

# APLICACIÓN DE MÉTODOS DE SUBSTRACCIÓN ESPECTRAL PARA REDUCCIÓN DE RUIDO

Alejo Cohen<sup>1</sup> y Benjamín Carisio<sup>2</sup>

<sup>1</sup>Universidad Nacional de Tres de Febrero  
cohen51134@estudiantes.untref.edu.ar

<sup>2</sup>Universidad Nacional de Tres de Febrero  
carisiobenjamin2@gmail.com

**Resumen** — *En el presente informe se presenta el Método de Supresión de Ruido, propuesto por Steven F. Boll, y se analiza la validez de una mejora del mismo que pretende reducir errores del original. Se desarrolló este algoritmo utilizando el lenguaje de programación Python 3.9. Utilizando una señal de habla con ruido, se analizan y comparan los resultados entre ambos métodos (el básico y el alternativo). Además, se emplean distintos parámetros objetivos para evaluar la efectividad de cada uno. Se concluye en que el método alternativo logra una mejora sustancial de la substracción, aunque se deben perfeccionar algunos aspectos de su procesamiento.*

## 1. INTRODUCCIÓN

En muchas aplicaciones resulta útil o hasta indispensable utilizar métodos de reducción de ruido para ciertas señales de audio. Existen diferentes entornos en los cuáles puede existir una inserción de ruido en señales con información valiosa.

Gracias a este tipo de algoritmos de supresión de ruido, se logra mejorar tanto la calidad como la inteligibilidad de la señal, teniendo así una gran importancia en diversas áreas de aplicación relacionadas al Procesamiento Digital de Señales.

### 1.1. Método de Substracción Espectral de Ruido

En el presente informe, se analizará el método de Supresión de Ruido presentado por Steven F. Boll [2], aplicado, en particular, para señales de habla. Este algoritmo, para su implementación, plantea que, dada una determinada señal que es resultado de la suma de una señal "limpia" de habla con una señal de ruido, se deben tener ciertas consideraciones.

- El ruido de fondo es agregado acústicamente (presentándose en el mismo entorno en el que se reproduce la señal de interés) o digitalmente a la señal limpia.

- El ruido de fondo es estacionario, de tal forma que la esperanza de su magnitud espectral  $\mu$  es la misma en todo momento de la señal (con o sin habla).
- Si el entorno en el cuál está presente la señal cambia a un nuevo estado estacionario, existe un intervalo de tiempo suficientemente largo como para poder calcular un nuevo  $\mu$ .
- Para variaciones lentas de estados no estacionarios, el algoritmo requiere un detector de habla para poder calcular el  $\mu$ .
- Es posible una reducción de ruido significativa simplemente quitando el ruido desde la magnitud espectral, es decir, eliminando el aporte del ruido a la señal original (en las frecuencias donde el ruido esté presente).

De esta forma, el procesamiento comienza aplicando a la señal de análisis  $x[n]$  (habla más ruido) un ventaneo con ventana de Hanning, con una superposición de media ventana, de acuerdo al cumplimiento de la restricción COLA (Constant OverLap Add), para la recuperación de la señal temporal filtrada. Luego, se aplica la Transformada Discreta de Fourier (DFT) a cada una de las ventanas, para obtener la información en frecuencia de cada intervalo, y luego, en particular, sus magnitudes.

Una vez hecho esto, se procede a estimar la esperanza

de la magnitud del ruido espectral  $\mu$ , obteniendo dicha información a partir de ventanas en las cuáles no está presente la señal de habla. Por consiguiente, se calcula el Estimador de Substracción Espectral, como la diferencia entre la magnitud de la DFT de la señal de cada frecuencia y la esperanza  $\mu$  del ruido espectral. Además, a este resultado se le asigna el mismo valor de fase  $\theta$  de la señal completa de análisis  $x[n]$ .

$$\widehat{S}(e^{j\omega}) = [|X(e^{j\omega})| - \mu(e^{j\omega})]e^{j\theta X(e^{j\omega})} \quad (1)$$

Donde  $\widehat{S}(e^{j\omega})$  es el estimador,  $|X(e^{j\omega})|$  la magnitud espectral de la señal total,  $e^{j\theta X(e^{j\omega})}$  su fase y  $\mu(e^{j\omega})$  la esperanza de la magnitud espectral del ruido.

Entonces, teniendo como entrada el espectro de la señal  $x[n]$ , y como salida el espectro de la señal filtrada  $\widehat{s}[n]$  (sin ruido), la respuesta  $H(e^{j\omega})$  del sistema de substracción de ruido, será:

$$\widehat{S}(e^{j\omega}) = H(e^{j\omega})X(e^{j\omega}) \quad (2)$$

$$H(e^{j\omega}) = 1 - \frac{\mu(e^{j\omega})}{|X(e^{j\omega})|} \quad (3)$$

Asimismo, si el resultado  $\widehat{S}(e^{j\omega})$  de la substracción para cierta frecuencia es negativo, entonces a la magnitud correspondiente se le asignará un valor igual a cero. Por otra parte, el error espectral  $\epsilon(e^{j\omega})$  de la estimación se entiende como la diferencia entre el espectro de la señal limpia y el de la señal filtrada. A su vez, puede verse como la diferencia entre el espectro de la señal de ruido y su esperanza  $\mu$ .

De aquí en adelante se presentan disintas etapas que forman parte del algoritmo de Substracción Espectral con el objetivo de reducir el mencionado error espectral.

### 1.1.1. Promedio de Magnitud

Se realiza un promedio local de las magnitudes espectrales. Esto corresponde a un promedio, para cada frecuencia, de las magnitudes en todas las ventanas de la señal. Siendo  $\overline{|X(e^{j\omega})|}$  el promedio local mencionado:

$$\overline{|X(e^{j\omega})|} = \frac{1}{M} \sum_{i=0}^{M-1} |X_i(e^{j\omega})| \quad (4)$$

Con  $X_i(e^{j\omega})$  la DFT de la ventana  $i$  y  $M$  la cantidad de ventanas.

Esta maniobra permite llegar a una aproximación del error espectral de forma tal que este sea igual a la diferencia entre el promedio de las magnitudes de cada frecuencia en todas las ventanas de la señal de ruido,  $\overline{N}(e^{j\omega})$  y la esperanza  $\mu$ .

### 1.1.2. Rectificación de Media Onda

Con el objetivo de disminuir el sesgo para la estimación de la magnitud espectral de cada frecuencia  $\omega$ , logrando también disminuir el piso de ruido por  $\mu$ , se implementa una rectificación de media onda. Para cada  $\omega$  cuyo valor de  $|X(e^{j\omega})|$  sea menor a  $\mu$ , se le asignará un valor de magnitud igual a cero. Esto puede conseguirse utilizando una respuesta al impulso  $H_R(e^{j\omega})$  para el sistema del Estimador  $S$ , tal que:

$$H_R(e^{j\omega}) = \frac{H(e^{j\omega}) + |H(e^{j\omega})|}{2} \quad (5)$$

Con  $H_R(e^{j\omega})$  siendo la rectificación de media onda para el sistema estimador  $\widehat{S}$  que tiene a  $H(e^{j\omega})$  como función transferencia.

### 1.1.3. Reducción de Ruido Residual

El ruido residual corresponde al ruido que permanece en la señal luego de ser procesada en la etapa de rectificación de media onda, que tendrá un valor de magnitud que estará entre cero y un valor máximo medido en un intervalo de la señal  $x$  donde no haya habla. Con el objetivo de reducirlo, para las frecuencias cuyos valores de magnitud de  $\widehat{S}(e^{j\omega})$  sean menores al valor máximo mencionado, se reemplaza su amplitud, en la ventana de análisis, por el mínimo valor obtenido de la comparación de la amplitud correspondiente para esa ventana y sus ventanas adyacentes (una anterior y otra posterior).

### 1.1.4. Atenuación Adicional de la Señal en Momentos Sin Habla

Esta última etapa surge a partir de una determinación empírica que establece que entre la potencia promedio en momentos de la señal con habla y en momentos sin habla existe una diferencia de aproximadamente 12 dB. De esta forma, se detectan los momentos de la señal en los que hay ausencia de habla y, por lo tanto, predomina el ruido residual que queda luego de aplicar los procesos ya mencionados. En esos intervalos, para obtener mejor calidad, se considera que la señal no debe eliminarse por completo, ya que si se hiciera, la percepción del ruido de fondo sería mayor en los momentos de presencia de habla. Entonces, se determina que en esos casos un valor de 30 dB de atenuación de ruido es aceptable.

## 1.2. Substracción Espectral Con Sobre-Substracción [3]

Con el objetivo de mejorar el desempeño del algoritmo de Boll presentado en la sección anterior, se propone evaluar un método alternativo que, con un enfoque distinto al original, pretende efectuar una substracción espectral de ruido más criteriosa.

La principal idea de esta mejora es sobre estimar la potencia espectral del ruido y de esta manera, prevenir

que el espectro resultante caiga por debajo de un mínimo espectral.

Este algoritmo se describe a continuación:

$$|\widehat{S}_{ej}(\omega)|^2 = |X_j(\omega)|^2 - |D_e(\omega)|^2 \quad (6)$$

si  $|X_j(\omega)|^2 > (\alpha + \beta)|D_e(\omega)|^2$ ,  
sino  $|\widehat{S}_{ej}(\omega)|^2 = \beta|D_e(\omega)|^2$ .

Donde  $\alpha \geq 1$  es un factor de sobre substracción y  $0 < \beta < 1$  un parámetro del piso espectral. Si  $\alpha$  es muy grande, la señal resultante estará severamente distorsionada e inteligible. Si  $\alpha$  es pequeño, no se realiza la suficiente substracción. Si  $\beta$  es muy chico, el ruido musical (referido a componentes frecuenciales aisladas) se volverá audible, pero el ruido residual será reducido. Si  $\beta$  es muy grande, el ruido musical será reducido, mientras que el residual se volverá audible.

### 1.3. Parámetros de Análisis

Se plantean 4 parámetros para el análisis y comparación de las señales resultantes.

- Zero-Crossing Rate (ZCR) <sup>[1]</sup>: Tasa de cruce por cero, que, para señales de ruido aleatorias, tiende a tener un valor constante y presenta poca desviación. En el caso de las señales de habla con ruido, esta tasa se ve alterada y por lo tanto se la puede analizar fácilmente, ya que se logra distinguir el efecto del ruido. Las señales con ruido presentan un alto valor de ZCR.
- Short Time Energy <sup>[1]</sup>: Se analiza la energía por ventanas en el dominio del tiempo, en el cual nuevamente para el caso del ruido presenta un constante (generalmente de bajo valor), mientras que al sumar la señal de habla, esta energía crece notablemente.
- Relación Señal/Ruido: Esta relación mide la diferencia de amplitudes entre la señal limpia y el ruido. Cuanto mayor sea el valor del parámetro, mayor diferencia entre la señal y el ruido, y por ende mejor calidad de señal. En este trabajo en particular, se comparará la amplitud máxima de la señal con la amplitud máxima del ruido en un intervalo de 1000 muestras al comienzo de la misma (asumiendo que esta no comienza con habla).
- Relación de Energía: Se propone un parámetro en el que se requiere tener la señal original limpia desde un principio. El parámetro consiste en tomar la señal con ruido procesada y restarle la señal limpia, de manera de obtener solo el ruido resultante luego del procesamiento.

$$s(n) - \widehat{s}(n) = N_r(n) \quad (7)$$

Siendo  $s(n)$  la señal original limpia,  $\widehat{s}(n)$  la señal procesada y  $N_r(n)$  el ruido restante (en el caso ideal sería cero).

Posteriormente, se calcula la energía de  $N_r(n)$  y se busca qué porcentaje representa de la energía total de la señal de salida  $\widehat{s}(n)$ . Por último, se puede aplicar este mismo cálculo para la señal de entrada, pudiendo finalmente comparar ambos porcentajes y analizar cuan efectivo resulta el algoritmo.

Cabe destacar que cualquier alinealidad que presente el algoritmo (distorsión o modificación de amplitudes de la señal) se verá reflejada en el porcentaje. Esto resulta positivo ya que también caracteriza cuánto se modifica la señal original.

## 2. DESARROLLO EXPERIMENTAL

Este algoritmo se implementó en el lenguaje de programación Python 3.9, utilizando el IDE Spyder. Además, se hizo uso de las librerías numpy, scipy, matplotlib, soundfile, tanto para el desarrollo de las funciones propias del algoritmo, como para la presentación y análisis de resultados.

Como práctica general, se procuró la optimización del software, a partir de la implementación de las herramientas mencionadas evitando, por ejemplo, el uso excesivo de ciclos for. A su vez, se comentó, paso a paso, la función de cada línea para una mejor comprensión y seguimiento del código a la hora de ser interpretado.

Asimismo, para una mejor organización del código, se desarrollaron las funciones del algoritmo en un archivo “*func.py*” y, por otro lado, estas fueron implementadas en un archivo separado “*main.py*”, donde también se realiza la presentación de resultados.

Con respecto a la ejecución del código para la obtención de resultados, se realizó con una señal de una frecuencia de muestreo de 48000 Hz, con una ventana de 50 ms.

## 3. ANÁLISIS Y RESULTADOS

A continuación, se realiza un análisis y comparación de los resultados obtenidos para los métodos de substracción de ruido implementados, a partir de los parámetros de análisis mencionados anteriormente. Vale aclarar que se dispuso de 3 archivos de audio .wav, de diferentes características tímbricas y frecuenciales, cada uno con su correspondiente señal limpia (sin ruido) y señal con ruido agregado. Sin embargo, se presentarán los resultados de solo uno de ellos, dado que el mismo se trata de una señal de voz humana, propósito principal del algoritmo en cuestión.

### 3.1. Zero-Crossing Rate

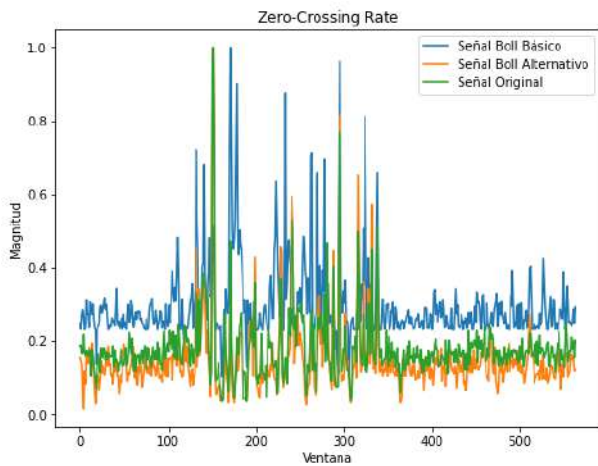


Figura 1: ZCR de las señales de análisis.

Se logra ver con facilidad en la Figura 1 que en la señal procesada con el algoritmo base de Boll, el valor medio del ruido es mayor que en la original. También, se deduce que en el mismo prevalecen componentes de alta frecuencia (lo que implica mayores cruces por el cero). Esto se puede apreciar auditivamente en los archivos de audio generados en el código. Sin embargo, gracias a la mejora aplicada, el algoritmo con el método alternativo logra una substracción exitosa del ruido, disminuyendo levemente el ZCR con respecto a la señal original en momentos sin habla.

### 3.2. Short Time Energy

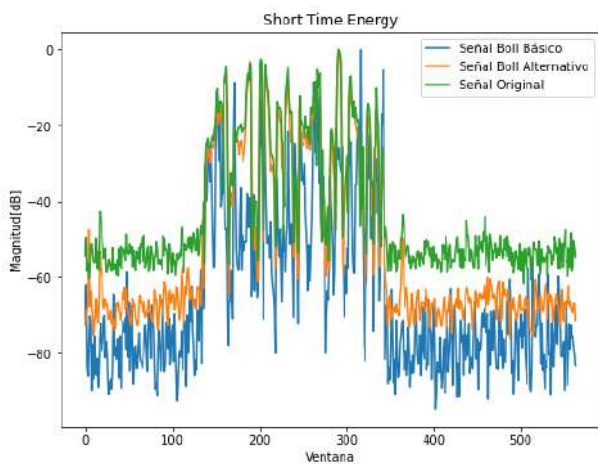


Figura 2: ZCR de las señales de análisis.

En este caso, ambos algoritmos demostraron una mejora en la substracción (Figura 2), siendo el algoritmo base el de mejor resultado (debido a sus valores más bajos

en momentos de la señal sin habla). Sin embargo, el algoritmo alternativo muestra un valor mas estable que el base (con menos fluctuaciones en dichos momentos).

### 3.3. Relación Señal/Ruido

Para poder visualizar fácilmente este parámetro, se cuenta con un gráfico de las formas de onda de cada señal obtenida con el fin de poder comparar la amplitud de los pisos de ruido en el dominio del tiempo.

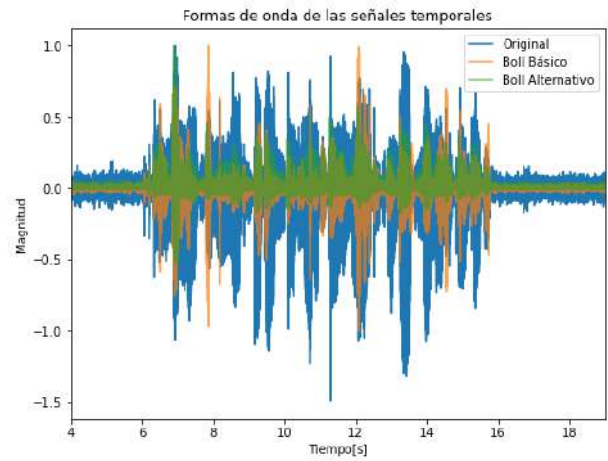


Figura 3: Formas de onda de las señales procesadas y original.

Es muy evidente en la Figura 3 cómo el algoritmo logra reducir el ruido para los momentos donde no hay habla. En la Tabla 1, se presentan los valores del S/R para cada caso:

Tabla 1: S/R calculado para cada señal obtenida.

Señal	S/R
Original	12.0
Boll Básico	38.5
Boll Alternativo	24.6

Ampliando el análisis de la forma de onda (Figura 4), se logra ver el tipo de ruido generado en la señal de salida del algoritmo alternativo. Es posible que este ruido sea generado por el ventaneo de la señal, ya que su período coincide con el largo de cada superposición de ventanas (en este caso, 25 ms).

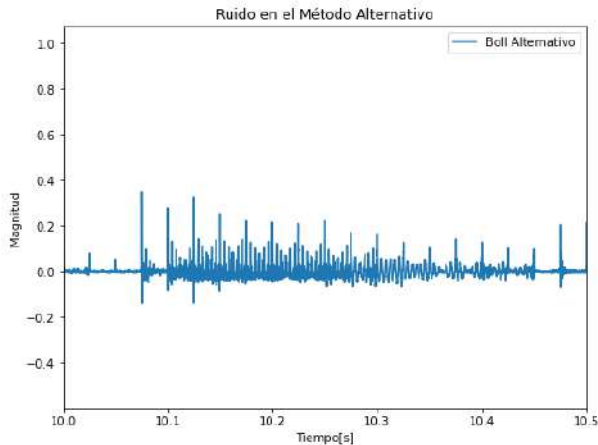


Figura 4: Fragmento de señal de salida del algoritmo alternativo (entre los momentos 10 s y 10.5 s).

### 3.4. Relación de Energía

Los resultados obtenidos relacionados al parámetro de Relación de Energía se presentan en la Tabla 2. Los altos valores de porcentaje para los métodos de filtrado pueden deberse a diferencias de fase (y de amplitud) entre la señal filtrada y la señal limpia original, ya que al no coincidir en ese aspecto, la resta de esta última a la señal procesada no se efectúa correctamente, generando “nuevos” valores que se sumarán en el cálculo de la energía. De esta manera, no se permite detectar cuánta energía relacionada al ruido fue filtrada. Por otro lado, sí se logra cuantificar la distorsión y alinealidades producidas por el procesamiento; se observa que el método de Boll Alternativo es el que genera la mayor distorsión de la señal original.

Tabla 2: Porcentaje de la energía total del ruido resultante para cada señal.

Señal	Porcentaje (%)
Original	7.333
Boll Básico	96.72
Boll Alternativo	216.4

## 4. CONCLUSIONES

En primer lugar, se concluye en que el método de Boll alternativo propuesto resulta en una mejora objetiva del método base.

En segundo lugar, resultó muy conveniente utilizar un promedio por frecuencia ( $\mu$ ) y no un promedio escalar fijo de magnitud espectral. De esta manera, se logró caracterizar mejor al ruido y llevar a cabo así una substracción más exitosa. Por lo tanto, se podría aplicar esta variable al algoritmo base.

Por otro lado, el parámetro de relación de energías resulta muy sensible a distorsiones de fase, ya que si esta

se altera mucho durante el procesamiento, este parámetro podría estar agregando energía y elevando el porcentaje por demás. Es por esto que se propone realizar una mejora al parámetro; forzar a la fase de la señal filtrada con la señal limpia original. De este modo, se podrá caracterizar y evaluar mejor la substracción en sí.

Por último, siendo el algoritmo alternativo el que mejor resultados presenta, se buscaría la manera de reducir el ruido producido por el ventaneo mencionado en la sección de análisis del S/R, para así aumentar su efectividad.

## REFERENCIAS

- [1] *Multifeature speech/music discrimination based on mid-term level statistics and supervised classifiers.*
- [2] *Suppression of Acoustic Noise in Speech Using Spectral Subtraction.*
- [3] *Different Approaches of Spectral Subtraction method for Enhancing the Speech Signal in Noisy Environments.*