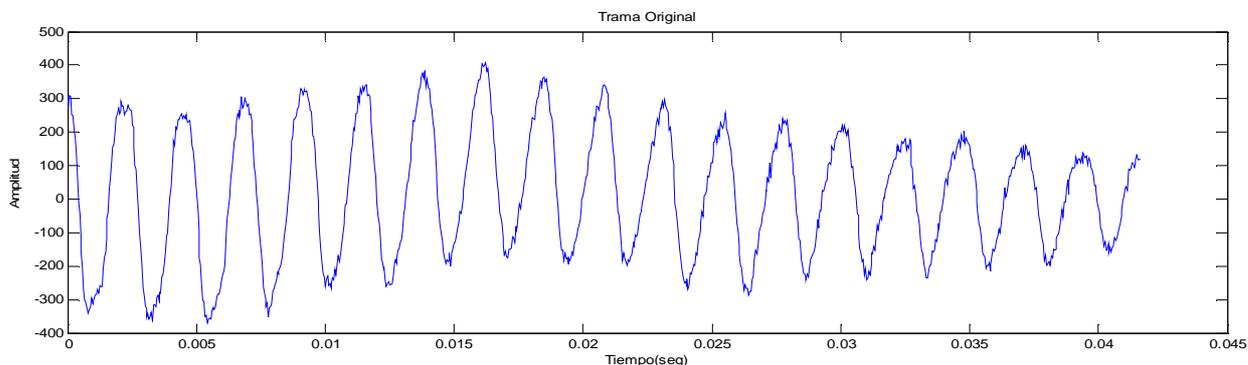
Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

- Número de parciales detectados: **10**
- Valor máximo: **0.1855**
- Valor mínimo: **0.1145**
- Valor medio: **0.15071**
- Varianza: **5.2172e-004**

Se observa en los primeros parciales fácilmente detectados, como **los valores del descriptor en conjunto siguen una tendencia plana**, ya que la trama tenía **poca modulación**.

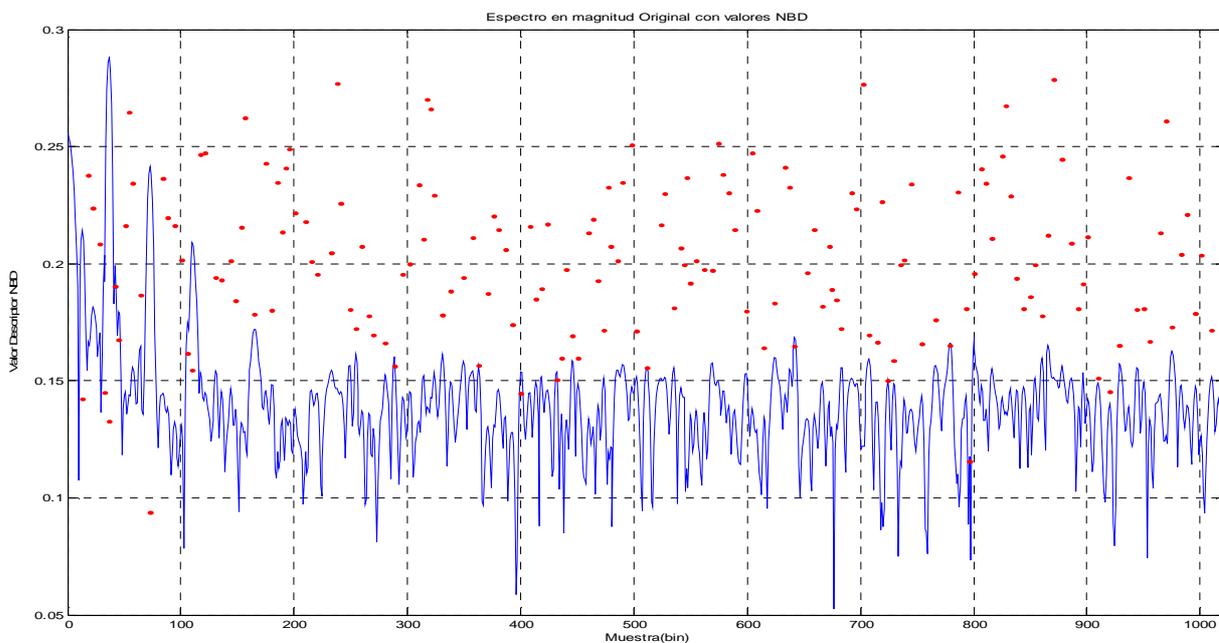
5.2.2.8 – TRAMA 5: cuasi estacionaria ruidosa y con mucha modulación. Shakuachi

Trama muy ruidosa y con mucha modulación AM-FM en el límite de lo cuasi estacionario.



(Fig. 5.2.2.8.1 Trama 5, 1000 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:



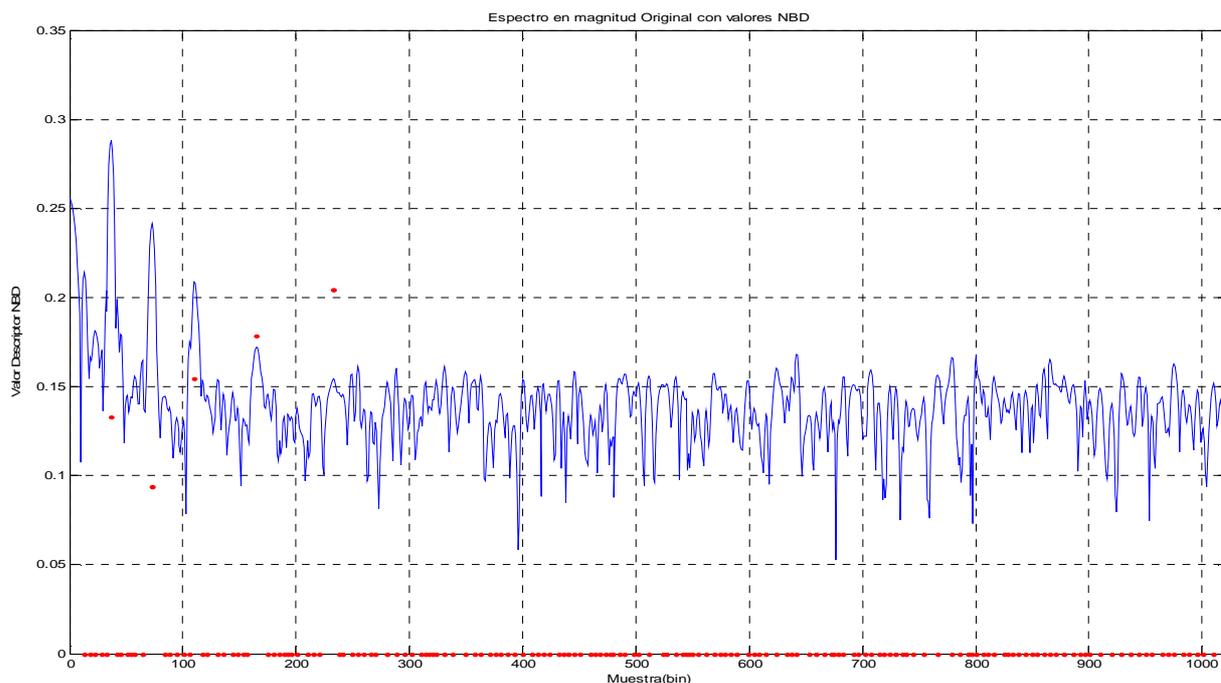
(Fig. 5.2.2.8.2 Valores del descriptor NBD sobre el espectro de la Trama 5)

Se detectan 166 picos. Destaca la gran cantidad de puntos con valores altos asociados a picos detectados como ruidosos, por el alto nivel de ruido y de modulación.

Realizamos el filtrado, quedándonos con los valores correspondientes a los picos que estimamos deterministas:

B(6,13,19,28,39)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.8.3 Valores del NBD para los parciales sobre el espectro de la Trama 5)

Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

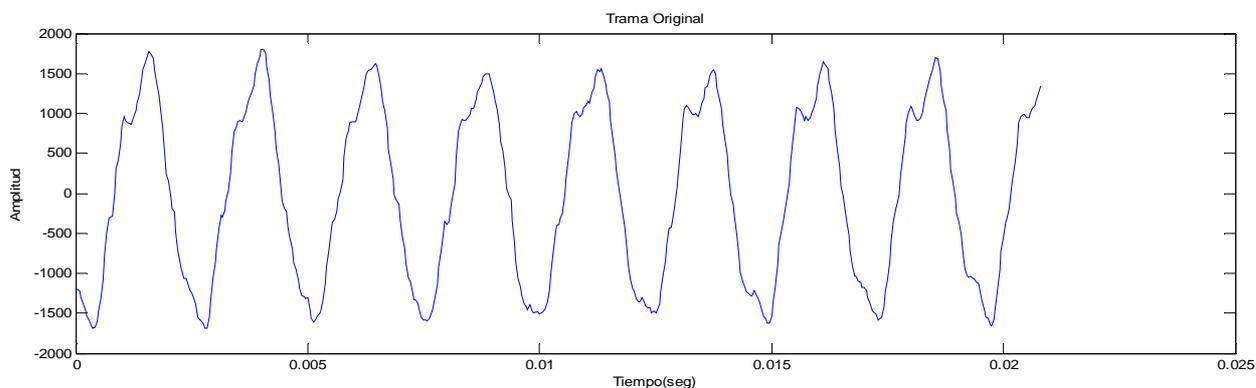
- Número de parciales detectados: **5**
- Valor máximo: **0.2046**
- Valor mínimo: **0.0939**
- Valor medio: **0.1528**
- Varianza: **0.0018**

Se observa claramente, para esta trama con mucho ruido y **mucha modulación**, como en vector con **los valores del descriptor para los picos deterministas tiene una tendencia creciente**. Nuevamente se verifica esta relación entre el grado de modulación y la **tendencia creciente de los valores del descriptor para los parciales**.

El hecho de que la trama tenga más nivel de ruido podría dificultar el proceso de medida, pero como se ha visto en estas tramas muy ruidosas, **incluso en los primeros parciales detectados ya se aprecia la tendencia de los valores del descriptor y se pueden sacar conclusiones sobre el grado de modulación presente en la trama**.

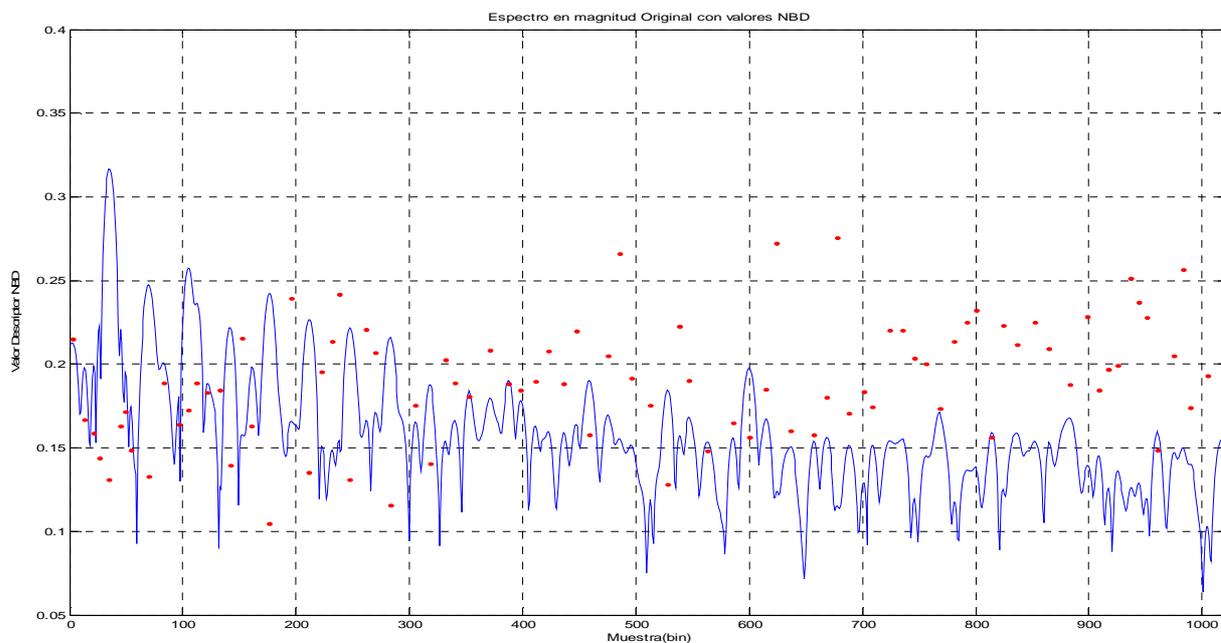
5.2.2.9 – TRAMA 8: ligeramente modulada. Shakuhachi

Trama ligeramente modulada y con ruido:



(Fig. 5.2.2.9.1 Trama 8, 500 muestras)

Aplicamos el algoritmo de clasificación de picos y asignación del valor NBD:

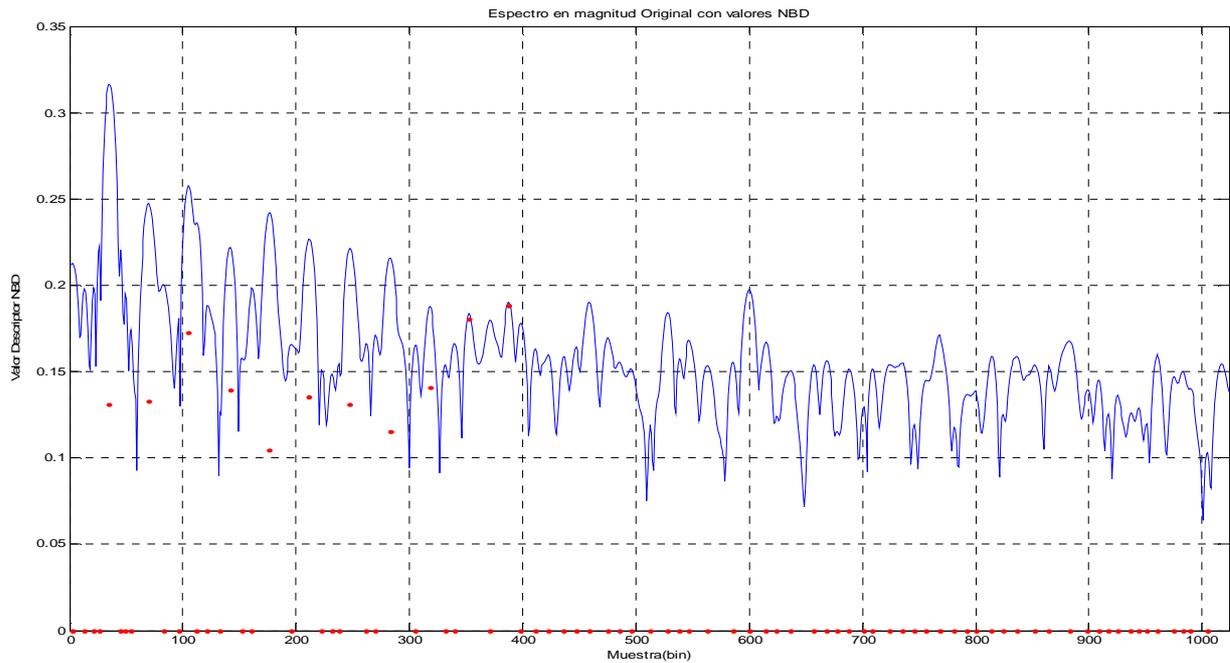


(Fig. 5.2.2.9.2 Valores del descriptor NBD sobre el espectro de la Trama 8)

Se detectan 86 picos. Realizamos el filtrado, quedándonos con los valores correspondientes a los picos que estimamos deterministas:

B(5,9,12,16,19,21,25,28,30,33,35)

Representamos el nuevo vector junto al espectro:



(Fig. 5.2.2.9.3 Valores del NBD para los parciales sobre el espectro de la Trama 3)

Calculamos los estadísticos del vector con los valores del descriptor para los picos deterministas:

- Número de parciales detectados: **11**
- Valor máximo: **0.1882**
- Valor mínimo: **0.1047**
- Valor medio: **0.1428**
- Varianza: **7.0035e-004**

Al tener esta trama **poca modulación** se observa nuevamente una **tendencia plana** en los valores del descriptor para los primeros parciales detectados correctamente.

5.2.2.10 – CONCLUSIONES A LAS TRAMAS DEL SHAKUHACHI

Comparamos los valores obtenidos para el vector NBD en las tramas del archivo en el que suena el **Shakuhachi**:

Trama 8: Poca modulación

- Número de parciales detectados: **11**
- Valor máximo: **0.1882**
- Valor mínimo: **0.1047**
- **Valor medio: 0.1428**
- Varianza: **7.0035e-004**

Trama 4: Ligera modulación

- Número de parciales detectados: **10**
- Valor máximo: **0.1855**
- Valor mínimo: **0.1145**
- **Valor medio: 0.15071**
- Varianza: **5.2172e-004**

Trama 5: Mucha modulación

- Número de parciales detectados: **5**
- Valor máximo: **0.2046**
- Valor mínimo: **0.0939**
- **Valor medio: 0.1528**
- Varianza: **0.0018**

Se observa de nuevo como el **valor medio del vector** con los valores del descriptor para los parciales **es mayor a medida que aumenta la modulación de la trama**.

5.2.3 CONCLUSIONES AL ESTUDIO DEL GRADO DE MODULACIÓN SEGÚN LOS VALORES NBD

Concluimos que el grado de modulación de una trama puede ser estimado a través del valor medio que toma el vector con los valores del descriptor NBD para los picos deterministas.

No profundizaremos más en el estudio de este método, a falta de determinar unos umbrales adecuados con los que realizar correctamente el baremo del grado de modulación. **Dejamos planteado para líneas futuras de investigación y otras aplicaciones este método, totalmente nuevo y lo damos por adecuado para estimar el grado de modulación presente en una trama.**

5.3 MÉTODO 3: SPARSITY DE LA DWT Y LA MDCT

El tercer y último método esta basado en un estudio planteado por Stéphane Molla y Bruno Torrèsani en su artículo “**Determining local transientness of audio signals**” [6]. Estudio basado en el “Sparsity” o esparcimiento de la señal expandida en bases de Wavelet y Coseno, mediante la transformada discreta de Wavelet (DWT, Discrete Wavelet Transform) y la transformada discreta modificada del coseno (MDCT, Modified Discrete Cosine Transform).

5.3.1 DESCRIPCIÓN

En el artículo mencionado se plantea descomponer la trama objeto de estudio en **3 componentes**, una estaría formada por la **parte transitoria**, otra por la **parte tonal** o armónica y la tercera sería la componente estocástica o **ruido**:

$$x = \sum_{\lambda \in \Delta\lambda} \alpha_{\lambda} \Psi_{\lambda} + \sum_{\delta \in \Delta\delta} \beta_{\delta} \omega_{\delta} + r = x_{trans} + x_{tonal} + r \quad (\text{Ecuación 5.3.1.1})$$

Donde los parámetros α_{λ} y β_{δ} son coeficientes que multiplican a las funciones base en las que se descompone la señal, Ψ_{λ} y ω_{δ} .

Para evaluar el grado de transitoriedad ó tonalidad de la trama bajo estudio, se definen dos índices complementarios entre si que determinan la proporción que existe de ambas componentes:

$$\hat{I}_{tonal} = \frac{\hat{N}_{\Psi}}{\hat{N}_{\Psi} + \hat{N}_{\omega}}$$

(Ecuación 5.3.1.2)

$$\hat{I}_{trans} = \frac{\hat{N}_{\omega}}{\hat{N}_{\Psi} + \hat{N}_{\omega}}$$

(Ecuación 5.3.1.3)

$$\hat{I}_{tonal} + \hat{I}_{trans} = 1$$

(Ecuación 5.3.1.4)

Donde:

$$\hat{N}_{\Psi}(x) = 2^{D_{\Psi}}$$

(Ecuación 5.3.1.5)

$$\hat{N}_{\omega}(x) = 2^{D_{\omega}}$$

(Ecuación 5.3.1.6)

Con:

$$D_{\Psi}(x) = \frac{1}{N} \sum_{n=1}^N \log_2 |\langle x, \Psi_n \rangle|^2$$

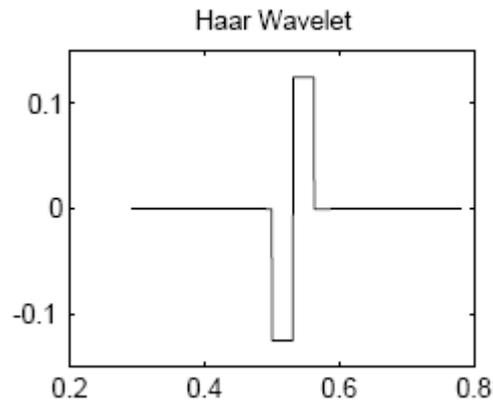
(Ecuación 5.3.1.7)

$$D_{\omega}(x) = \frac{1}{N} \sum_{n=1}^N \log_2 |\langle x, \omega_n \rangle|^2$$

(Ecuación 5.3.1.8)

Donde: $\langle x, \omega_n \rangle$ es la DWT clásica (Haar) cuya base de descomposición se asemejan a rectángulos y es adecuada para las partes transitorias, y $\langle x, \Psi_n \rangle$ es la MDCT similar a la FFT pero solo con cosenos y adecuada para las partes tonales o armónicas.

Implementamos el método para calcular los índices de transitoriedad y tonalidad (\hat{I}_{trans} , \hat{I}_{tonal}) y los evaluamos para varias tramas⁶.



(Fig. 5.3.1 Base de descomposición Haar para la transformada discreta de Wavelet)

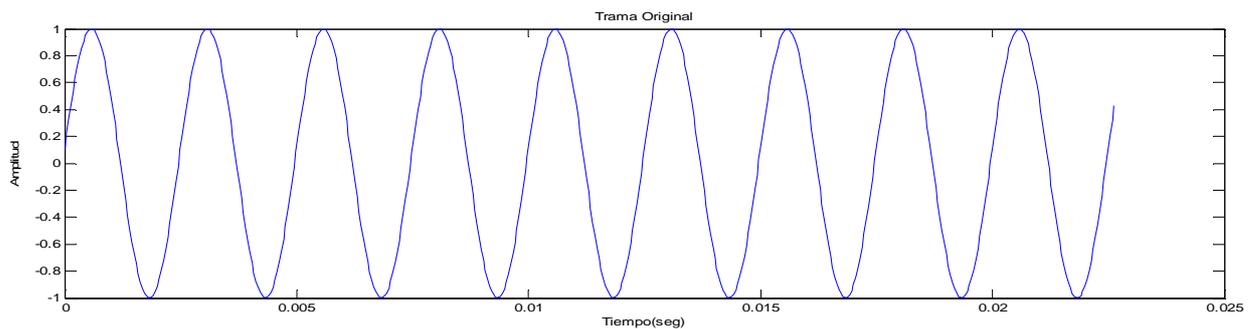
⁶ El método se encuentra implementado en la función “**sparsity.m**” que devuelve los índices de tonalidad y transitoriedad.

5.3.2 PRUEBA DEL MÉTODO

Procederemos a probar el tercer método sobre algunas de las tramas estudiadas anteriormente.

5.3.2.1 – Referencia 1. Tono puro.

Veremos que sucede si tomamos como referencia un tono puro, por definición tendrá únicamente componente tonal. ($f_0=400\text{Hz}$, $\text{SR}=22050$ y 500 muestras)



(Fig. 5.3.2.1 Tono puro, 500 muestras, $f_0=400\text{Hz}$, $\text{SR}=22050$)

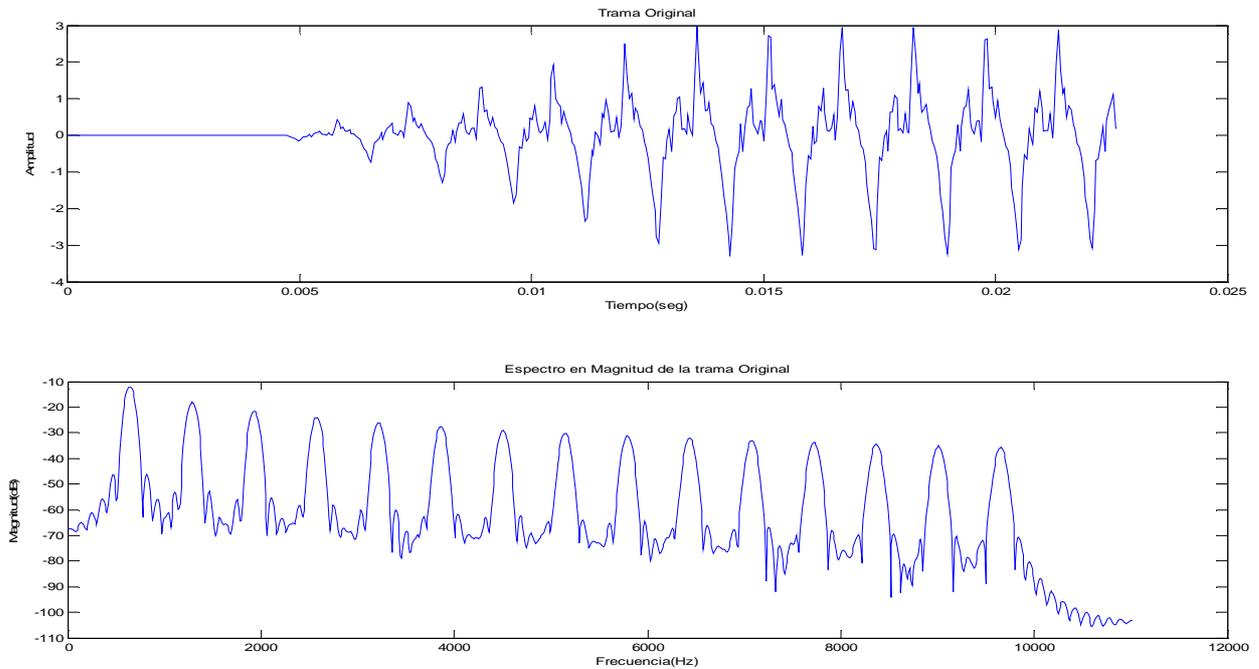
Obtenemos los siguientes índices:

- $\hat{I}_{trans} = 1.8539\text{e-}007$
- $\hat{I}_{tonal} = 1.0000$

El método funciona, **determina que la trama es tonal.**

5.3.2.2 – Referencia 2. Señal sintética.

Veremos que sucede si tomamos como referencia una señal sintetizada con la función “**TransientModel.m**” que implementa señales con parte tonal y transitoria en proporciones dadas por los parámetros de entrada. Creamos primeramente una señal sin ruido, con **SR=22050, 15 armónicos, 500 muestras, historia de 100 muestras y transitorio de 200.**



(Fig. 5.3.2.2 Señal sintética, 500 muestras, SR=22050, 15 armónicos, hist. 100, trans. 200)

Aplicamos el algoritmo y obtenemos los siguientes valores:

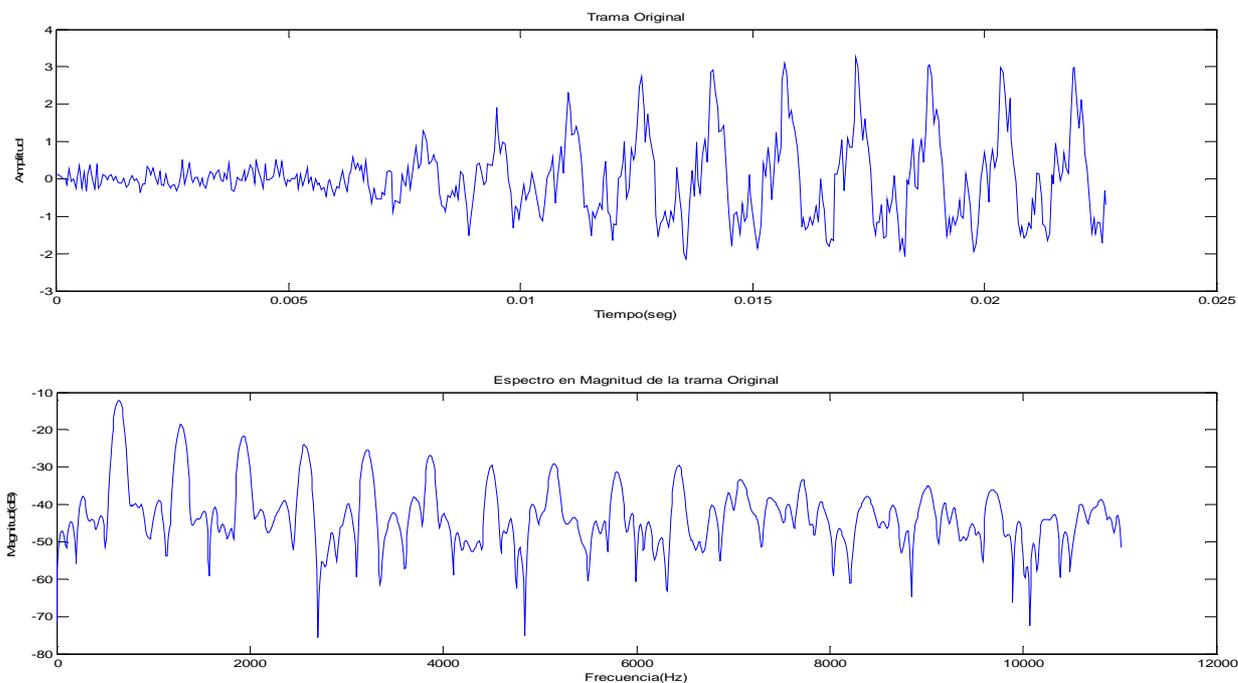
- $\hat{I}_{trans} = 0.7219$
- $\hat{I}_{tonal} = 0.2781$

Determina que la trama es **transitoria**.

Establecemos como umbral para decidir si la trama es transitoria o tonal el 0.5.

5.3.2.3 – Referencia 3. Señal sintética transitoria ruidosa. “Caso anterior con ruido”.

Veremos que sucede si tomamos ahora la señal transitoria sintetizada de **referencia 2 y le añadimos ruido**. Asignamos a la función “TransientModel.m” los siguientes parámetros: **SR=22050, 15 armónicos, 500 muestras para la ventana, SNRp=10, historia de 100 muestras y transitorio de 200.**



(Fig. 5.3.2.3 Señal sintética, 500 muestras, SR=22050, 15 armónicos, hist. 200, trans. 5)

Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.8956$

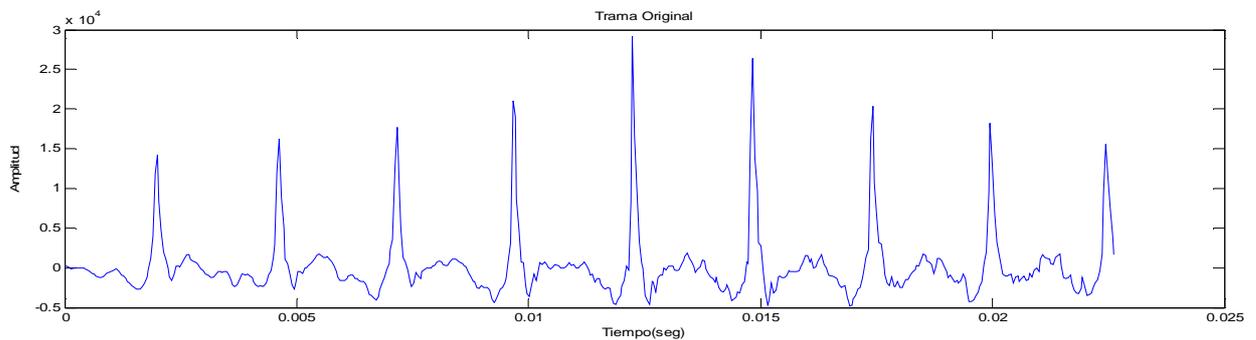
- $\hat{I}_{tonal} = 0.1044$

Se obtiene un resultado algo diferente en presencia de ruido pero se sigue concluyendo que la trama es **transitoria**.

A continuación estudiaremos tramas de señales musicales.

5.3.2.4 – TRAMA 2: cuasi estacionaria poco ruidosa con fuerte modulación. Trumpet

Trama con una modulación AM en el límite de lo cuasi estacionario.



(Fig. 5.3.2.4 Trama 2, 500 muestras)

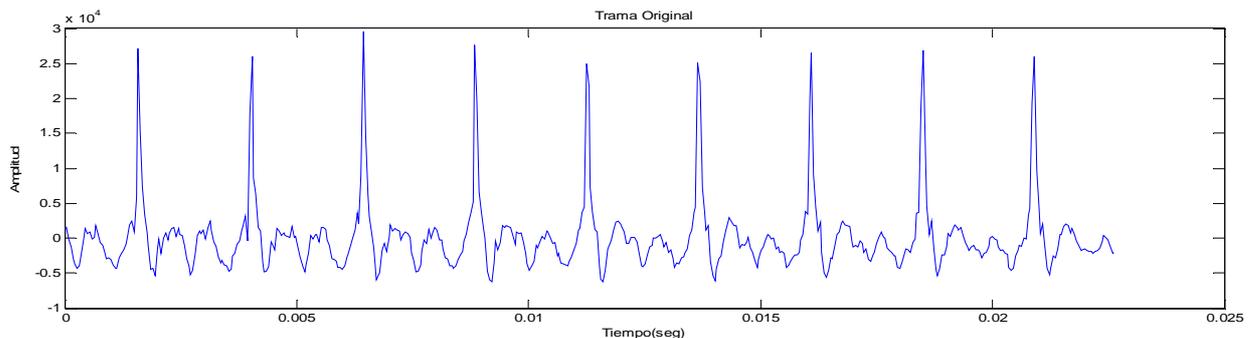
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.3945$
- $\hat{I}_{tonal} = 0.6055$

Determina que la trama es **tonal**.

5.3.2.5 – TRAMA 1: cuasi estacionaria poco ruidosa con poca modulación. Trumpet

Trama cuasi estacionaria poco ruidosa y con poca modulación.



(Fig. 5.3.2.5 Trama 1, 500 muestras)

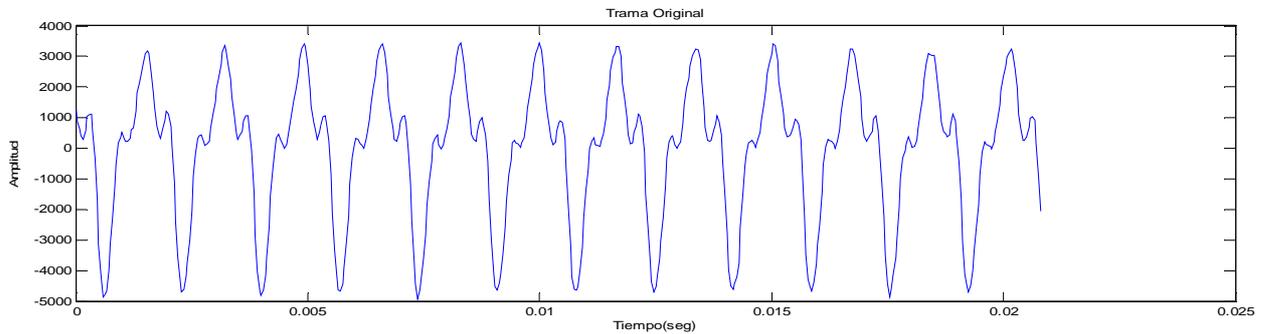
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.0524$
- $\hat{I}_{tonal} = 0.9476$

Determina que la trama es **tonal**.

5.3.2.6 – TRAMA 4: cuasi estacionaria ruidosa y con poca modulación. Shakuhi

Trama muy ruidosa con poca modulación.



(Fig. 5.3.2.6 Trama 4, 500 muestras)

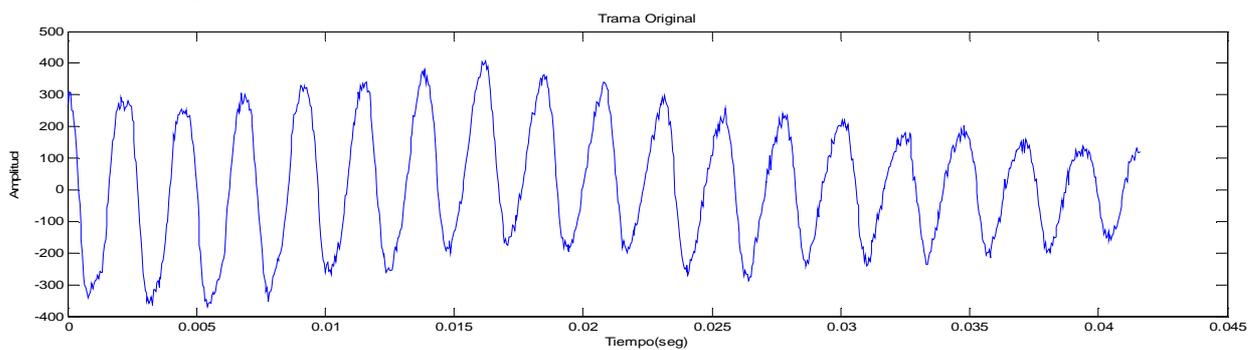
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.0077$
- $\hat{I}_{tonal} = 0.9923$

Determina que la trama es **tonal**.

5.3.2.7 – TRAMA 5: cuasi estacionaria ruidosa y con mucha modulación. Shakuhi

Trama muy ruidosa y con mucha modulación AM-FM en el límite de lo cuasi estacionario.



(Fig. 5.3.2.7 Trama 5, 500 muestras)

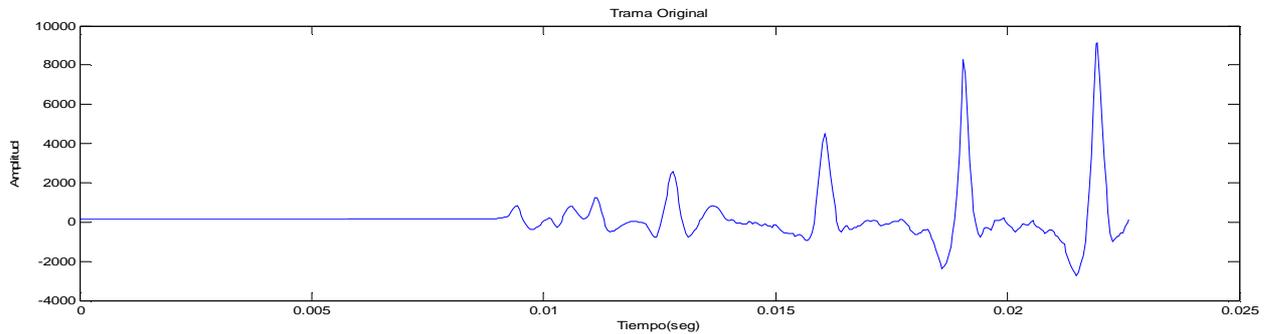
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.2561$
- $\hat{I}_{tonal} = 0.7439$

Determina que la trama es **tonal**.

5.3.2.8 – TRAMA 3: transitoria poco ruidosa. Trumpet

Veremos que sucede con la trama transitoria.



(Fig. 5.3.2.8 Trama 3, 500 muestras)

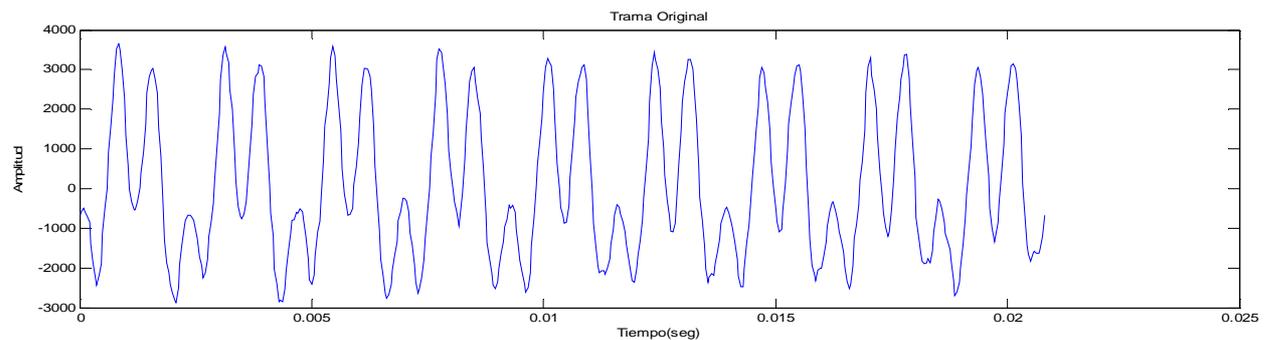
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.9827$
- $\hat{I}_{tonal} = 0.0173$

Determina que la trama es **transitoria**.

5.3.2.9 – TRAMA 6: poca modulación y ruidosa. Shakuhachi

Trama bastante estacionaria.



(Fig. 5.3.2.9 Trama 6, 500 muestras)

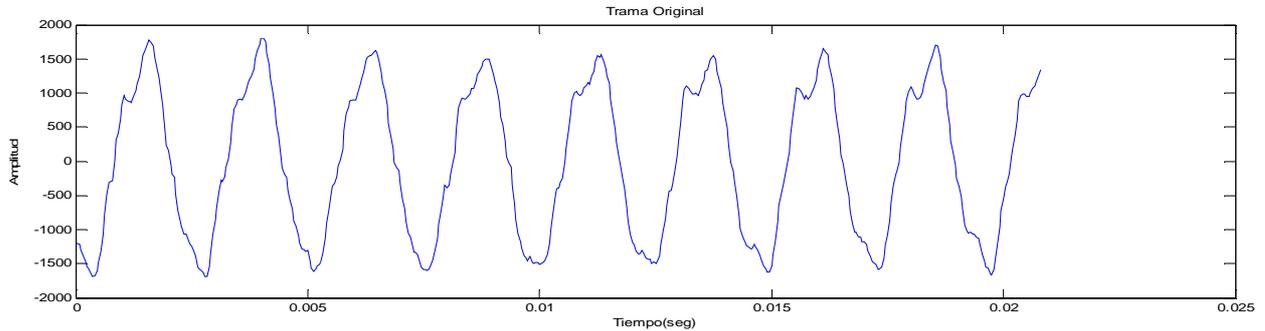
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.0235$
- $\hat{I}_{tonal} = 0.9765$

Determina que la trama es **tonal**.

5.3.2.10 – TRAMA 8: ligeramente modulada. Shakuhachi

Trama ligeramente modulada y con ruido:



(Fig. 5.3.2.10 Trama 8, 500 muestras)

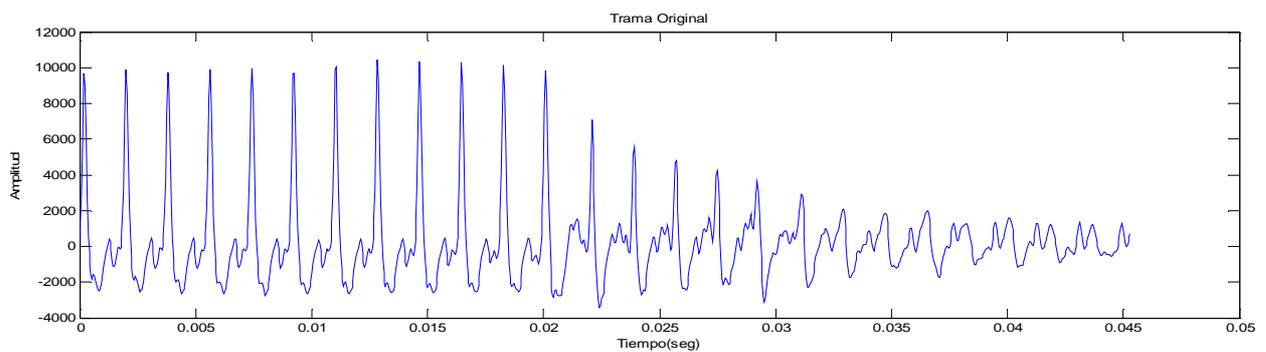
Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.0990$
- $\hat{I}_{tonal} = 0.9010$

Determina que la trama es **tonal**.

5.3.2.11 – TRAMA 9: desvanecimiento. Trumpet

Trama con mucha modulación y ruido. Veremos que nos dicen los índices.



(Fig. 5.3.2.11 Tono puro, 500 muestras, $f_0=400\text{Hz}$, $SR=22050$)

Aplicamos el algoritmo y obtenemos los siguientes valores:

- $\hat{I}_{trans} = 0.7411$
- $\hat{I}_{tonal} = 0.2589$

Determina que la trama es **transitoria**.

5.3.3 CONCLUSIONES AL MÉTODO BASADO EN EL SPARSITY

Se ha demostrado como el método funciona y determina para las tramas estudiadas si éstas son transitorias o tonales. No profundizaremos más en este método a falta de determinar unos umbrales con los que poder establecer los distintos grados de modulación presentes en una trama.

5.4 CONCLUSIONES A LA BÚSQUEDA DE UN MÉTODO DE BAREMO DEL GRADO DE MODULACIÓN

Tras estudiar los 3 métodos propuestos, concluimos quedarnos con el primero de ellos ya que calcular el tiempo medio y la duración temporal de una trama es muy rápido computacionalmente y hemos podido estimar unos umbrales claros con los que definir los distintos grados de modulación presentes en una trama.

Es por tanto el **método 1** el que utilizaremos para implementar el algoritmo de **modelado optimizado**.

6. Búsqueda de un método para estimar el impacto del ruido en una trama

A continuación se describirá e implementará un método con el cual **decidir si una trama es ruidosa o no**, con el fin de junto al grado de modulación, poder clasificarla en uno de los 5 grupos establecidos anteriormente y asignarle unos parámetros de modelado óptimos.

Como el título indica, se tratará de estimar el **impacto del ruido en una trama**, según **como afecte el ruido al conjunto de todos los parciales del espectro**. No se tratará de medir el nivel de ruido. Se clasificarán las tramas como **ruidosas o poco ruidosas**.

El método está basado en la información proporcionada por el algoritmo de clasificación de picos, para una trama dada aplicaremos el algoritmo ya implementado de asignación del valor del descriptor NBD a cada pico. **Separaremos los picos ruidosos** estableciendo un **umbral fijo de 0.15** (utilizado habitualmente en otras aplicaciones). Se considerarán picos ruidosos los que obtengan un valor del NBD por encima de dicho umbral.

Contaremos el número de picos ruidosos detectados en toda la trama y dividiremos dicho resultado entre el número total de picos detectados (tanto ruidosos como deterministas). Multiplicaremos el cociente por 100 y **obtendremos una estimación en tanto por ciento del grado de picos ruidosos detectados en la trama**, y por tanto del grado o **medida en la que el ruido afecta a los parciales en la trama**.

Inicialmente se pensó en calcular una estimación del ruido realizando un cociente entre la energía asociada a los picos deterministas y la energía asociada a los picos ruidosos, similar al cálculo de la **SNR** (Signal to Noise Ratio) de la trama. Hay que mencionar que este planteamiento **para estimar el grado en el que el ruido afecta a la totalidad de los parciales en la trama no es correcto**. Puede darse el caso según la envolvente concreta del instrumento, en el que por ejemplo los 3 primeros parciales tengan mucha energía, estando el resto inmersos en el ruido y se obtenga una SNR alta (debido a la alta energía los 3 primeros parciales) por tanto no nos informaría de cómo el ruido está afectando a la totalidad de los parciales.

Implementamos el método descrito basado en clasificación de picos y el descriptor NBD y lo aplicamos a varias de las tramas estudiadas anteriormente⁷.

⁷ El método se encuentra implementado en la función **“Estima_ruido.m”** que devuelve la estimación del ruido en tanto por ciento.

6.1 PRUEBA DEL MÉTODO Y ESTIMACIÓN DE UN UMBRAL

Aplicamos el método para determinar el impacto del ruido en una trama a varias de las estudiadas anteriormente. Estimaremos un umbral a partir del cual decidir si la trama es **poco ruidosa** o **ruidosa**.

El método se aplica a las tramas que se han considerado cuasi estacionarias, ya que en las tramas transitorias no hacemos distinción entre ruidosas o poco ruidosas.

	TRAMA	VALOR ESTIMADOR	GRADO RUIDOSO DETERMINADO
Trama 1	Poco ruido y poca modulación. Trumpet	69.14%	Poco ruidosa
Trama 2	Poco ruido y mucha modulación. Trumpet	72.41%	Poco ruidosa
Trama 4	Mucho ruido y poca modulación. Shakuhachi	89.01%	Ruidosa
Trama 5	Mucho ruido y mucha modulación. Shakuhachi	95.78%	Ruidosa
Trama 6	Mucho ruido y poca modulación. Shakuhachi	89.65%	Ruidosa
Trama 7	Ruidosa Pura. Shakuhachi	96.29%	Ruidosa
Trama 8	Mucho ruido y poca modulación. Shakuhachi	84.88%	Ruidosa

Se observa como el método diseñado determina correctamente el impacto del ruido en las tramas. Podemos establecer como umbral para decidir si una trama es poco ruidosa o ruidosa un valor de nuestro estimador del **80%**.

7. Implementación del modelado optimizado y estudio comparativo

Implementamos el algoritmo de modelado optimizado, de modo que según el grado determinado de ruido y de modulación, se apliquen los parámetros y tipo de modelado óptimos a las características de cada trama en concreto.

Para **determinar el grado de modulación de la trama** emplearemos el primer método, basado en el **tiempo medio y la duración temporal**, donde ya tenemos estimados unos umbrales de decisión.

Para **determinar el impacto del ruido en la trama** emplearemos el método diseñado en el apartado anterior, basado en el descriptor **NBD** y la **proporción de picos ruidosos detectados**.

El algoritmo mejorado **modela tanto la energía armónica como la transitoria** de manera optimizada. Se implementa clasificando primeramente las tramas en uno de los 5 grupos establecidos y posteriormente modelando cada una de ellas con los parámetros y tipo de modelado que se han determinado óptimos en el estudio.

TIPO DE TRAMA	MÉTODO DE EXTRACCIÓN DE LA PARTE DETERMINISTA
Poca modulación y poco ruido	Modelado no estacionario polinomial grado 3
Mucha modulación y poco ruido	Modelado no estacionario polinomial grado 6
Poca modulación y mucho ruido	Modelado no estacionario polinomial grado 2
Mucha modulación y mucho ruido	Modelado no estacionario polinomial grado 6
Transitoria	Modelado lineal por tramos Breaks= 5

En la separación de la secuencia de entrada en tramas se hace uso de la técnica de solapamiento entre tramas (Overlap) para evitar discontinuidades al pasar de una trama a otra.

⁸ El algoritmo de modelado optimizado se encuentra implementado en la función **“Modelado_optimo_overlap.m”** que devuelve la parte determinista modelada y el residuo obtenido.

7.1 ESTUDIO COMPARATIVO

Para realizar el estudio comparativo entre el nuevo algoritmo de modelado optimizado y los algoritmos clásicos de modelado en los que se aplicaban unos parámetros fijos para toda la secuencia, utilizaremos una de las secuencias audio de las que hemos extraído tramas para los estudios anteriores. En concreto la secuencia correspondiente a la Trompeta, archivo ‘trumpet4.aiff’.

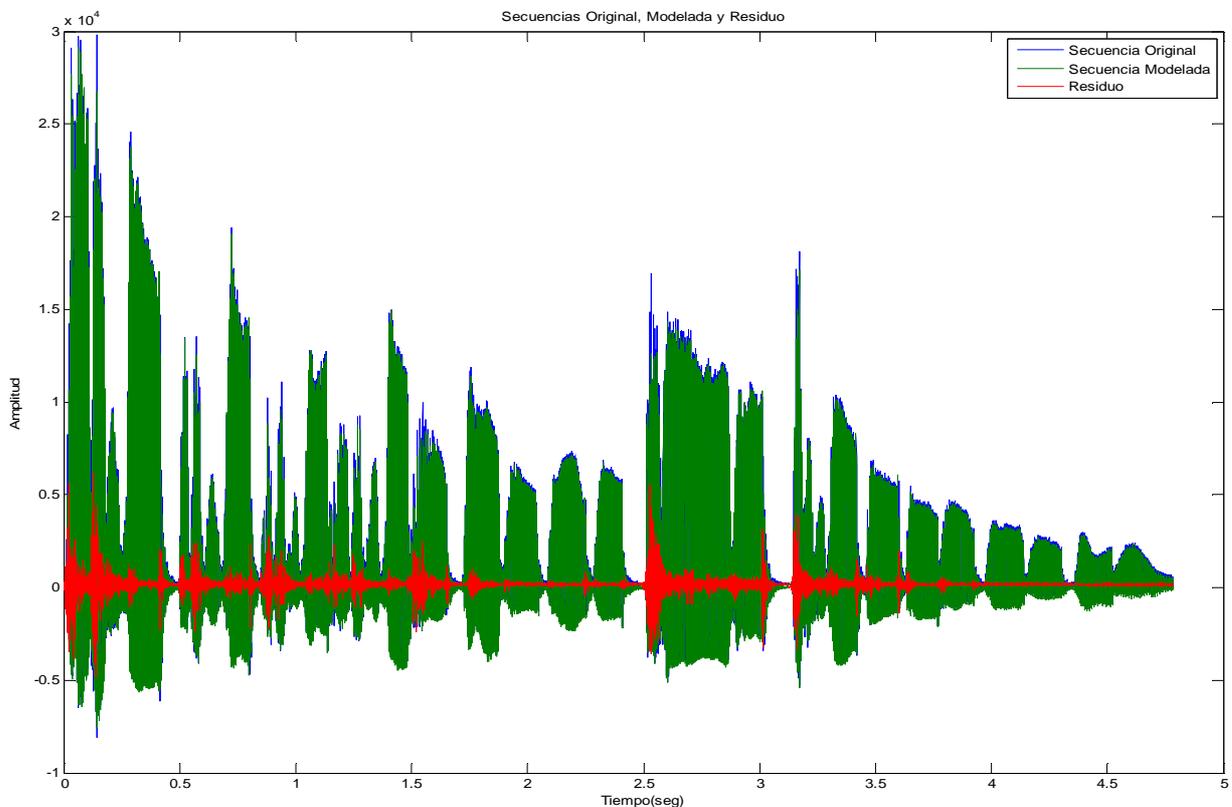
Modelamos la secuencia entera con nuestro algoritmo optimizado, con el algoritmo clásico de modelado polinomial de grado fijo 6, 3, 2, 1 y por último con el algoritmo de modelado lineal por tramos.

Estudiaremos el residuo obtenido para todos los tipos de modelado, calculando la energía total de ruido extraída con cada algoritmo y describiendo la percepción sonora tanto del residuo como de la parte determinista modelada.

En todos los modelados aplicaremos ventana de 500 muestras y $N_{harm}=28$.

7.1.1 – Modelado Clásico Polinomial Grado 6:

Modelamos la secuencia entera y obtenemos las siguientes señales:



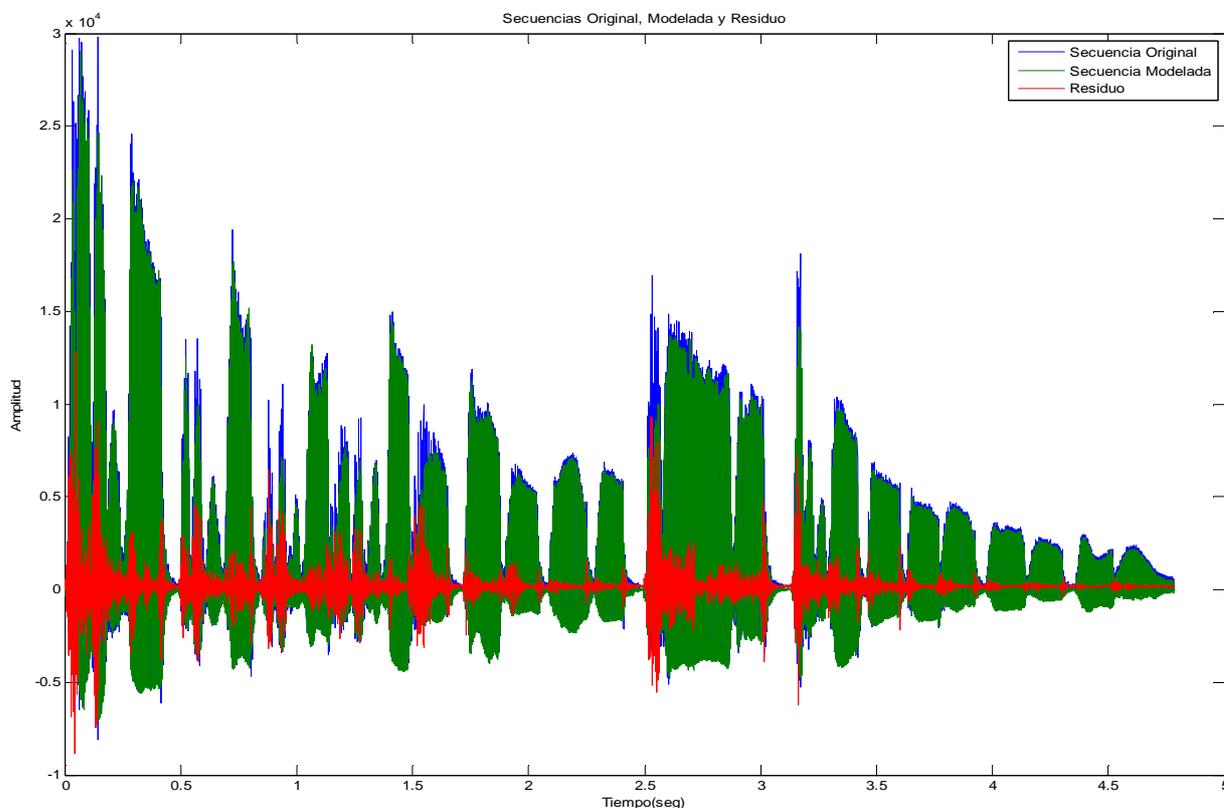
(Fig. 7.1.1 Secuencias obtenidas con el Modelado clásico, $P=6$, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **4.6056e+011**
- Energía del residuo: **7.2492e+009**

7.1.2 – Modelado Clásico Polinomial Grado 3:

Modelamos la secuencia entera y obtenemos las siguientes señales:



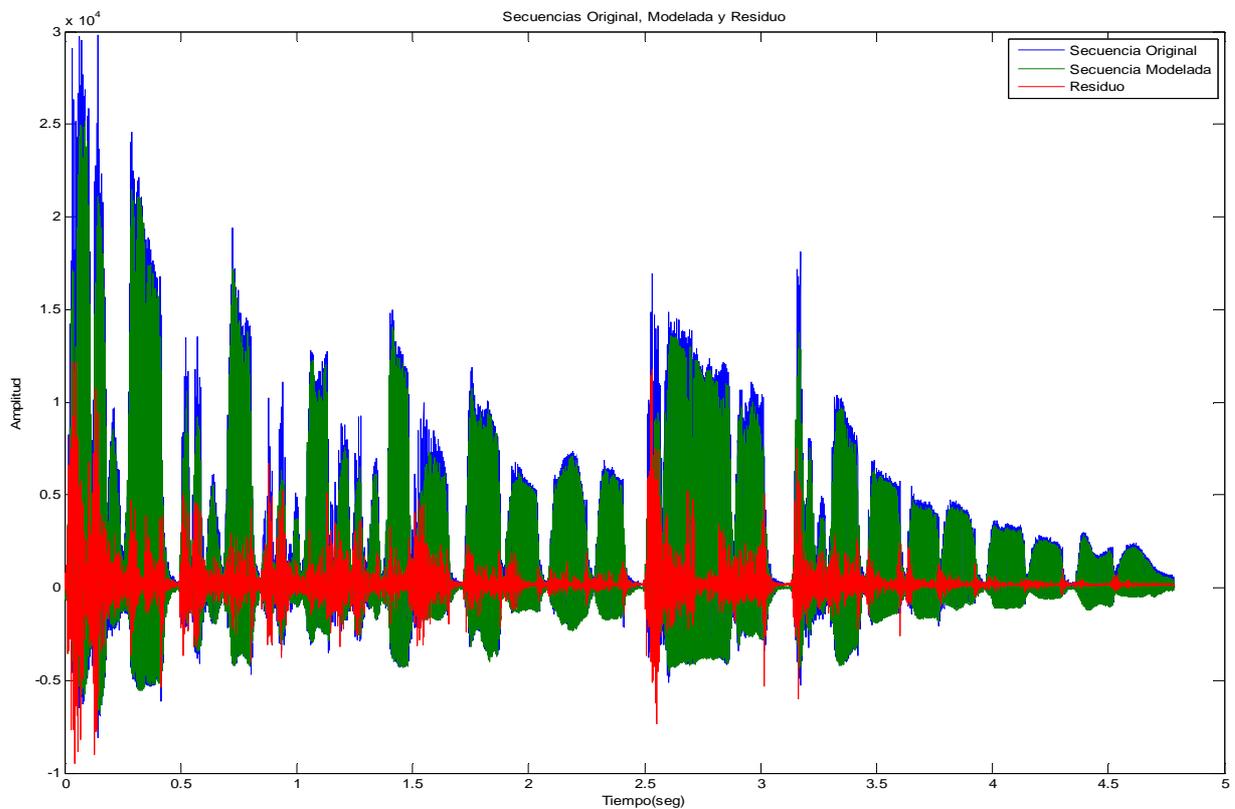
(Fig. 7.1.2 Secuencias obtenidas con el Modelado clásico, $P=3$, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **4.4597e+011**
- Energía del residuo: **1.8750e+010**

7.1.3 – Modelado Clásico Polinomial Grado 2:

Modelamos la secuencia entera y obtenemos las siguientes señales:



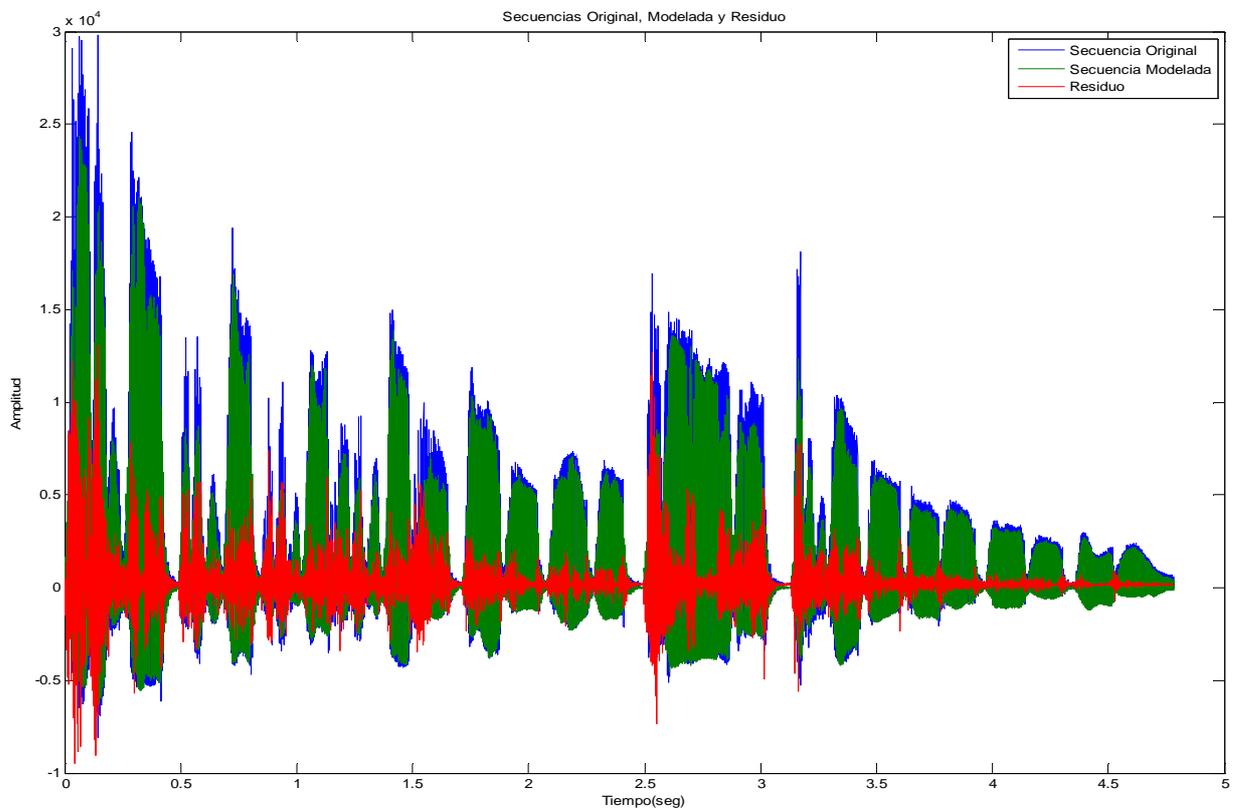
(Fig. 7.1.3 Secuencias obtenidas con el Modelado clásico, $P=2$, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **4.1361e+011**
- Energía del residuo: **2.7559e+010**

7.1.4 – Modelado Clásico Polinomial Grado 1:

Modelamos la secuencia entera y obtenemos las siguientes señales:



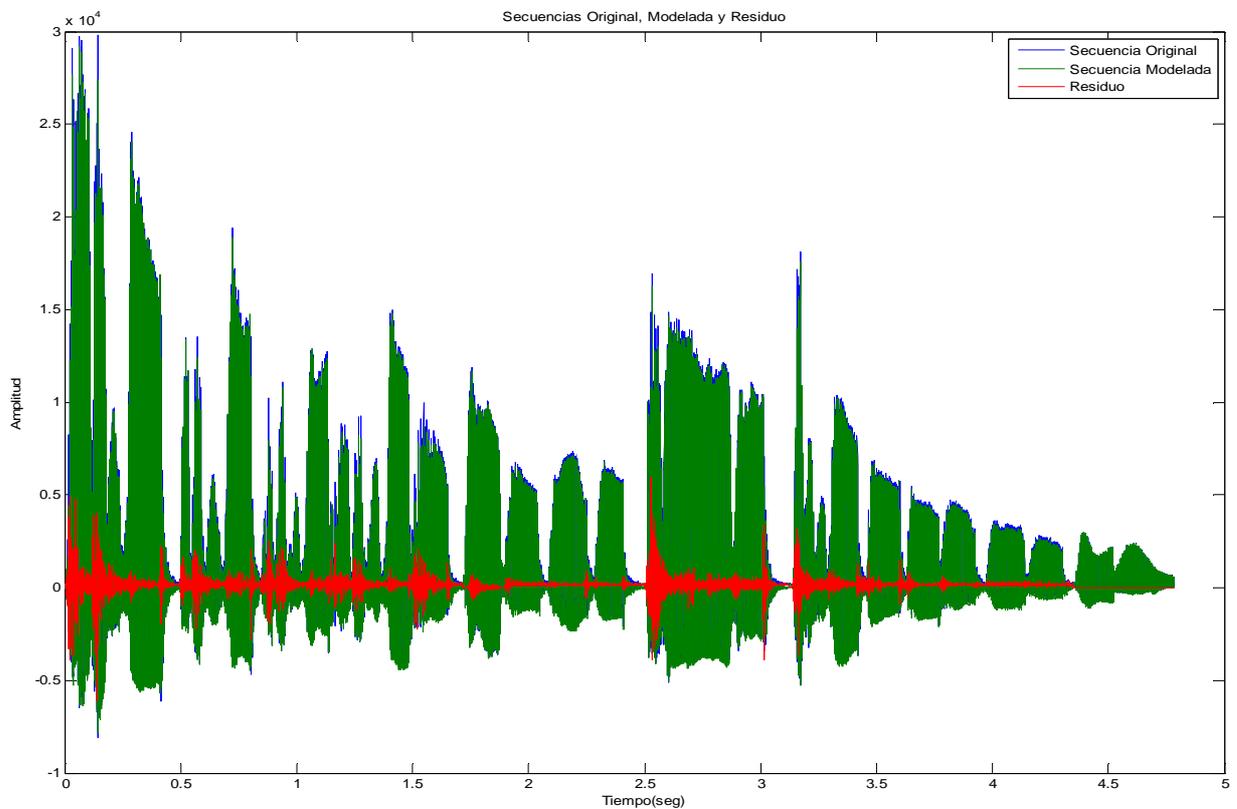
(Fig. 7.1.4 Secuencias obtenidas con el Modelado clásico, $P=1$, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **3.8689e+011**
- Energía del residuo: **3.8808e+010**

7.1.5 – Modelado Clásico Lineal por tramos Breaks=5:

Modelamos la secuencia entera y obtenemos las siguientes señales:



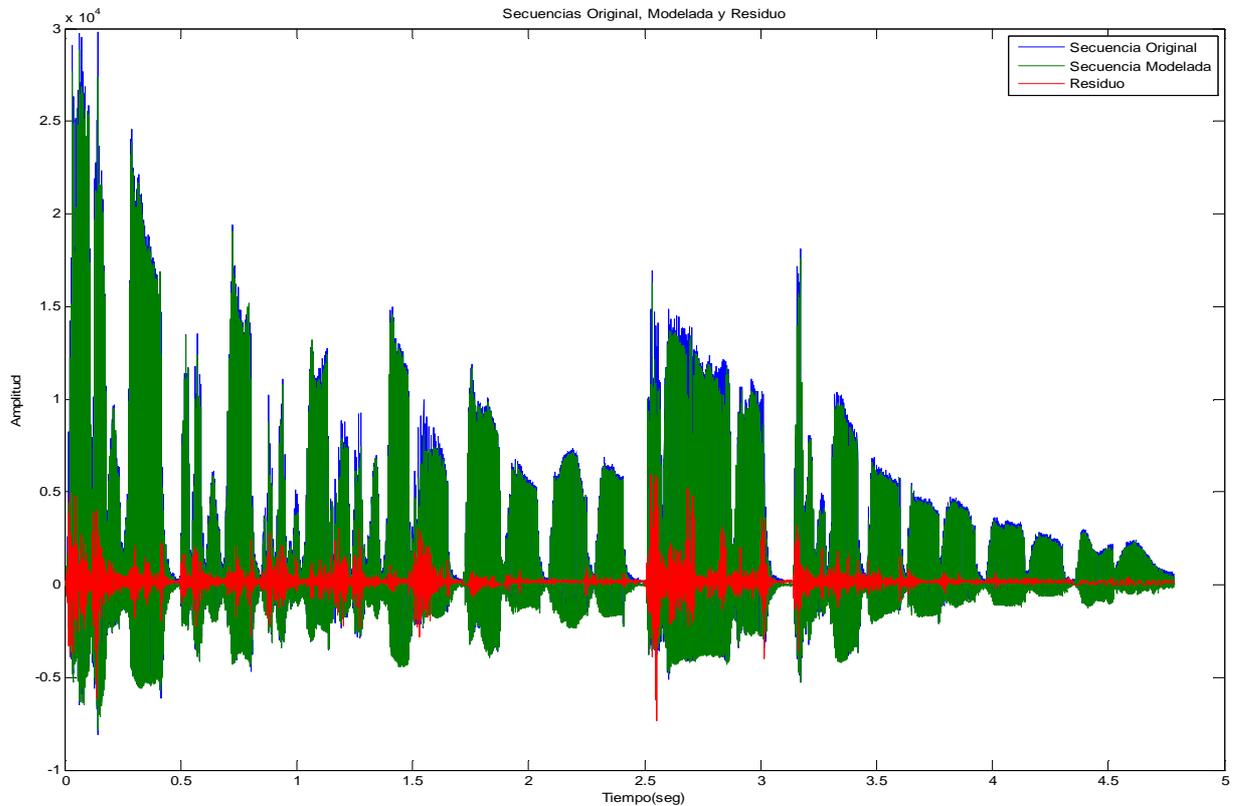
(Fig. 7.1.5 Secuencias obtenidas con el Modelado clásico, Breaks=5, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **4.6226e+011**
- Energía del residuo: **6.1669e+009**

7.1.6 – Nuevo Modelado Optimizado:

Modelamos la secuencia entera y obtenemos las siguientes señales:



(Fig. 7.1.6 Secuencias obtenidas con el Modelado Optimizado, Trompeta)

Calculamos la energía de la secuencia original, de la secuencia modelada (energía determinista) y de la secuencia residuo (energía estocástica):

- Energía de la secuencia original: **4.7274e+011**
- Energía de la secuencia modelada: **4.5538e+011**
- Energía del residuo: **9.1555e+009**

7.1.7 – Conclusiones al estudio comparativo:

Recopilamos las energías obtenidas para las secuencias resultantes de los distintos tipos de modelado:

Modelado Clásico P= 6:

- Energía secuencia original: **4.7274e+011**
- Energía secuencia modelada: **4.6056e+011**
- Energía residuo: **7.2492e+009**

Modelado Clásico P=3:

- Energía secuencia modelada: **4.4597e+011**
- Energía residuo: **1.8750e+010**

Modelado Clásico P=2:

- Energía secuencia modelada: **4.1361e+011**
- Energía residuo: **2.7559e+010**

Modelado Clásico P=1:

- Energía secuencia modelada: **3.8689e+011**
- Energía residuo: **3.8808e+010**

Modelado Clásico Breaks=5:

- Energía secuencia modelada: **4.6226e+011**
- Energía residuo: **6.1669e+009**

Nuevo Modelado Optimizado:

- Energía secuencia modelada: **4.5538e+011**
- Energía residuo: **9.1555e+009**

Se observa como en el **modelado polinomial clásico**, a medida que baja el grado del polinomio, el residuo obtenido es mayor. Cuanto más baja el grado del polinomio mayor Subfit se produce, la secuencia modelada tiene menos energía y el residuo tiene más componente determinista (más energía determinista) y la separación es peor. El caso extremo se tiene con **P=1** donde su residuo tiene mucha energía determinista.

Escuchando los residuos podemos darnos cuenta de este fenómeno, el obtenido con **P=6** no tiene parte determinista y a medida que se baja el grado del polinomio, en el residuo obtenido se escucha mayor componente determinista.

El **modelado lineal por tramos** obtiene un **residuo sin parte determinista**, pero al escuchar la secuencia modelada se aprecia cierta **textura robótica** debida a la aproximación mediante funciones triangulares, **no se modela bien la secuencia original**.

El **Nuevo modelado optimizado** obtiene un **residuo con una energía intermedia entre el modelado polinomial clásico de grado 6 y el de grado 3**, pero **sin producirse Overfit ni Subfit**, por lo tanto **optimizamos la extracción del residuo**. Escuchando el **residuo obtenido con el modelado óptimo** se aprecia una **mayor calidad**, **no se detecta componente determinista y se tiene mucha energía extraída**. Escuchando la secuencia modelada (parte determinista) también se aprecia una mejor aproximación.

El **modelado optimizado** obtiene mejoras en cuanto al nivel de extracción de ruido sin producir Overfit ni Subfit por tanto se **extrae más energía ruidosa sin llevarse consigo componente determinista**. Se obtienen una mejor separación entre la componente determinista y estocástica apreciable auditivamente.

8. Conclusiones Finales

En el presente proyecto hemos cumplido con creces los objetivos planteados inicialmente:

- Se ha realizado un estudio y se han obtenido los parámetros y tipo óptimos de modelado no estacionario de cada trama en concreto atendiendo a sus características en cuanto a modulación y ruido.
- Se ha realizado un estudio sobre la posible mejora en cuanto a eliminación del ruido y mitigación del Overfit de aplicar un modelo híbrido de modelado y clasificación de picos conjuntamente, concluyendo que no se obtiene mejora al aplicar modelado sobre la trama ya procesada con clasificación de picos.
- Se ha concluido así mismo que para tramas con fuerte inarmonicidad el modelado no es capaz de aproximar correctamente la parte determinista y se han propuesto soluciones alternativas para este tipo de tramas poco frecuentes.
- Se ha investigado e implementado 3 métodos con los que medir el grado de modulación presente en una trama, siendo uno de ellos totalmente nuevo.
- Se ha diseñado e implementado un método totalmente nuevo para medir el impacto del ruido sobre una trama, basado en un descriptor utilizado en clasificación de picos.
- Se ha implementado un algoritmo nuevo de modelado no estacionario optimizado que obtiene mejores resultados en cuanto a separación de las partes determinista y estocástica, obteniéndose mayor energía estocástica extraída sin producirse Overfit ni Subfit. La mejora en la calidad de la separación de ambas componentes es apreciable auditivamente.

9. Líneas futuras

En el presente proyecto se han estudiado tramas y secuencias en las que se tiene una única fuente armónica. El modelado no estacionario es capaz de separar la parte determinista y la estocástica en secuencias donde aparecen varias fuentes armónicas (polifónicas) pero debería realizarse un estudio sobre el comportamiento real en este tipo de señales.

Todo lo estudiado y descubierto en el presente proyecto puede ser útil en otras áreas relacionadas directamente con el modelado de señales musicales, como la separación de fuentes armónicas en señales polifónicas.

Los métodos obtenidos para el baremo del grado de modulación de una trama destinados a separar las tramas según su grado de modulación o de tonalidad son aplicables en otras áreas de la ingeniería y de la biomedicina donde interese estudiar señales transitorias, o extraer las partes transitorias de las tonales para estudiarlas por separado. Así mismo son aplicables todos los resultados obtenidos en cuanto a modelado y separación del ruido tanto en la energía transitoria como en la armónica para el estudio de cualquier tipo de señal producida por una fuente armónica (motor, corazón, pulmón, etc..).

10. Bibliografía

- [1] Xavier Serra, “Musical Sound Modelling with Sinusoids Plus Noise”, Musical Signal Processing 1997.
- [2] Miroslav Zivanovic and Johan Schoukens, “Time-Variant Harmonic Signal Modeling by using Polynomial Approximation and Fully Automated Spectral Analysis”.
- [3] Miroslav Zivanovic, Axel Röbel and Xavier Rodet, “A New Approach to Spectral Peak Classification”.
- [4] Miroslav Zivanovic, Axel Röbel and Xavier Rodet, “Adaptative Threshold Determination for Spectral Peak Classification”, 2008.
- [5] L. Cohen, “Time-frequency analysis”, Prentice Hall Signal Processing Series, 1995.
- [6] Stéphane Molla and Bruno Torrèsani, “Determining local transientness of audio signals”.