



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**IMPLEMENTACIÓN DE ALGORITMOS
BASADOS EN LA LÓGICA DIFUSA PARA
UN CONTROL AUTOMÁTICO DE
GANANCIA DIGITAL**

TESIS

Que para obtener el título de

Ingeniero Eléctrico Electrónico

P R E S E N T A

Carlos Herrera Luna

DIRECTOR DE TESIS

Dr. Jaime Octavio Guerra Pulido



Ciudad Universitaria, Cd. Mx., 2017

Jurado aprobado:

- Presidente: Dr. Heriberto de Jesús Aguilar Juárez
- Vocal: Dr. Jaime Octavo Guerra Pulido
- Secretario: Ing. Roberto Federico Mandujano Wild
- 1er. Suplente: Dra. Fátima Moumtadi
- 2do. Suplente: Dr. Pablo Roberto Pérez Alcázar

Esta tesis se realizó en la División de Ingeniería Eléctrica de la Facultad de Ingeniería, UNAM en Octubre de 2017.

AGRADECIMIENTOS

Mi mayor agradecimiento es a Dios por brindarme el don de disfrutar cada instante y oportunidad de este camino llamado vida en donde mis principales objetivos son: amar, aprender y servir.

Agradezco también a mis hermanos Tere y Rodrigo quienes han sido motivo para no detenerme, por su apoyo expresado en oraciones, palabras de aliento, bromas, risas y tantos momentos compartidos que sin ellos todo sería diferente.

A mi papi, mi mami, Diana, Angie y Mario por ser el ejemplo de cuál puede ser el camino hacia la superación, por esas llamadas de atención, diferencias en la manera de pensar, pero sobre todo por su cariño y amor que han procurado en mí, también a Rous porque con ella crecí y compartimos muchas alegrías y logros y superamos tristezas y momentos difíciles.

A mis abuelitos que me han tenido presente siempre en sus oraciones y que me regalaron el tesoro más valioso para mí: mi mamá. A todos mis tíos y primos porque, cuando tenemos la oportunidad de vernos, es un gusto poder platicar y recibir palabras de aliento para no desistir y superarme cada día.

De manera especial agradezco mucho a mi gran amigo Eder, a quien desde que lo conocí, me ha transmitido sus conocimientos, consejos, experiencias y también sus gustos, entre ellos, el audio, razón que influyó en mí para tomar la decisión de estudiar una ingeniería que se relacionara con el tema y gracias a la asesoría de mi profesora Edith Zepeda decidí que esa ingeniería fuera eléctrica electrónica.

Gracias también a mis amigos que conocí en EIAK, por compartir esta etapa, ha sido un honor poder coincidir en esta oportunidad de superación, a Iván, Zuriel, Rolfi, Sandra, Vicky, Giovanni, Toño, Ana, Idania, Pacheco, Vero, Dianita, Mayin, Jazz, Kary, Kafer, Wendy, Anaí, Mariana, Annie, Juan, Ximena, aún faltan muchos por mencionar pero de todos tengo los mejores recuerdos compartiendo vivencias de nuestra vida.

Agradezco a EIAK en especial por aquellos amigos con los he tenido la oportunidad de convivir y crear un lazo de amistad más fuerte: Daniel, Gaby D., Ricardo, Lupita, Omar, Gaby J., Adriana y Ale; es un honor poder contar con ustedes y una fortuna poder ser parte de sus vidas, gracias por estar hombro a hombro dando lo mejor de nosotros para buscar una nueva y mejor versión personal. Incluyo en este agradecimiento también a los señores Socorro, Martha y Ernesto, quienes han tenido una atención especial conmigo y no tengo palabras para explicar el cariño y aprecio que tengo hacia ustedes.

De la misma manera a Gabriela y Fernando, papás de mi amiga Fery, quienes han estado al pendiente de mi trabajo de tesis, entre otros logros, siempre estaré agradecido por esas atenciones hacia mí.

Así como Fery se convirtió en una amiga muy especial de la universidad, también agradezco su amistad a Eder, Aldo, Abraham, Arturo, Fanny, Tamara, Majo y Abán, ha sido un honor compartir esta inolvidable etapa, gracias por ser parte de esta etapa y por todas las vivencias que tuvimos, también incluyen mis agradecimientos a Dianita, Carmen, Kary, y mis amigas desde la preparatoria Stephanie y Vianey.

Ahora es momento de incluir en mis agradecimientos a mis amigos que no dejaron de impulsarme para no desistir y seguir esforzándome hasta concluir este ciclo: Said, Sam, Monti, Sol y Quique, acepto que si no hubiera sido por su insistencia, todo habría sido diferente. Gracias a Aline y Raúl que los conocí también durante este pedazo de camino llamado tesis, es una fortuna tener su compañía en este momento de mi vida, así como también de los señores Juan José y Rocío.

Siempre estaré agradecido con todo el equipo de ADDER, por el apoyo brindado y los permisos para poder concluir este trabajo, a mi profesor Roberto Mandujano por ver algo en mí que, aún no descubro que fue pero, me permitió ser parte de la empresa Cibernética y Electrónica.

Finalmente, no porque tenga menos importancia, sino porque todo debe cerrar con broche de oro. Mi mayor agradecimiento es para mi mamá, quien ha sido la persona que siempre ha apostado todo por mí, a ella le doy las gracias por su amor y confianza que me ha impulsado a esforzarme, a buscar una mejor versión de mi persona, y que cuando he sentido que ya no puedo o quiero rendirme basta con traerla a mi pensamientos para inspirarme y dar el máximo esfuerzo, ella me ha demostrado con hechos que “imposible” es sólo una palabra que limita la mente pero es desplazada por la fe y el empeño en nuestras acciones. Muchas gracias mamá por dejarme sentir el amor de Dios en tus acciones, siempre serás mi ejemplo a seguir.

Agradezco a todas las personas con quien he coincidido en este camino durante mi vida, y que en algún momento han conocido quien soy y qué estudiaba y al final no fallaron sus mejores y sinceros deseos de éxito.

Desde lo más profundo de mi alma y con toda la energía de mi existir...

¡Muchas gracias!

DEDICATORIA

A mi Madre por brindarme su apoyo incondicional, a la Universidad por otorgarme la oportunidad de ser mi máxima casa de estudios para mi superación académica y a mi director de tesis por su paciencia, diligencia y dedicación para este proyecto.

ÍNDICE

AGRADECIMIENTOS.....	iii
DEDICATORIA.....	vii
ÍNDICE DE FIGURAS	xi
I. INTRODUCCIÓN	1
II. ANTECEDENTES	7
II.1 SONIDO, AUDIO Y RUIDO	7
II.1.1 SONIDO	7
II.1.2 AUDIO	8
II.1.3 RUIDO	17
II.2 SISTEMAS DIFUSOS	20
II.2.1 LÓGICA DIFUSA.....	20
II.2.2 OPERACIONES DE CONJUNTOS DIFUSOS	23
II.2.3 FUNCIONES DE PERTENENCIA	25
II.2.4 MÁQUINA DE INFERENCIA DIFUSA.....	25
II.3 CONTROL DE GANANCIA AUTOMÁTICA	32
II.4 DETECCIÓN DE ACTIVIDAD DE VOZ (VAD, VOICE ACTIVITY DETECTION).....	35
III. METODOLOGÍA.....	39
III.1 DEFINICIÓN DEL PROBLEMA	39
III.2 PROYECTO GENERAL.....	39
III.3 BLOQUE DE CONTROL AUTOMÁTICO DE GANANCIA	40
PARÁMETROS	40
DISEÑO DEL CONTROLADOR	41
III.4 BLOQUE DE DETECCIÓN DE ACTIVIDAD DE AUDIO.....	47
PARÁMETROS	47
DISEÑO DEL CONTROL.....	48
IV. RESULTADOS.....	59
IV.1 BLOQUE DE DETECCIÓN DE ACTIVIDAD DE AUDIO.....	59
IV.2 BLOQUE DE CONTROL DE GANANCIA AUTOMÁTICA	60
IV.3 ACOPLAMIENTO DE AMBOS BLOQUES.....	76
V. CONCLUSIONES	85
VI. REFERENCIAS	89

ÍNDICE DE FIGURAS

II. ANTECEDENTES

II.01	Densidad espectral de potencia para el ruido blanco	17
II.02	Densidad espectral de potencia para el ruido rosa	18
II.03	Densidad espectral de potencia para el ruido café	19
II.04	Estructura de máquina de inferencia difusa	26
II.05	Esquema de neurona artificial	31
II.06	Configuración de un AGC con a) realimentación y con b) prealimentación	33
II.07	Diagrama de bloques de un AGC con lazo de realimentación	33
II.08	Diagrama de bloques de un AGC digital	34

III. METODOLOGÍA

III.01	Diferencia normalizada LBE Ruido Blanco vs Audio	49
III.02	Diferencia normalizada LBE Ruido Rosa vs Audio	49
III.03	Diferencia normalizada LBE Ruido Café vs Audio	50
III.04	Diferencia normalizada MBE Ruido Blanco vs Audio	50
III.05	Diferencia normalizada MBE Ruido Rosa vs Audio	51
III.06	Diferencia normalizada MBE Ruido Café vs Audio	51
III.07	Diferencia normalizada HBE Ruido Blanco vs Audio	52
III.08	Diferencia normalizada HBE Ruido Rosa vs Audio	52
III.09	Diferencia normalizada HBE Ruido Café vs Audio	53
III.10	Cruce por ceros normalizado Ruido Blanco vs Audio	54
III.11	Cruce por ceros normalizado Ruido Rosa vs Audio	54
III.12	Cruce por ceros normalizado Ruido Café vs Audio	55

IV. RESULTADOS

IV.01	Función de pertenencia del Detector de Actividad de Audio	60
IV.02	Valor RMS Caso 1	61
IV.03	Valor RMS Caso 2	61
IV.04	Valor RMS Caso 3	62

IV.05	Señal error Caso 1	63
IV.06	Señal error Caso 2	63
IV.07	Señal error Caso 3	64
IV.08	Señal derivada del error Caso 1	64
IV.09	Señal derivada del error Caso 2	65
IV.10	Señal derivada del error Caso 3	65
IV.11	Incremento de ganancia Caso 1	66
IV.12	Valor de la ganancia Caso 1	66
IV.13	Incremento de ganancia Caso 2	67
IV.14	Valor de la ganancia Caso 2	67
IV.15	Incremento de ganancia Caso 3	68
IV.16	Valor de la ganancia Caso 3	68
IV.17	Valor RMS Caso 4	69
IV.18	Valor RMS Caso 5	70
IV.19	Valor RMS Caso 6	70
IV.20	Señal error Caso 4	71
IV.21	Señal error Caso 5	71
IV.22	Señal error Caso 6	72
IV.23	Señal derivada del error Caso 4	72
IV.24	Señal derivada del error Caso 5	73
IV.25	Señal derivada del error Caso 6	73
IV.26	Incremento de ganancia Caso 4	74
IV.27	Valor de la ganancia Caso 4	74
IV.28	Incremento de ganancia Caso 5	75
IV.29	Valor de la ganancia Caso 5	75
IV.30	Incremento de ganancia Caso 6	76
IV.31	Valor de la ganancia Caso 6	76
IV.32	Función de pertenencia de detección de actividad de audio del bloque AAD del proyecto final	77

IV.33	Valor RMS de entrada y salida para la el archivo de campana tibetana del bloque AGC del proyecto final	78
IV.34	Valor RMS de entrada y salida para el archivo de música comercial del bloque AGC del proyecto final	78
IV.35	Prueba de proyecto completo: Función de Detección de Actividad de Audio	80
IV.36	Prueba de proyecto completo: Valor de la ganancia	82
IV.37	Prueba de proyecto completo: Valor RMS de entrada y salida	83

I. INTRODUCCIÓN

Entre las principales funciones de los sistemas de audio está reproducir, grabar, amplificar y procesar señales de audio, entendiendo éstas como señales sonoras que son agradables al ser humano y, se encuentran en el rango de ancho de banda de 20 a 20,000 Hz. El procesamiento que se lleva a cabo en este tipo de sistemas es, por ejemplo: la ecualización y el control de volumen.

Una de las problemáticas que se tienen en los sistemas de audio es la interferencia que existe cuando se sintoniza una señal, por ejemplo, de radio o televisión, ya sea por otras señales o por ruido del ambiente. La consecuencia de este problema se manifiesta como sonidos incómodos, que también pudieran dañar los equipos. De igual manera, al cambiar de estación o canal cambia el nivel de volumen en el que están grabadas las diferentes pistas de audio de manera repentina, lo que es también incómodo debido a que hay que ajustar nuevamente el nivel deseado. Actualmente existen programas que utilizan algoritmos para mantener el mismo nivel de volumen durante la reproducción de las pistas que fueron adquiridas con niveles diferentes entre cada una de ellas [1]. Algunas aplicaciones comerciales que usan este tipo de algoritmos son: Spotify, iTunes, entre otras. Sin embargo, estas aplicaciones no son capaces de discernir entre el ruido y el audio.

Existen también algoritmos que tienen como finalidad la detección de actividad de voz; siendo una de sus aplicaciones la transmisión digital inalámbrica, donde son utilizados durante las llamadas telefónicas y donde se aplica un protocolo para el envío de la información, tal que únicamente cuando existe actividad de voz se hace transmisión de datos. De esta manera se evita transmitir ruido o información que es inútil a los usuarios [2].

Considerando lo expuesto previamente, se plantea la siguiente pregunta: ¿Qué sucedería si los algoritmos que controlan el nivel de volumen de las señales de audio, tomaran en cuenta la identificación de audio en las señales adicionadas con ruido o de señales no deseadas? Podría ser que al sintonizar una estación de radio donde se transmite un concierto, o en la televisión al sintonizar un partido de futbol o una película, una vez que el algoritmo identificara presencia de ruido, o silencio, nombrando a estos episodios como “inactividad de audio”, éste permitiera que el controlador de volumen modificara el nivel del sonido para que no se escuchara, por tratarse de una señal indeseada. El resultado sería que el usuario además de poder definir un nivel de volumen y que éste permaneciera independiente de la estación, canal o señal que se sintonice, también tendría la tranquilidad de que cuando existiera una interferencia que afectara de manera importante la recepción, el nivel cambiaría para que no se escuchara y, además, el equipo de audio quedara protegido de los cambios repentinos en el volumen de las pistas y que, en un caso grave, pudieran dañarlo.

Este proyecto tiene como objetivo desarrollar un algoritmo que permita identificar la actividad de audio en una señal y, además, conservar un nivel deseado definido por el usuario a través de un control automático de ganancia con el cual al momento de no existir actividad de audio, ya sea por presencia de ruido o una señal que por su amplitud pueda considerarse como un silencio, la ganancia del sistema disminuya para que el ruido no se amplifique.

Este algoritmo que procesa señales de audio digitales, específicamente en formato WAV, se divide en dos etapas: el bloque de detección de actividad de audio y, el segundo, un control de ganancia automática. Ambos bloques utilizan controles difusos implementados a través de máquinas de inferencia difusa para determinar la detección de actividad de audio y el incremento de la variable que controla la ganancia, respectivamente.

La manera en que trabaja el algoritmo es por medio de un vector formado por varias colecciones de muestras, que permitan analizar la detección de actividad de audio para cada colección, calculando el valor cuadrático medio (RMS) de dicho conjunto, debido a que la identificación de actividad de audio es más más efectiva cuando se realiza con una colección de pocas muestras; en cambio, el control de ganancia se hace con un conjunto de varias colecciones ya que, de esta manera, el valor RMS al tener un número mayor de muestras, tiene un valor más constante sin variaciones abruptas.

El bloque de detección de actividad de audio opera con cuatro variables de entrada, las cuales son: energía normalizada en banda baja, energía normalizada en banda media, energía normalizada en banda alta y el número de cruces por cero que tiene la señal en un periodo de tiempo. Las energías normalizadas se utilizan para ser independientes del nivel de volumen que tiene la señal a procesar. Este proceso fue implementado a través de un sistema de inferencia de Sugeno, cuyas funciones de pertenencia fueron obtenidas por medio de un método de optimización basado en redes neuronales. Como resultado se obtiene a la salida el valor de pertenencia de la actividad de audio que tiene la señal, siendo el valor 0 designado para una actividad de audio nula, significando que existe ruido en la señal, o es un periodo de silencio; y el valor de 1 que significa que la señal es completamente audio y no tiene o no es posible detectar interferencia de ruido.

El bloque de control automático de ganancia usa el valor de pertenencia de actividad de audio, calculado con el bloque anterior, para determinar cuándo debe amplificar o atenuar la señal. Además, éste considera el nivel RMS a lo largo del tiempo para determinar los niveles de ganancia para cada conjunto de muestras. Este bloque está realimentado por lo que calcula el nivel RMS de la señal a la salida para que en conjunto con el nivel de ganancia deseado se obtenga la señal de error y, también, su derivada; ésta última representa el cambio de nivel de volumen que va teniendo la

señal a través del tiempo. Con estas variables se determina, por medio de un controlador incremental difuso [3] y usando el método de inferencia de Mamdani, el incremento o decremento de la ganancia por la cual se multiplica la ventana de muestras de la señal de salida. Dependiendo de los casos que ocurran en el bloque; por ejemplo, si el error es positivo que representa que el nivel RMS de la señal de salida es menor al nivel de volumen deseado y además la derivada del error es negativa, entonces, significa que el nivel está disminuyendo; esto tendrá como resultado que el incremento que se le dé a la ganancia sea positivo para que el nivel de volumen de la señal aumente.

Las ventajas que ofrece este algoritmo son que, como amplificador de audio, mantiene un nivel de volumen deseado por el usuario es decir, independientemente de la señal que sea procesada, respeta el nivel al que se desea escuchar; a diferencia de otros amplificadores en donde el usuario define directamente la ganancia, tienen como consecuencia que se proporcione la misma amplificación a diferentes señales, pudiendo tener éstas un nivel de volumen diferente ya sea desde el momento en que se grabaron, o porque se les hizo alguna modificación posterior.

Otra ventaja que ofrece el algoritmo es que la detección de actividad de audio y el efecto que tiene en el control de ganancia se hace sin tener que recurrir a una ecualización, en donde primero se deben identificar las frecuencias en las que se encuentra el ruido para que sean suprimidas por un filtro supresor de bandas; además de que, considerando que al suprimir estas frecuencias no se garantiza que aparezcan nuevas interferencias o ruido en la señal de salida y, también, la ecualización de una señal digital requiere de algoritmos computacionalmente más exhaustivos debido a que se involucra el parámetro de la frecuencia.

En el siguiente capítulo del presente escrito se abordan conceptos y temas requeridos para el diseño del algoritmo; posteriormente, en el capítulo de la metodología, se

explica cada uno de los dos bloques principales que conforman el sistema y se muestran las ecuaciones que se utilizan en cada cálculo; luego, en el capítulo de resultados, se reporta el comportamiento que tiene cada uno de los bloques trabajando por separado o en conjunto; y además, se determina la eficiencia del bloque de detección de audio. Finalmente, en las conclusiones se discute sobre los resultados obtenidos, sus características más importantes y se proponen mejoras que se pueden hacer o probar para obtener un mejor desempeño.

II. ANTECEDENTES

II.1 SONIDO, AUDIO Y RUIDO

II.1.1 SONIDO

El sonido es la percepción de nuestro cerebro de ondas mecánicas, que son una perturbación propagándose en un medio, por ejemplo el aire. En este caso, la perturbación se da en las moléculas del aire, cuando un objeto o cuerpo entra en vibración, y viaja a través de él hasta ser percibidas por nuestro oído, que en un proceso interno, la convierte en impulsos eléctricos para ser interpretada por nuestro cerebro.

La velocidad a la que viaja el sonido en el aire, a una temperatura de 21°C, es de 344m/s, pero aumentará o disminuirá 0.6 m/s por cada grado que aumente o disminuya la temperatura.

Por ser una vibración cíclica posee dos propiedades relevantes: la amplitud, que es la presión máxima que alcanza dicha oscilación, y ésta se mide comúnmente en decibeles [dB] tomando como presión de referencia en el aire 20 μ Pa. La otra propiedad es la frecuencia, que es la cantidad de perturbaciones por segundo que contiene la onda sonora, la cual se caracteriza de la manera siguiente: sonidos graves contienen frecuencias bajas, mientras que frecuencias altas pertenecen a sonidos agudos; su unidad es el Hertz [Hz].

El sistema auditivo del ser humano tiene ciertas limitaciones sobre la amplitud y la frecuencia del sonido. La amplitud máxima que el oído del ser humano puede llegar a tolerar sin sufrir una lesión es de hasta 120 dB, sin embargo una conversación a un metro de distancia llega a tener una amplitud de 70 dB [4]. Sobre la frecuencia, el rango audible está entre los 20 y 20,000 Hz, pero puede variar según cada persona con la edad; a medida que el ser humano envejece, va disminuyendo este rango;

también, la percepción cambia dependiendo de la amplitud, necesitando las frecuencias bajas una mayor amplitud a diferencia de las frecuencias altas. Además, los sonidos con frecuencias menores a 20 Hz se les clasifican como infrasonidos, y por encima de los 20 kHz son ultrasonidos.

II.1.2 AUDIO

Para procesar los diferentes tipos de sonido son necesarios los sistemas de audio los cuales se definen como la combinación de sistemas locales compatibles para poder ser integrados con otros sistemas locales o en red, con la finalidad de dar algún tratamiento, como puede ser grabación, reproducción, transmisión o mejora en la calidad del sonido [4]; algunos ejemplos pueden ser:

- Sistemas de reforzamiento: normalmente caracterizados por brindar amplificación en tiempo real a una señal sonora en un entorno acústico.
- Sistemas de reproducción de sonido: estos sistemas reproducen y amplifican sonido desde un sistema de almacenamiento o desde un origen distante, por ejemplo, transmisión de radio.
- Sistemas sintetizadores: La fuente de sonido de un sintetizador no es humana y permite alterar un sonido mientras se graba o también modificarlo posteriormente.
- Sistemas de almacenamiento: Estos sistemas permiten grabar el sonido.
- Sistemas de medición: Analizan características del sonido en diferentes ambientes.
- Sistemas de control: Permiten configurar diversas características en un sistema sonoro.
- Sistemas de comunicación: Tienen como objetivo la transmisión y la recepción de señales sonoras.

Los sistemas de audio pueden ser analógicos o digitales, siendo la diferencia principal entre éstos el tipo de señal con la que se trabaja: una señal analógica varía con el

tiempo en forma continua y puede tomar cualquier valor, mientras que las señales digitales cambian en cantidades discretas, es decir se muestrea con un determinado periodo y los valores que puede tomar ya están definidos.

En los sistemas de audio analógicos se le da tratamiento a una señal eléctrica, por lo que por medio de circuitos eléctricos y electrónicos se puede conseguir:

- Modificar la ganancia para amplificar o reducir la señal
- Ecuilibrar la señal
- Distorsionar la señal
- Transmitir la señal
- Grabar la señal

Del mismo modo, se pueden lograr los mismos objetivos con los sistemas digitales al muestrear las señales análogas, con la diferencia que éstos procesan y realizan operaciones mediante dígitos que representan a las señales por números binarios.

La diferencia entre los sistemas digitales y analógicos es principalmente en el manejo de dispositivos electrónicos: mientras que en la electrónica analógica el almacenamiento, manipulación, y recuperación de información requieren de mayor número de componentes en comparación con los que se utilizan en la electrónica digital, ya que un conjunto de componentes de electrónica analógica pueden ser reemplazados con algún dispositivo microprocesador o microcontrolador además, gracias a estos componentes, el tiempo que tardan los procesos en los sistemas digitales es menor.

Ahora se hablará sobre algunos procesos que se pueden hacer para caracterizar señales digitales, los cuales permitirán obtener parámetros que pueden ser utilizados por algoritmos con un propósito definido. A continuación se mencionan algunos.

VALOR PROMEDIO

Para una función periódica discreta, el valor promedio se obtiene con la sumatoria de todos los valores instantáneos dividida entre el número total de muestras:

$$X_{prom} = \frac{1}{N} \sum_{i=1}^N x(i) \quad (II, 01)$$

VALOR RMS

El valor medio cuadrático o RMS (root mean square) es la media cuadrática de los valores instantáneos en una colección de N muestras y está dado por:

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N |x(i)|^2} \quad (II, 02)$$

CONVOLUCIÓN

Es un operador matemático que transforma dos funciones f y g en una tercera función, lo que representa esta tercera función es la magnitud en la que se superponen f y una versión trasladada e invertida de g .

A continuación se describe la fórmula para secuencias de duración finita [5].

Teorema: Dada la secuencia de duración finita f , de longitud N_f y cuyo dominio está en el intervalo $n \in [0, N_f]$. Dada también la secuencia g , de longitud N_g y cuyo dominio está en el intervalo $n \in [0, N_g]$. La convolución queda representada en el dominio discreto como:

$$f(n) * g(n) = \sum_{m=0}^{N_f} f(m)g(-(m - n)); \forall n \in [0, N_f + N_g - 2] \quad (II, 03)$$

CORRELACIÓN Y AUTOCORRELACIÓN

Es una operación binaria entre dos funciones que indican la fuerza y dirección de la relación entre ambas funciones [5].

Sean dos funciones reales e variable entera $f(n)$ y $g(n)$, la correlación de ambas funciones denotada por $f(n)**g(n)$ o H_{fg} , ambas en el intervalo $n \in [0, N-1]$ es otra función denotada como $f**g(n)$. Se define:

$$H_{fg} = f ** g(n) = \sum_{m=0}^{N-1} f(m)g(m-n); \forall n \in [-(N-1), (N-1)] \quad (II, 04)$$

En el procesamiento digital de señales se usa la correlación para cuantificar la similitud entre dos señales, y permitir la detección e identificación de señales. [5]

La correlación de una señal consigo misma se denomina autocorrelación:

$$H_{ff} = f ** f(n) = \sum_{m=0}^{N-1} f(m)f(m-n); \forall n \in [-(N-1), (N-1)] \quad (II, 05)$$

FILTROS DIGITALES

Un filtro digital puede ser un sistema lineal e invariante en el tiempo discreto, el cual es utilizado para seleccionar un intervalo de frecuencias en una colección de muestras de la señal de entrada. [6]

En procesamiento digital de señales los filtros digitales tienen múltiples aplicaciones, por ejemplo: separar dos señales que han sido mezcladas, remover ruido indeseable, realizar análisis espectral, atenuar o amplificar algunas frecuencias, entre otras. Por ejemplo, se puede implementar un sistema digital para controlar los tonos graves y agudos en un sistema digital de audio.

Entre sus múltiples clasificaciones se mencionan algunas:

- De acuerdo a las frecuencias en bandas de paso:
 - Filtro paso bajas (FPB)
 - Filtro paso altas (FPA)
 - Filtro paso banda (FPBW)
 - Filtro supresor de banda (FSBW)

- De acuerdo a la aproximación polinómica:
 - Butterworth
 - Chebyshev
 - Bessel
- Lineales
 - Filtros FIR (de respuesta finita al impulso)
 - Filtros IIR (de respuesta infinita al impulso)

Un filtro puede estar dentro de varias categorías, por ejemplo un filtro supresor de banda FIR o un filtro paso altas IIR con aproximación Butterworth.

ESPECIFICACIONES DE UN FILTRO DIGITAL

Un filtro ideal no es realizable, debido a que tiene una respuesta infinita al impulso y es no causal, esto quiere decir que durante el diseño de un filtro no es posible obtener respuestas planas en la banda de paso, ni una pendiente infinita en la banda de transición ni características de fase lineal; sin embargo, deben considerarse diversos compromisos para que el filtro resultante tenga la mejor aproximación posible a las especificaciones ideales.

Las especificaciones de un filtro pueden estar dadas en dos formas: absolutas y relativas.

- Especificaciones absolutas: Proveen los requerimientos de la magnitud de la respuesta en frecuencia: banda de paso, banda suprimida, banda de transición, tolerancia o rizo aceptable en la banda de paso y la banda suprimida, frecuencia de corte.
- Especificaciones relativas: proveen los requerimientos de frecuencia en la escala de decibeles: rizo en la banda de paso (R_p), rizo en la banda suprimida (R_s), atenuación de la banda suprimida (A_s), ganancia en la banda de paso (A_p).

FILTROS FIR

En este proyecto se usan filtros FIR, por su estabilidad y sus características de fase lineal, este punto es de suma importancia en aplicaciones como voz y audio. Estos filtros son no recursivos, es decir, la salida de este tipo de filtro sólo depende de la muestra actual de entrada y N-1 retardos de la entrada.

ESTRUCTURA DE LOS FILTROS FIR

La estructura de un filtro digital indica la manera en que se van a efectuar las operaciones matemáticas. Un problema básico en el diseño de filtros es el cálculo de coeficientes que aproximen las características ideales de una respuesta en frecuencia.

A continuación, en la ecuación (II, 06), se aprecia la respuesta en frecuencia de un filtro FIR tiene la característica de ser un sistema no recursivo, lo que le da también estabilidad:

$$H(\omega) = \frac{Y(\omega)}{X(\omega)} = \sum_{k=0}^q h_k e^{-i\omega k} \quad (II, 06)$$

El diseño de los filtros FIR se lleva a cabo mediante la localización de ceros con respecto al círculo unitario en el plano z. Los polos del filtro se encuentran en el origen. Una desventaja de este tipo de filtros es la necesidad de utilizar un orden mayor para lograr pendientes de corte pronunciadas en comparación de los IIR.

La función de transferencia de un filtro FIR tiene la forma:

$$H(z) = h_0 + h_1 z^{-1} + \dots + h_{N-1} z^{-N+1} = \sum_{i=0}^{N-1} h(i) z^{-i} \quad (II, 07)$$

siendo su respuesta al impulso

$$h(n) = \begin{cases} h_i, & 0 \leq i \leq N - 1 \\ 0, & \text{COC} \end{cases} \quad (II, 08)$$

y la ecuación en diferencias:

$$y(n) = h_0x(n) + h_1x(n-1) + \dots + h_{N-1}x(n-N+1) = \sum_{i=0}^{N-1} h(i)x(n-i) \quad (II, 09)$$

La ecuación (II, 09) se obtiene al efectuar una convolución lineal de la entrada $x(n)$ con la respuesta al impulso $h(n)$ del filtro FIR, N es la longitud del filtro y es igual al número de coeficientes del filtro.

DISEÑO DE FILTROS FIR

Entre los métodos para el diseño de filtros FIR con fase lineal, se pueden presentar los siguientes métodos:

- Método de ventanas: este método se basa en acotar la respuesta infinita al impulso de un filtro ideal, su uso puede generar una mejor respuesta en frecuencia con menor número de oscilaciones en la banda de paso y en la banda rechazada. Sin embargo, las transiciones entre ambas bandas no son tan abruptas como las que se obtienen con una ventana rectangular, otra desventaja de este método es que no se pueden diseñar filtros con atenuaciones diferentes en las diferentes bandas.
- Muestreo de frecuencia: se basa en el muestreo de la amplitud deseada en la frecuencia, para el cálculo de los coeficientes de $h(n)$, donde la respuesta en frecuencia deseada $H(\omega)$ es muestreada a razón de $\omega_k=2\pi$ para $k= 0,1,2,\dots, N-1$. La principal desventaja de este método es la falta del control de precisión de las frecuencias críticas.
- Rizado constante (*equiripple*): Es una familia de métodos que se basan en definir la respuesta en frecuencia ideal del filtro y, fijado un orden, obtener los coeficientes que generan la respuesta más aproximada, en particular, los más comunes se basan en la aproximación de Chebyshev [7]. Una de las ventajas que

ofrece este método es que permite un control total de las características del filtro en cuanto a frecuencias y ganancia.

FILTROS IIR

Otros filtros que se usan son los filtros de respuesta infinita al impulso (IIR), también llamados recursivos o autorregresivos de movimiento promedio. Estos constan de una parte que efectúa la suma ponderada de la entrada $x(n)$ y retardos de $x(n-i)$, y la otra que efectúa una suma ponderada de las salidas $y(n-i)$ retrasadas [6]. Este tipo de sistema discreto tiene una respuesta al impulso de carácter recursivo en la que se realimenta con la salida atrasada. Estos filtros pueden emular las características de un filtro analógico, que también tiene una respuesta infinita al impulso [6].

Algunas ventajas que ofrecen los filtros IIR son que debido a su recursividad con pocos coeficientes pueden generar filtros de una pendiente muy pronunciada en la banda de transición, así como también generar respuestas con una buena aproximación al comportamiento de un filtro analógico. También los filtros IIR tienen desventajas, como su posible inestabilidad debida a los errores introducidos en el cálculo de la salida $y(n)$ por la precisión finita con la que se efectúan los cálculos matemáticos, esto puede provocar que se muevan los polos fuera del círculo unitario. Otras desventajas son: debido a que $h(n)$ es infinita, no es posible efectuar la convolución entre la salida y su respuesta al impulso ni tampoco son filtro con una fase lineal en la banda de paso como los filtros FIR.

ESTRUCTURA DE LOS FILTROS IIR

Recordando que la salida de un sistema lineal y discreto puede describirse como la convolución de la entrada con su respuesta al impulso, que para el caso de los filtros IIR es infinita, tenemos:

$$y(n) = x(n) * h(n) = \sum_{i=0}^{\infty} h(i)x(n-i) \quad (II, 10)$$

Para los filtros IIR esta operación no se puede computar porque se necesita un número infinito de coeficientes, sumas y productos y, por lo tanto, para que genere una salida infinita con una entrada impulso, el sistema debe ser recursivo; entonces, para poder realizar un filtro IIR, es necesario una ecuación en diferencias que realmente la salida retardada:

$$y(n) = \sum_{i=0}^q b(i)x(n-i) - \sum_{i=0}^p a(i)y(n-i) \quad (II, 11)$$

Los filtros IIR tienen diferentes estructuras de implementación, considerando la ecuación en diferencias que se utiliza para implementar un filtro IIR y aplicando la transformada Z, se obtiene la función de transferencia $H(z)$ del filtro:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^q b(i)z^{-i}}{\sum_{i=0}^p a(i)z^{-i}} = \frac{b_0 + b_1z^{-1} + \dots + b_qz^{-q}}{a_0 + a_1z^{-1} + \dots + a_pz^{-p}}, a_0 = 1 \quad (II, 12)$$

Donde p es el número de polos, q es el número de ceros de $H(z)$, a_i y b_i son los coeficientes del filtro.

Existen varios métodos para el diseño de filtros IIR, uno de los más utilizados es el de la transformación analógica a digital a través de la transformada bilineal [6]. Entre los métodos de diseño se encuentran:

- Indirectos o transformaciones de filtros analógicos a digitales: Están basados en el desarrollo de funciones de transferencia digitales $H(z)$ a partir de funciones de transferencia analógicas $H(s)$ normalizadas que corresponden a especificaciones de un filtro analógico y a las cuales se les aplican diversas transformaciones como son: la transformada bilineal, transformaciones por impulso invariante, aproximación por diferencias infinitas, transformación Z-matched, entre otras.
- Directos: con estos métodos el diseño se realiza directamente en el plano z para obtener directamente la función de transferencia $H(z)$, entre los diversos métodos destacan: Diseño en el plano z, minimización del error o por fórmulas.

II.1.3 RUIDO

Después de haber tratado sobre los sistemas de audio, cómo es que funcionan y cómo le dan tratamiento al sonido, es conveniente considerar una característica indeseada pero que hay en todos los sistemas: el ruido. Para los sistemas de audio el ruido se puede agregar como señal sonora al momento de grabarse, o como ruido eléctrico efecto de los circuitos. El ruido acústico se percibe como sensaciones que son incómodas para el oído o que no permite escuchar con claridad la señal de interés.

Existe una clasificación del ruido la cuál según su densidad espectral de potencia (PSD), lo clasifica en tres tipos:

- Ruido blanco: La característica de este tipo de ruido es que su PSD es constante, esto quiere decir que todas las frecuencias de sonido tienen la misma potencia, similar a la luz blanca, que contiene todas las longitudes de onda del espectro visible. Este tipo de ruido tiene algunas aplicaciones como puede ser análisis de funciones de transferencia o síntesis de audio.

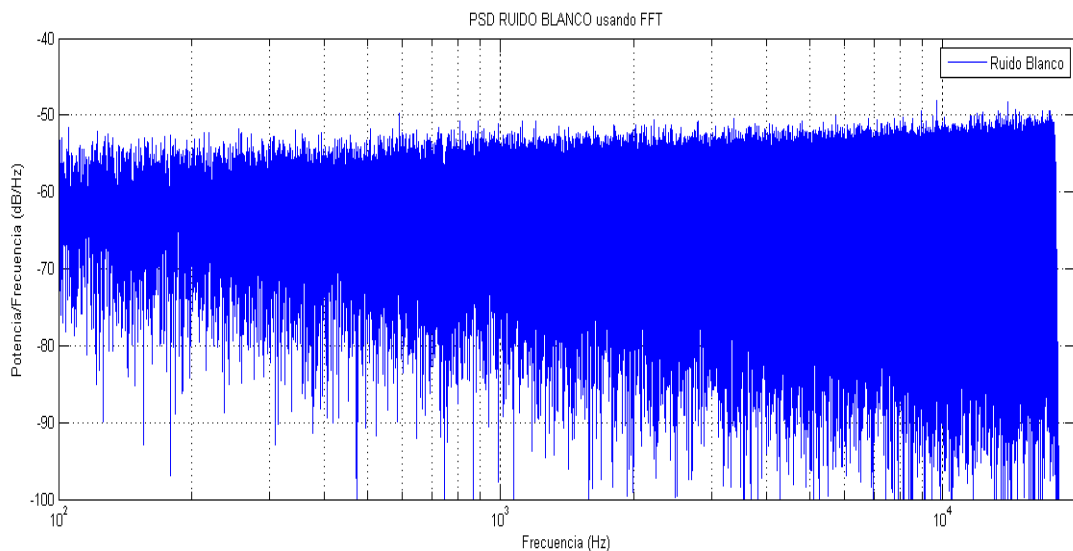


Figura II.01 Densidad espectral de potencia para el ruido blanco

- Ruido rosa: tiene la característica que su PSD es inversamente proporcional a la frecuencia, es decir que tiene una pérdida de 3 dB por octava; sonoramente se escuchan las frecuencias bajas con un nivel mayor en comparación con las frecuencias altas. Se utiliza para analizar el comportamiento de salas, altavoces, equipos de sonido etc. Es una señal conocida en ingeniería de audio, y si lo amplificamos con un altavoz dentro de una sala podemos conocer datos sobre el comportamiento acústico de un altavoz o una sala. Normalmente se genera entre 20 Hz y 20 kHz. Su sonido es muy parecido al que podemos oír cuando se sintoniza entre dos emisoras de FM, en el espacio que se recibe únicamente el ruido, es como un soplido.

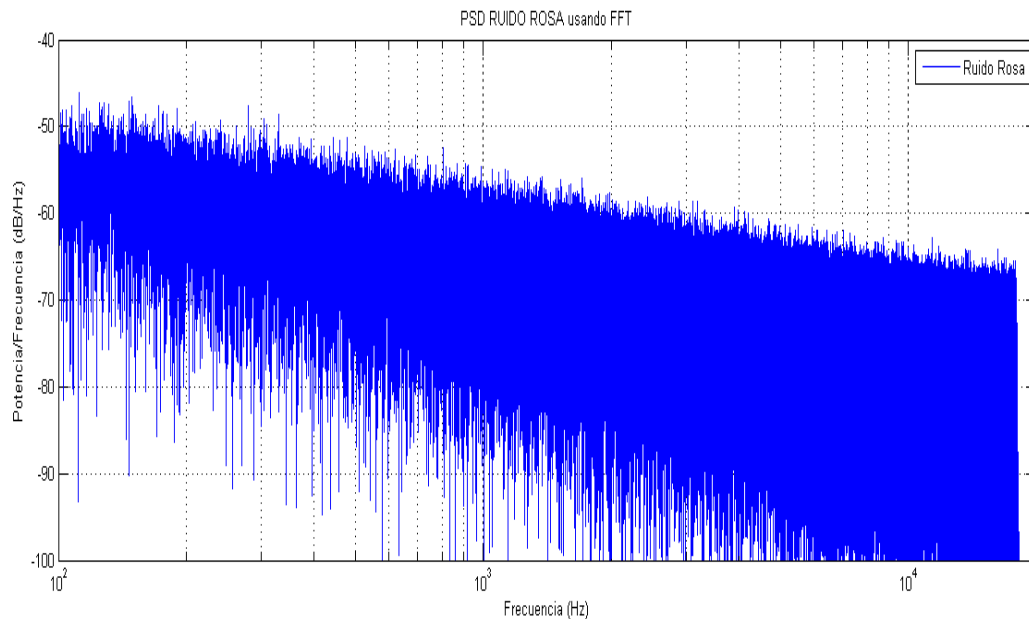


Figura II.02 Densidad espectral de potencia para el ruido rosa

- Ruido café, marrón o rojo: Su PSD es directamente proporcional a $1/f^2$ o dicho de otra forma decae 6 dB por octava a medida que se incrementa la frecuencia. A veces se le designa como "marrón" del inglés "brown" lo que puede llevar a equívoco con su verdadera naturaleza: esto no tiene nada que ver con que su espectro se equipare al color marrón, estando más cercano al rojo, sino con el científico Robert Brown que estudió el movimiento browniano. Este tipo de ruido

puede ser generado por un algoritmo que simule dicho movimiento. Este ruido está compuesto por frecuencias graves y medias, por esta característica su comportamiento es parecido a una señal de audio.

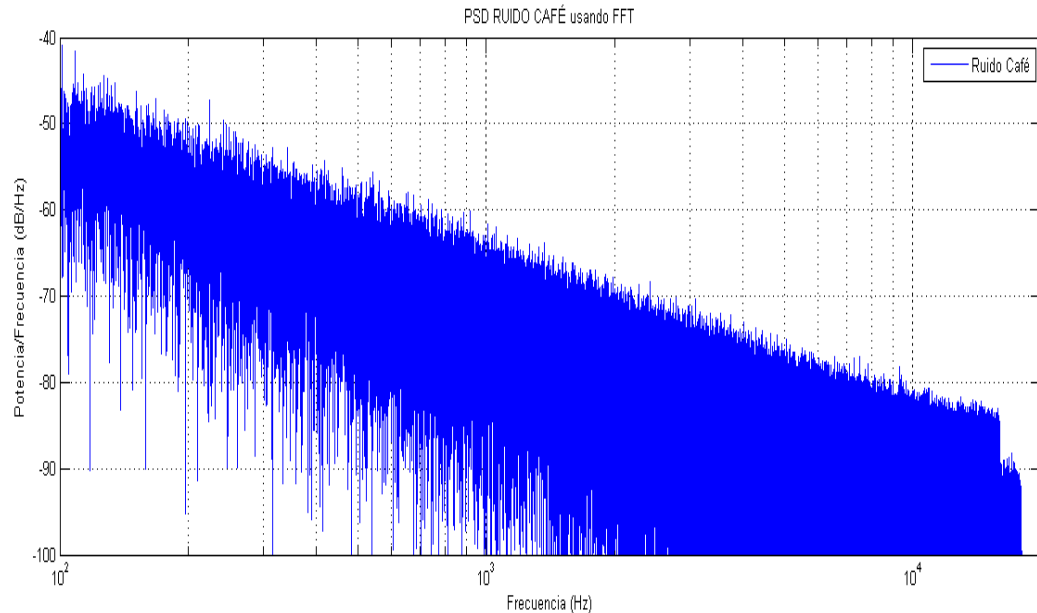


Figura II.03 Densidad espectral de potencia para el ruido café

Otro tipo de ruido es el ruido eléctrico, el cual es una señal eléctrica que se suma a una señal útil a la cual se le quiere dar tratamiento, que en nuestro caso es una señal de audio, y se manifiesta al escucharse en forma de silbidos o zumbidos, los cuales son diferentes al ruido ambiental, tratándose de ruido blanco y rosa.

El ruido eléctrico puede agregarse de dos maneras:

- Por el exterior: otra fuente de señal u otro circuito que haga interferencia en nuestra señal de interés.
- Generado por el propio circuito: la electrónica del circuito puede generar ruidos aleatorios: el ruido térmico, que es generado por la agitación térmica de los electrones, se llega a manifestar como ruido blanco, y comúnmente se genera en

elementos como las resistencias; el ruido de disparo, se genera siempre que una corriente atraviesa una barrera de potencial, por ejemplo, en las uniones semiconductoras pn , por las fluctuaciones de corriente producidas por la generación y difusión aleatorias de portadores de carga; y finalmente el ruido de baja frecuencia, se genera debido a que la materia no es homogénea, y se clasifica como ruido rosa, un ejemplo puede ser el ruido de la base de un transistor BJT [8].

- También, una mala conexión a tierra puede generar ruido.

II.2 SISTEMAS DIFUSOS

II.2.1 LÓGICA DIFUSA

Antes de hablar de los sistemas difusos, es pertinente conocer la lógica difusa, la cual, es un método de razonamiento aproximado, no probabilista, que facilita el modelado de información cualitativa de forma aproximada.

La lógica difusa (LD) es una rama de la lógica multivaluada que permite representar matemáticamente la incertidumbre y la vaguedad, proporciona herramientas para su tratamiento, esto quiere decir, que permite tratar con información imprecisa, con variables de entrada definidas en términos de conjuntos difusos que al combinarse por medio de reglas definen acciones y con uno o varios valores de salida.

La LD permite establecer el análisis de variables de entrada difusa y sus salidas de forma adecuada, atendiendo a criterios de significado, no de precisión.

La LD fue investigada por primera vez a mediados de los años sesenta en la Universidad de Berkeley, California, por el ingeniero Lofti A. Zadeh [9].

La LD permite representar el conocimiento común, que en su mayoría es del tipo lingüístico cualitativo y no necesariamente cuantitativo, por medio de un lenguaje matemático, es decir, la teoría de la LD permite trabajar con datos numéricos y términos lingüísticos al mismo tiempo.

EL PRINCIPIO DE INCOMPATIBILIDAD

Este principio se refiere a que la descripción del comportamiento de un sistema complejo no puede realizarse de forma absolutamente precisa. Zadeh expresa su necesidad de herramientas capaces de manejar de forma rigurosa y fiable información imprecisa, lo que induce al desarrollo de dos aspectos:

- Representación de la información imprecisa: ésta es una manera de ligar las variables de entrada con las variables de salida mediante el empleo de una teoría de conjuntos difusos, así como la descripción de los sistemas complejos y sus relaciones entrada-salida mediante proposiciones condicionales *si-entonces*.
- Inferencia sobre información imprecisa: ante la necesidad de una forma de combinar esta información para obtener nuevos hechos, Zadeh establece la necesidad de un método de inferencia generalizado, conocido como Regla Composicional de Inferencia. A partir de este principio se pueden describir las principales características de la lógica difusa y de los sistemas difusos:
 - El razonamiento exacto, puede verse como un caso particular del razonamiento aproximado en el que cualquier sistema lógico puede ser *fuzzificado* [9]. Mediante la LD se puede formular conocimiento humano de una forma sistemática, y así ser incluido en sistemas de ingeniería.
 - El conocimiento se interpreta como una colección de restricciones difusas sobre una colección de variables. Los sistemas difusos son especialmente interesantes para la definición de sistemas cuyo modelo exacto es difícil de obtener y es necesario introducir una aproximación.
 - La inferencia puede verse como un proceso de propagación de restricciones difusas y su aplicación en los sistemas ayuda a la toma de decisiones, por que la LD permite obtener decisiones con valores incompletos o información incierta.

CONJUNTOS DIFUSOS

Tras el intento de desarrollar una serie de conceptos para tratar de un modo sistemático el tipo de imprecisión que aparece cuando los límites de las clases de objetos no están claramente definidos, podemos definir un *Conjunto Difuso* como una clase en la que hay una progresión gradual desde la pertenencia al conjunto hasta la no pertenencia; es decir, en la que un objeto puede tener un grado de pertenencia definido entre la pertenencia total o no pertenencia. Desde esta perspectiva, los conjuntos convencionales (de la lógica clásica, *crisp*) pueden verse como un caso particular de los conjuntos difusos: un conjunto difuso que sólo admite dos grados de pertenencia (uno y cero).

En la lógica clásica los conjuntos *crisp*, son unos conjuntos C definidos sobre X mediante las *funciones características* de C como f_C .

$$f_C = \begin{cases} 1 & \text{cuando } x \in C \\ 0 & \text{cuando } x \notin C \end{cases} \quad (II, 13)$$

Este conjunto mapea el universo X en un conjunto de dos elementos donde la función $f_C(x)$ es 1 si el elemento x pertenece al conjunto C y 0 si el elemento x no pertenece al conjunto C .

Al generalizar esta función para que los valores asignados a los elementos del conjunto caigan en un rango particular y así indicar el grado de pertenencia de los elementos a ese conjunto, se define la **función de pertenencia** de un determinado conjunto difuso. La función de pertenencia μ_A por la que se define un conjunto difuso A sería:

$$\mu_A = X \rightarrow [0, 1] \quad (II, 14)$$

Donde $\mu_A(x) = 1$ si x está totalmente en A , $\mu_A(x) = 0$ si x no está en A y $0 < \mu_A(x) < 1$ si x está parcialmente en A . Este valor entre 0 y 1 representa el *grado de pertenencia* o *valor de pertenencia* de un elemento x a un conjunto A .

El intervalo de la ecuación anterior está contenido por números reales e incluye los extremos.

II.2.2 OPERACIONES DE CONJUNTOS DIFUSOS

Así como en la lógica clásica existen tres operaciones fundamentales para los conjuntos, las cuales son: UNIÓN, INTERSECCIÓN Y COMPLEMENTO, en la lógica difusa, de manera semejante, existen estas tres operaciones y, adicionalmente, existen las operaciones DIFERENCIA y CONTENCIÓN.

Las operaciones se definen de la siguiente manera:

UNIÓN:

$$\begin{aligned}\mu_{A \cup B}(x) &= \mu_A(x) \vee \mu_B(x) \\ \mu_{A \cup B}(x) &= \text{Max}(\mu_A(x), \mu_B(x))\end{aligned}\quad (II, 15)$$

INTERSECCIÓN:

$$\begin{aligned}\mu_{A \cap B}(x) &= \mu_A(x) \wedge \mu_B(x) \\ \mu_{A \cap B}(x) &= \text{min}(\mu_A(x), \mu_B(x))\end{aligned}\quad (II, 16)$$

COMPLEMENTO:

$$\mu_{\bar{A}}(x) = 1 - \mu_A(x) \quad (II, 17)$$

DIFERENCIA:

$$\mu_{A/B}(x) = \mu_{A \cap \bar{B}}(x) = \text{min}(\mu_A(x), \mu_{\bar{B}}(x)) = \text{min}(\mu_A(x), 1 - \mu_B(x)) \quad (II, 18)$$

CONTENCIÓN:

$$\text{Si } \mu_A(x) \leq \mu_B(x), \text{ entonces } \mu_A(x) \subseteq \mu_B(x) \quad (II, 19)$$

PROPIEDADES DE LOS CONJUNTOS DIFUSOS

Al igual que los conjuntos crisp, los conjuntos difusos tienen las mismas propiedades:

CONMUTATIVA:

$$\mu_{A \cap B}(x) = \mu_{B \cap A}(x) = \mu_B(x) \cap \mu_A(x) \quad (II, 20)$$

ASOCIATIVA:

$$\mu_{A \cup B}(x) \cup \mu_C(x) = \mu_A(x) \cup \mu_{B \cup C}(x) \quad (II, 21)$$

DISTRIBUTIVA:

$$\mu_A(x) \cup \mu_{B \cap C}(x) = \mu_{A \cup B}(x) \cap \mu_{A \cup C}(x) \quad (II, 22)$$

IDEMPOTENCIA:

$$\mu_{A \cup A}(x) = \mu_A(x) \text{ y } \mu_{A \cap A}(x) = \mu_A(x) \quad (II, 23)$$

INVOLUCIÓN

$$\overline{\mu_A(x)} = \mu_A(x) \quad (II, 24)$$

TRANSITIVA:

$$\text{If } (\mu_A(x) \subset \mu_B(x)) \cap (\mu_B(x) \subset \mu_C(x)), \text{ then } (\mu_A(x) \subset \mu_C(x)) \quad (II, 25)$$

PRINCIPIO DE MORGAN:

$$\mu_{\overline{A \cap B}}(x) = \mu_{\overline{A}}(x) \cup \mu_{\overline{B}}(x) \text{ y } \mu_{\overline{A \cup B}}(x) = \mu_{\overline{A}}(x) \cap \mu_{\overline{B}}(x) \quad (II, 26)$$

II.2.3 FUNCIONES DE PERTENENCIA

REPRESENTACIÓN DE FUNCIONES DE PERTENENCIA DE CONJUNTOS DIFUSOS

Como se explicó anteriormente sobre los conjuntos difusos, como la clase gradual desde la pertenencia hasta la no pertenencia de una característica en un objeto, esta gradualidad se puede representar de diferentes maneras:

- Funciones rectas que a su vez pueden ser: triangulares, trapezoidales, con sólo una pendiente
- Gaussianas
- Exponenciales
- Asintóticas

La única condición que deben cumplir es que su rango sea entre 1 y 0; sin embargo, en ocasiones estas funciones se pueden modelar por intuición donde no importa la forma de la función pero, en otros casos, se utiliza también el orden jerárquico (mediante encuestas, pruebas sucesivas) y, dependiendo de la distribución obtenida, se modela la función de pertenencia. Algunas otras herramientas pueden ser: redes neuronales, algoritmos genéticos y los modelados físicos.

II.2.4 MÁQUINA DE INFERENCIA DIFUSA

ESTRUCTURA DE MÁQUINA DE INFERENCIA DIFUSA

En la figura I.04 se muestra que la entrada de la máquina de inferencia puede tener n entradas, éstas pasarán al bloque de difusión donde serán transformadas en valores que pueda usar el bloque de inferencia difusa, éste se ayuda con el bloque de reglas y la base de datos que contienen los conjuntos difusos y las reglas, respectivamente. Su diseño y desarrollo provienen del conocimiento del sujeto por su experiencia, finalmente las variables internas pasarán al bloque de desdifusión, donde se obtienen m variables de salida.

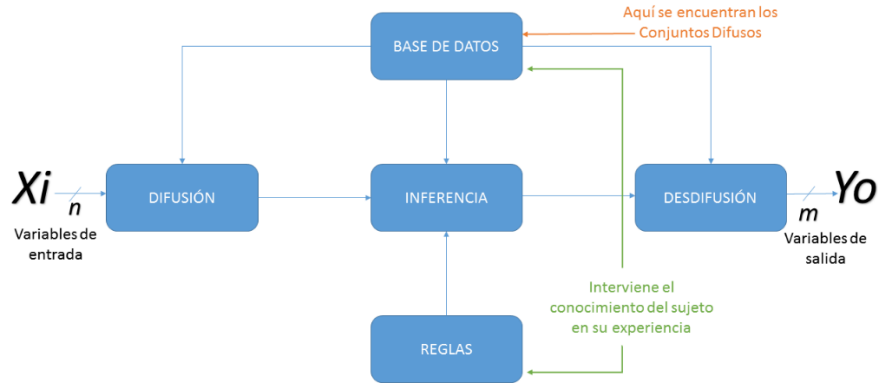


Figura 1.04 Estructura de máquina de inferencia difusa

MÉTODOS GRÁFICOS DE INFERENCIA

Los pasos a seguir en la inferencia difusa son los siguientes:

1. Difusión: es la etapa en donde la variable que entra a la máquina de inferencia es transformada, del dominio en que se encuentra, al dominio de las funciones de pertenencia que serán utilizadas por el controlador.

$$\mu_{A_1^1}(x_{10}) = \begin{cases} f_{A_1}(x), & a_1 \leq x < a_2 \\ f_{A_2}(x), & a_2 \leq x < a_3 \\ f_{A_n}(x), & a_n \leq x < a_{n+1} \\ 0, & \text{COC} \end{cases} \quad (II, 27)$$

2. Cálculo del antecedente total: dependiendo de las reglas establecidas, se aplican los operadores correspondientes mínimos para las operaciones AND, y máximo para los operadores OR.

$$\mu_{A_1} = \min \text{ o } \max(\mu_{A_1^1}(x_{10}), \mu_{A_2^1}(x_{20})) \quad (II, 28)$$

3. Cálculo de la implicación: Es la etapa en donde las funciones de pertenencia de las reglas se mezclan con sus respectivos consecuentes, dependiendo del método que se utilice se sigue un procedimiento determinado.

4. Agregación: en esta etapa se relacionan todos los consecuentes de las reglas, al igual que el paso anterior, dependiendo del método se procede de diferente manera para cada uno de ellos.
5. Desdifusión: Es el proceso en el cuál se transforma la variable difusa de salida, resultado de la máquina de inferencia, a la dimensión del usuario de modo que éste la comprenda. Existen diversos métodos de desdifusión, entre ellos se encuentran:

- Principio de pertenencia máxima: El valor de salida se obtiene encontrando el valor único máximo de la función de pertenencia generada a través del paso de la agregación, es decir, que sólo aparece una vez, ya que de este valor se obtiene el valor z^* que será el valor de salida:

$$\mu_C(z^*) \geq \mu_C(z); z \in Z \quad (II, 29)$$

- Método del promedio-máximo: Cuando la función de pertenencia es trapezoidal, y tiene uno o más catetos, se puede utilizar este método localizando el valor promedio máximo de la función, el valor z^* se obtendrá mediante la fórmula:

$$z^* = \frac{a+b}{2} \quad (II, 30)$$

donde a es el valor inicial y b el valor final del promedio máximo en la función de pertenencia

- Primero o último del máximo: De manera similar al método del promedio máximo, localizando este valor, el valor z^* desfuzificado puede ser:

$$z^* = a; \text{ para el primer valor donde inicia el valor máximo} \quad (II, 31)$$

$$z^* = b; \text{ para el último valor donde termina el valor máximo} \quad (II, 32)$$

- Método del centroide: llamado también primer momento de área, centro de gravedad o centro de área, una vez definida la función de pertenencia se aplica la siguiente fórmula:

$$z^* = \frac{\int z \mu_C(z) dz}{\int \mu_C(z) dz} \quad (II, 33)$$

- Centro del área más grande: cuando la función de pertenencia está dividida en dos o más grandes subregiones convexas, se aplica al centroide de la subregión convexa con el área más grande.
- Método del promedio ponderado: éste método sólo aplica para funciones simétricas, y se obtiene mediante la fórmula:

$$z^* = \frac{\sum_{k=1}^n (\mu_{C_k}(z) \bar{z}_k)}{\sum_{k=1}^n \mu_{C_k}(z)} \quad (II, 34)$$

- Centro de sumas: a diferencia del método anterior no es necesario que las funciones sean simétricas ya que se obtiene el centroide de cada función y se hace un promedio ponderado, por lo que la ecuación [10] para obtener z^* es:

$$z^* = \frac{\sum_{l=1}^k \bar{z}_l (\int \mu_{B_l}(z) dz)}{\sum_{l=1}^k (\int \mu_{B_l}(z) dz)} \quad (II, 35)$$

Para mayor conocimiento de estos métodos de desfusión se puede consultar el capítulo: PROPERTIES OF MEMBERSHIP FUNCTIONS, FUZZIFICATION, AND DEFUZZIFICATION del libro Fuzzy Logic with Engineering Applications, del autor Timothy J. Ross.

Una vez que ya conocemos todos los pasos que se siguen en la inferencia difusa, conoceremos algunos métodos de implicación, de los cuales algunos son:

- Método de Mamdani: existen dos casos, el máx-min, y el max-producto
 - o En el caso máx-min se utiliza el mínimo entre el antecedente total de la regla con su consecuente y para la agregación el máximo entre cada uno de las funciones de implicación de cada regla.

$$\text{Implicación: } \mu_{B_1}(z) = \min(\mu_{A_1}(x_{10}), \mu_{B_1}(y)) \quad (II, 36)$$

$$\text{Agregación: } \mu_C(z) = \max(\mu_{B_1}(z), \mu_{B_2}(z), \mu_{B_3}(z), \dots, \mu_{B_n}(z)) \quad (II, 37)$$

- Para el caso máx-producto el antecedente total se obtiene mediante la multiplicación del antecedente total por el consecuente de la regla, y la agregación se hace mediante el máximo entre cada una de las funciones de implicación de cada regla.

$$\text{Implicación: } \mu_{B_1}(z) = [\mu_{A_1}(x_{10})][\mu_{B_1}(y)] \quad (II, 38)$$

$$\text{Agregación: } \mu_C(z) = \max(\mu_{B_1}(z), \mu_{B_2}(z), \mu_{B_3}(z), \dots, \mu_{B_n}(z)) \quad (II, 39)$$

Ya que se obtiene la agregación, se procede con la desdifusión usando cualquiera de los métodos vistos previamente.

- Método de Sugeno: Para este método, se sigue la siguiente forma:

$$IF \ x \text{ es } A \text{ AND } y \text{ es } B \text{ THEN } z = f(x, y)$$

en donde la función es una constante, la función es de orden 0, si la función es la ecuación de un plano, la función es de orden 1

Se explicará brevemente el método para funciones de orden 1:

$$R1: IF \ x \text{ es } A_1 \text{ AND } y \text{ es } B_1 \text{ THEN } z_1 = p_1x + q_1y + r_1$$

$$R2: IF \ x \text{ es } A_2 \text{ OR } y \text{ es } B_2 \text{ THEN } z_2 = p_2x + q_2y + r_2$$

$$Rn: IF \ x \text{ es } A_n \text{ AND/OR } y \text{ es } B_n \text{ THEN } z_n = p_nx + q_ny + r_n$$

- El antecedente total de cada regla nos servirá para las variables w_1 , w_2 y w_n , dependiendo del número de reglas, que nos servirán para la fórmula para la desdifusión.
- Para la implicación, se obtiene z_1 , z_2 , o z_n con la fórmula definida en cada una de las reglas, éstas se diseñan mediante estadística, algoritmos genéticos, redes neuronales, entre otros métodos.
- Finalmente se hace la agregación y la desdifusión mediante la siguiente fórmula:

$$Z^* = \frac{w_1z_1 + w_2z_2 + \dots + w_nz_n}{w_1 + w_2 + \dots + w_n} \quad (II, 40)$$

Mientras que el método de Sugeno, es computacionalmente más eficaz y, por tanto, es más rápido, garantiza la continuidad de la superficie de salida, mientras que el método de Mamdani es más intuitivo y se ajusta a la forma del pensamiento del ser humano, con la condición de que la continuidad depende del traslapamiento de los conjuntos de entrada.

Existe también el método de Tsukamoto, y es parecido al Método de Mamdani, con la diferencia de que las funciones consecuentes de las reglas son funciones monotónicamente crecientes o decrecientes.

REDES NEURONALES

Como se mencionó previamente, para el diseño de las funciones de implicación, en el Método de Sugeno, se utilizan diversos métodos, entre ellos las redes neuronales. Éstas se definen como un sistema de computación compuesto por elementos simples de procesos interconectados, los cuales procesan información por medio de su estado dinámico como respuesta a entradas externas. [11]

Una neurona artificial es una unidad procesadora con cuatro elementos funcionales:

1. El elemento receptor, a donde llegan una o varias señales de entrada
2. El elemento sumador, efectúa la suma algebraica ponderada de las señales de entrada
3. El elemento de función activadora, aplica una función no lineal de umbral a la salida del sumador para decidir si la neurona se activa
4. El elemento de salida, es el que produce la señal y que constituye la salida de la neurona [12]

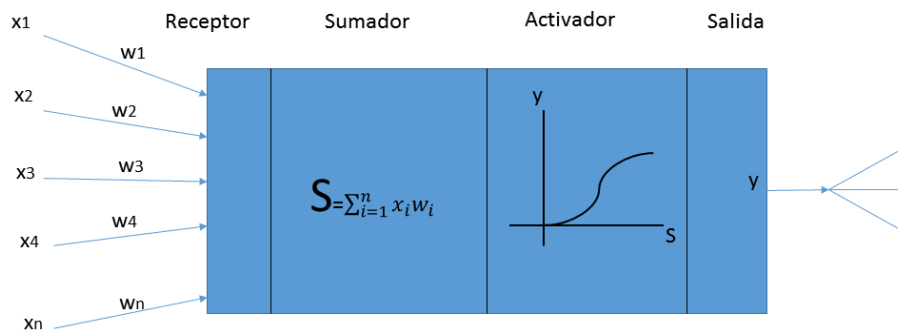


Figura II.05 Esquema de neurona artificial

NIVELES O CAPAS DE UNA RED NEURONAL: La distribución de neuronas dentro de la red se realiza formando niveles o capas, con un número determinado de dichas neuronas entre cada una de ellas. Se distinguen principalmente tres tipos de capas:

- De entrada: es la capa que recibe directamente la información proveniente de fuentes externas a la red.
- Ocultas: Son capas internas a la red y no tienen contacto directo con el entorno exterior.
- De salidas: que se encargan de transferir información de la red hacia el exterior.

TIPOS DE NEURONAS ARTIFICIALES: Además, las neuronas se pueden clasificar de acuerdo a los valores que pueden tomar, se distinguen dos tipos principalmente:

- Neuronas binarias: únicamente pueden tomar los valores $\{0, 1\}$ o $\{-1, 1\}$.
- Neuronas reales: pueden hacerlo dentro del rango $[0, 1]$ o $[-1, 1]$.

El aprendizaje de una red neuronal es el proceso por el cual se modifican sus pesos en respuesta a una información de entrada, pudiendo llegar a tener otros cambios como son destrucción, modificación y creación de conexiones entre las neuronas que

conforman la red. La destrucción de una conexión entre dos neuronas es cuando el peso de su entrada es cero.

Durante la etapa de entrenamiento de una red neuronal se distinguen dos métodos de aprendizaje [11]:

- Aprendizaje supervisado: el entrenamiento es controlado por un agente externo, llamado supervisor maestro, que determina la respuesta que debería generar la red a partir de una entrada determinada, éste, también, controla la salida de la red y, en caso de que ésta no coincida con la deseada, se procede a la modificación de los pesos de las conexiones. A su vez este tipo de aprendizaje distingue tres formas de llevarlo a cabo:
 - Aprendizaje por corrección de error
 - Aprendizaje por refuerzo
 - Aprendizaje estocástico

- Aprendizaje no supervisado: se le conoce también como aprendizaje autosupervisado, no requiere de una influencia externa para ajustar los pesos de las conexiones entre sus neuronas.

II.3 CONTROL DE GANANCIA AUTOMÁTICA

Un Amplificador con Control de Ganancia Automática (de sus siglas en inglés, AGC) realiza una función esencial en diversas aplicaciones modernas en las cuales se deben de procesar señales con un alto rango dinámico [13, 14], tales como: los canales de lectura de una unidad de disco [15], sistemas médicos [16, 17] y multimedia, comunicaciones alámbricas e inalámbricas [13], interfaz de sensores [18], dispositivos de carga acoplada (CCD), sólo por mencionar algunos. La finalidad principal de estos dispositivos es ajustar el nivel de una señal a través de un amplificador de ganancia variable o programable hasta obtener un nivel adecuado según la aplicación o las condiciones en las que está operando el sistema. Por ejemplo, en aquellos sistemas

donde se procesan señales analógicas que luego serán digitalizadas, este dispositivo sirve para ajustar la señal de interés de acuerdo al rango dinámico del ADC para evitar que éste se sature [13]. Existen dos estructuras básicas para los AGC: con realimentación o con prealimentación, éstas se muestran en la figura II.3.01.

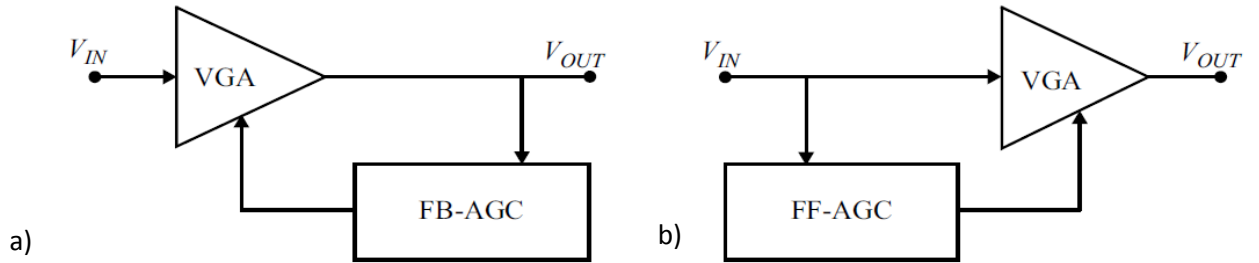


Figura II.06 Configuración de un AGC con a) realimentación y con b) prealimentación

Para la configuración con realimentación, la señal de entrada cambia de amplitud por medio de un amplificador con ganancia variable (VGA). La ganancia del VGA es controlada por el lazo de control formado por un detector, un filtro y una señal de referencia, tal como se muestra en la figura II.07. Para el caso del AGC con prealimentación, el proceso es similar, sin embargo, la señal que se usa para el ajuste de la ganancia es la señal de entrada y no la de la salida. Entonces, para ambos procesos el voltaje a la salida está dado por:

$$V_{out} = G_{VGA} * V_{in} \quad (II, 41)$$

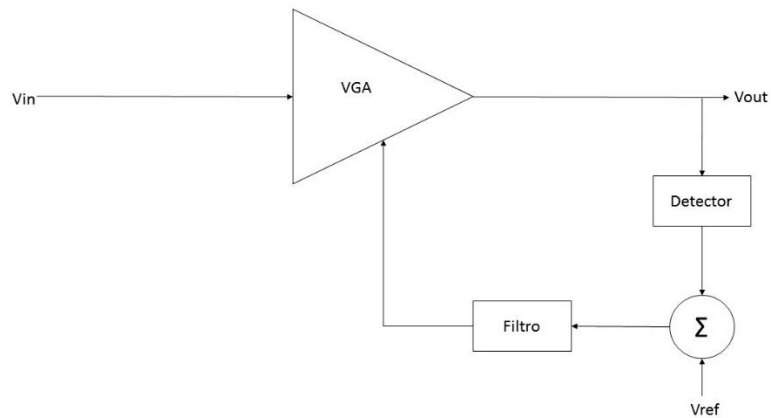


Figura II.07. Diagrama de bloques de un AGC con lazo de realimentación [18]

La teoría de operación de estos sistemas se conoce desde hace muchos años [19], incluso antes de se extendiera el uso de las computadoras digitales. Con la miniaturización de los sistemas digitales y la reducción de sus costos, su uso se ha extendido al procesamiento digital de señales. De este modo, se han implementado algoritmos que operan de una manera muy parecida a como lo haría un AGC analógico, es decir, se han desarrollado AGCs digitales [20]. En la figura II.3.03 se muestra el diagrama de bloques de un AGC digital con realimentación. Este dispositivo opera de la siguiente manera: primero, se toma la muestra de la salida y se estima su amplitud (R'), ésta se compara con la referencia R y, con esta información, se hace un ajuste de la ganancia G para la muestra $X(n+1)$ [20]. Nótese que podría implementarse un AGC digital con prealimentación.

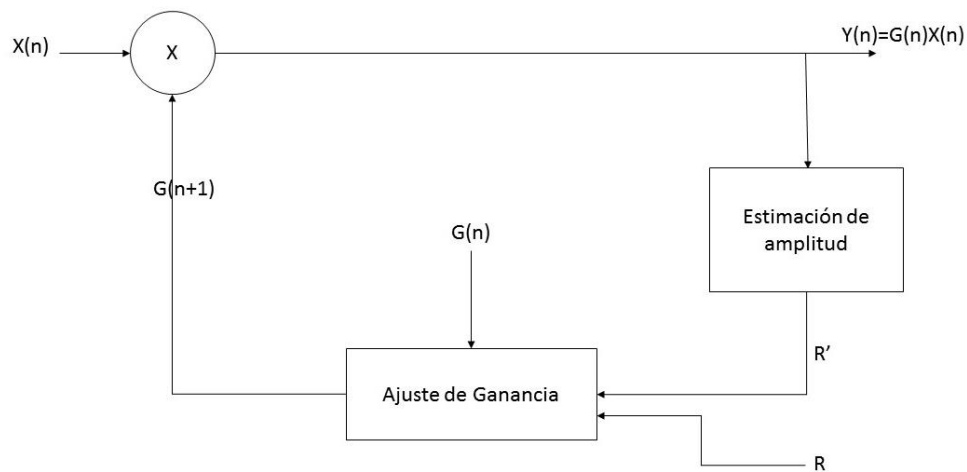


Figura II.08: Diagrama de bloques de un AGC digital [20].

En general, los bloques ubicados en la realimentación de un AGC son dos: el estimador de amplitud y el ajuste de ganancia tal como se puede observar en la figura 3. Se han explorado diversas maneras de implementar estos bloques de control [21, 22].

II.4 DETECCIÓN DE ACTIVIDAD DE VOZ (VAD, VOICE ACTIVITY DETECTION)

La detección de actividad de voz (VAD por sus siglas en inglés), tiene como propósito distinguir acústicamente entre el discurso (habla) y diversos tipos de ruido acústico de fondo, incluso con bajos niveles de SNR (relación señal a ruido). [2] Esto quiere decir que durante el procesamiento se tendrá a la entrada la señal que contiene la mezcla de voz y de ruido acústico y, a la salida, se tendrá una señal donde se hace una atenuación del ruido, para que únicamente sea fácil de identificar la actividad y la inactividad de voz. [23]

Entre sus aplicaciones, se encuentra el campo de comunicaciones multimedia, donde al aplicarse el VAD, se utiliza el modo de transmisión discontinua (DTX) como el que se aplica en el sistema global para comunicaciones móviles (GSM). Como ventajas se obtiene que se reduce la interferencia entre canales, incrementado el número de canales de radio y el consumo de potencia de equipos portátiles. [2]

Existen diversos diseños aplicados en VAD, una de las desventajas de algunos diseños de VAD, es que aún no es lo suficientemente eficiente, especialmente cuando este está operando en condiciones acústicas adversas. [24] A continuación se analizará el diseño estandarizado por la ITU-T VAD por G.729 Anexo B, en donde la tasa de transmisión es de 8 kbit/s, además el código trabaja en dos modos: si la bandera de VAD es igual a uno, se llama al codificador de discurso para codificar o decodificar la señal de voz y, si es cero, se llama al DTX para codificar/decodificar señales de voz inactiva. [2]

Los parámetros principales, trabajando con ventanas de 80 muestras (10 ms, con frecuencia de muestreo de 8kHz) son los siguientes:

- Diferencia de energía en banda completa
- Diferencia de energía en banda baja

- Diferencia de cruces por cero
- Distorsión espectral

Estas cuatro diferencias se obtienen del valor obtenido en tiempo real (muestra de la señal en curso) y del promedio de la señal (que se actualiza únicamente ante presencia de sólo ruido, es decir voz inactiva).

A continuación se muestran los parámetros utilizados para obtener las diferencias obtenidas [23]:

- Frecuencias Espectrales lineales: Se deriva d los primeros 11 términos de la autocorrelación usando el procedimiento de G.729 Anexo A.

$$\{LSF_i\}_{i=1}^p, p = 10 \quad (II, 42)$$

- Energía en banda completa: Es el logaritmo del primer coeficiente de la autocorrelación normalizada.

$$E_f = 10 * \log_{10} \left[\frac{1}{10} R(0) \right] \quad (II, 43)$$

En donde R(0) es el primer coeficiente normalizado de autocorrelación para la colección de muestras

- Energía de banda baja: la señal a analizar se somete a un filtro FIR paso bajas con una frecuencia de corte de 1 kHz:

$$E_l = 10 * \log_{10} \left[\frac{1}{N} h^T R h \right] \quad (II, 44)$$

En donde h es la respuesta al impulso del filtro FIR, y R es la matriz de autocorrelación de Toeplitz.

- Tasa de cruce por cero: se obtiene de la ecuación:

$$ZC = \frac{1}{2M} \sum_{i=0}^{M-1} [|sgn[x(i)] - sgn[x(i - 1)]|] \quad (II, 45)$$

Donde M=80 por el tamaño de la muestra a analizar

El algoritmo tiene principalmente las siguientes reglas: Los valores promedio de los parámetros, únicamente se actualizan sólo en presencia de ruido (voz inactiva). Para las primeras 32 ventanas de muestras, la bandera de VAD es forzada a ser igual a 1 (modo activo) si el parámetro de energía en banda completa es mayor a 15 dB, de igual manera si el número de ventana es igual al número total de ventanas la bandera será forzada a ser igual a cero (modo inactivo).

Inicialmente la decisión de voz activa se hace utilizando el cálculo de la región de decisión de límites múltiples, en el espacio de las cuatro diferencias calculadas, la decisión de voz activa está dada por la unión de las regiones de decisión y la decisión de voz inactiva es su decisión lógica complementaria. La decisión genérica del límite en sus cuatro dimensiones está dada por:

$$IF \Delta P_i < a * \Delta P_k + b THEN Flag = 1 \quad i, k = 1, \dots, 4.$$

En donde ΔP_i y ΔP_k son dos de las cuatro diferencias de los parámetros, a y b son constantes, si ninguna de las 14 condiciones es VERDADERA, la bandera es igual a cero. La decisión final está dada por la consideración de niveles de energía en conjunto con las dos decisiones aplicadas a sus respectivas ventanas anteriores.

III. METODOLOGÍA

III.1 DEFINICIÓN DEL PROBLEMA

Ahora que sabemos que la presencia de ruido en los sistemas de audio es natural e inevitable, procesar una señal de audio, para atenuar el ruido añadido, es de los principales retos a resolver. El presente proyecto tiene como objetivo identificar en una señal de audio, en formato WAV la presencia o ausencia de ruido, definiendo de esta manera, la detección de actividad de audio (AAD, por sus siglas en inglés) y, dependiendo del rango porcentual de esta detección, el sistema asigna una ganancia controlada para que a la salida se obtenga un valor RMS deseado, evitando que una señal se amplifique con ruido o que disminuya el nivel de la señal cuando ésta no está contaminada con señales indeseadas.

III.2 PROYECTO GENERAL

El algoritmo total toma de la señal de entrada, que es un archivo en formato de audio con frecuencia de muestro F_s , muestras de n datos y, también un valor RMS deseado, este valor debe ser mayor a 0 y menor o igual a 1, las muestras serán analizadas por el bloque de Detección de Actividad de Audio, y nos dará el valor de pertenencia a la Actividad de Audio, siendo 1 para aquellos sonidos en los que no se identificó ruido sin ruido en la señal y 0 sonidos identificados como ruido.

Dependiendo de este porcentaje al pasar al bloque de control automático de ganancia y, tomando en cuenta el valor RMS deseado, si el valor de pertenencia de la AAD es mayor a 0.75, entonces, el bloque trabajará para amplificar o reducir la ganancia para que al ser multiplicada por las n muestras y calcular el valor RMS sea igual al valor deseado; pero si el valor de pertenencia está entre 0.5 y 0.75, se multiplicará el valor RMS deseado por el porcentaje de pertenencia para reducir la ganancia y evitar que se amplifique a la salida, debido a que la presencia de ruido en la señal podría ser

molesta al usuario. Finalmente, si el valor de pertenencia es menor a 0.5, no se le dará tratamiento el bloque de Control de ganancia sino que directamente se reducirá el valor de la variable de la ganancia para que al multiplicarse por la muestra, se disminuya el valor RMS a la salida.

III.3 BLOQUE DE CONTROL AUTOMÁTICO DE GANANCIA

PARÁMETROS

En este bloque a la entrada se tiene las n muestras de la señal, el valor RMS deseado (V_{RMSd}), y también se utiliza la frecuencia de muestreo (F_s), con estos valores se calculan el valor RMS de la señal (V_{RMSi}), y en conjunto con el valor RMS deseado se obtendrán los parámetros Error y Derivada del Error para que sean utilizados por el control difuso que a la salida tendrá un incremento de ganancia, que al final se sumará a la ganancia instantánea para que al ser multiplicada por la señal, obtenga el valor RMS a la salida (V_{RMSo}) y sirva para los siguientes cálculos:

- Error normalizado:

$$E_n = \frac{V_{RMSd} - V_{RMSo}(k-1)}{V_{RMSd}} \quad (III, 01)$$

- Derivada del Error normalizado: La derivada de la función Error normalizado, en función del nivel deseado y nivel de RMS a la salida se define como:

$$\frac{\Delta E_n}{\Delta k} \equiv -F_s * \frac{V_{RMSo}(k-1) - V_{RMSo}(k-2)}{V_{RMSd} * n} \quad (III, 01)$$

Este parámetro representa si el nivel RMS está ascendiendo o descendiendo, físicamente se percibe si el volumen aumenta o baja.

A la salida de este bloque se obtendrá el valor de la ganancia que multiplicará a las muestras para obtener el valor RMS deseado.

DISEÑO DEL CONTROLADOR

Una vez que se obtienen estos dos valores, serán utilizados como variables de entrada en un control difuso usando el método de Mamdani:

DIFUSIÓN

Para cada uno de los parámetros se tienen cinco funciones de pertenencia las cuales son:

FUNCIONES DE PERTENENCIA DEL ERROR

- Error negativo máximo: Representa que el nivel RMS a la salida es relativamente muy elevado al valor deseado, se define con la siguiente función:

$$\mu_{EnMax}(E) = \begin{cases} 1, & E < -0.8 \\ -2.5 * E - 1, & -0.8 \leq E < -0.4 \\ 0, & COC \end{cases} \quad (III, 02)$$

- Error negativo mínimo: En esta función el nivel RMS es mayor al deseado, se define:

$$\mu_{Enmin}(E) = \begin{cases} 1.25 * E + 1.25, & -1 \leq E < -0.2 \\ -5 * E, & -0.2 \leq E < 0 \\ 0, & COC \end{cases} \quad (III, 03)$$

- Error nulo: El nivel RMS a la salida es igual al valor deseado, su función de pertenencia se define:

$$\mu_{Enulo}(E) = \begin{cases} 5 * E + 1, & -0.2 \leq E < 0 \\ -5 * E - 1, & 0 \leq E < 0.2 \\ 0, & COC \end{cases} \quad (III, 04)$$

- Error positivo mínimo: El nivel RMS es menor al valor deseado, su función de pertenencia es:

$$\mu_{Epmin}(E) = \begin{cases} 5 * E, & 0 \leq E < 0.2 \\ -1.25 * E + 1.25, & 0.2 \leq E < 1 \\ 0, & COC \end{cases} \quad (III, 05)$$

- Error positivo máximo: El nivel RMS es mucho menor que el valor deseado, se define con la siguiente función:

$$\mu_{EpMax}(E) = \begin{cases} 2.5 * E - 1, & 0.4 \leq E < 0.8 \\ 1, & E \geq 0.8 \\ 0, & COC \end{cases} \quad (III, 06)$$

FUNCIONES DE PERTENENCIA DERIVADA DEL ERROR

Así mismo las funciones de pertenencia que representan el estado del nivel dl valor RMS son las siguientes:

- Nivel desciende máximo: Esta función de pertenencia representa que el valor RMS está disminuyendo de manera muy rápida, se define de la siguiente manera:

$$\mu_{NdMax}(\Delta E) = \begin{cases} 1, & \Delta E < -4 \\ -0.5 * \Delta E - 1, & -4 \leq \Delta E < -2 \\ 0, & COC \end{cases} \quad (III, 07)$$

- Nivel desciende mínimo: El valor RMS está disminuyendo moderadamente:

$$\mu_{Ndmin}(\Delta E) = \begin{cases} 0.25 * E + 1.25, & -5 \leq E < -1 \\ -E, & -1 \leq E < 0 \\ 0, & COC \end{cases} \quad (III, 08)$$

- Nivel mantenido: Significa que el nivel RMS sigue teniendo el mismo valor, su función es:

$$\mu_{Nmant}(\Delta E) = \begin{cases} E + 1, & -1 \leq E < 0 \\ -E + 1, & 0 \leq E < 1 \\ 0, & COC \end{cases} \quad (III, 9)$$

- Nivel asciende mínimo: El nivel RMS está ascendiendo, su función es:

$$\mu_{NaMin}(\Delta E) = \begin{cases} E, & 0 \leq E < 1 \\ -0.25 * E + 1.25, & 1 \leq E < 5 \\ 0, & COC \end{cases} \quad (III, 10)$$

- Nivel asciende máximo: En esta función define que el valor RMS está incrementando su valor de manera muy rápida, se define de la siguiente manera:

$$\mu_{NaMax}(\Delta E) = \begin{cases} 0.5 * E - 1, & 2 \leq E < 4 \\ 1, & E \geq 4 \\ 0, & COC \end{cases} \quad (III, 11)$$

En esta etapa el bloque de ventana asegura que todos los valores sean definidos en alguna función de pertenencia está tomado en cuenta en la definición de las funciones de los valores de frontera en error negativo y positivo máximo y en nivel asciende y descende máximo.

Debido a que en nuestro bloque de control tenemos dos entradas y estas tienen cinco clasificaciones cada una, siguiendo con el método haremos una relación con cada clasificación de la primer entrada: el error, con cada una de la segunda entrada: derivada del error:

1. IF E_{nMax} AND N_{dMax} THEN G_{ds}
2. IF E_{nMax} AND N_{dmin} THEN G_{dmin}
3. IF E_{nMax} AND N_{mant} THEN G_{dmin}
4. IF E_{nMax} AND N_{imin} THEN G_{dMax}
5. IF E_{nMax} AND N_{iMax} THEN G_{dMax}
6. IF E_{nmin} AND N_{dMax} THEN G₀
7. IF E_{nmin} AND N_{dmin} THEN G_{ds}
8. IF E_{nmin} AND N_{mant} THEN G_{ds}
9. IF E_{nmin} AND N_{imin} THEN G_{dmin}

10. IF E_{nmin} AND N_{iMax} THEN G_{dMax}
11. IF E_{nulo} AND N_{dMax} THEN G_{imin}
12. IF E_{nulo} AND N_{dmin} THEN G_{is}
13. IF E_{nulo} AND N_{mant} THEN G_0
14. IF E_{nulo} AND N_{imin} THEN G_{ds}
15. IF E_{nulo} AND N_{iMax} THEN G_{dmin}
16. IF E_{pmin} AND N_{dMax} THEN G_{imin}
17. IF E_{pmin} AND N_{dmin} THEN G_{is}
18. IF E_{pmin} AND N_{mant} THEN G_{is}
19. IF E_{pmin} AND N_{imin} THEN G_0
20. IF E_{pmin} AND N_{iMax} THEN G_{ds}
21. IF E_{pMax} AND N_{dMax} THEN G_{iMax}
22. IF E_{pMax} AND N_{dmin} THEN G_{imin}
23. IF E_{pMax} AND N_{mant} THEN G_{imin}
24. IF E_{pMax} AND N_{imin} THEN G_{is}
25. IF E_{pMax} AND N_{iMax} THEN G_0

Siendo los consecuentes los incrementos de las reglas las clasificaciones de ganancia que son cinco:

- Decremento máximo de ganancia (G_{dMax})
- Decremento mínimo de ganancia (G_{dmin})
- Decremento suave de ganancia (G_{ds})
- Ganancia cero (G_0)
- Incremento suave de ganancia (G_{is})
- Incremento mínimo de ganancia (G_{imin})
- Incremento máximo de ganancia (G_{iMax})

ANTECEDENTE TOTAL E IMPLICACIÓN

Siguiendo con la teoría de los conjuntos difusos, la aplicamos a las reglas para obtener el antecedente total de cada una de las reglas:

1. $ATR01 = \min(E_{nMax}, N_{dMax})$
2. $ATR02 = \min(E_{nMax}, N_{dmin})$
3. $ATR03 = \min(E_{nMax}, N_{mant})$
4. $ATR04 = \min(E_{nMax}, N_{imin})$
5. $ATR05 = \min(E_{nMax}, N_{iMax})$
6. $ATR06 = \min(E_{nmin}, N_{dMax})$
7. $ATR07 = \min(E_{nmin}, N_{dmin})$
8. $ATR08 = \min(E_{nmin}, N_{mant})$
9. $ATR09 = \min(E_{nmin}, N_{imin})$
10. $ATR10 = \min(E_{nmin}, N_{iMax})$
11. $ATR11 = \min(E_{nulo}, N_{dMax})$
12. $ATR12 = \min(E_{nulo}, N_{dmin})$
13. $ATR13 = \min(E_{nulo}, N_{mant})$
14. $ATR14 = \min(E_{nulo}, N_{imin})$
15. $ATR15 = \min(E_{nulo}, N_{iMax})$
16. $ATR16 = \min(E_{pmin}, N_{dMax})$
17. $ATR17 = \min(E_{pmin}, N_{dmin})$
18. $ATR18 = \min(E_{pmin}, N_{mant})$
19. $ATR19 = \min(E_{pmin}, N_{imin})$
20. $ATR20 = \min(E_{pmin}, N_{iMax})$
21. $ATR21 = \min(E_{pMax}, N_{dMax})$
22. $ATR22 = \min(E_{pMax}, N_{dmin})$
23. $ATR23 = \min(E_{pMax}, N_{mant})$
24. $ATR24 = \min(E_{pMax}, N_{imin})$
25. $ATR25 = \min(E_{pMax}, N_{iMax})$

Una vez obtenido los antecedentes totales de las reglas procederemos a la implicación de cada uno de los antecedentes con su respectiva salida:

1. $\mu_{R01} = \min(ATR01, G_{ds})$
2. $\mu_{R02} = \min(ATR02, G_{dmin})$
3. $\mu_{R03} = \min(ATR03, G_{dmin})$
4. $\mu_{R04} = \min(ATR04, G_{dMax})$
5. $\mu_{R05} = \min(ATR05, G_{dMax})$
6. $\mu_{R06} = \min(ATR06, G_0)$
7. $\mu_{R07} = \min(ATR07, G_{ds})$
8. $\mu_{R08} = \min(ATR08, G_{ds})$
9. $\mu_{R09} = \min(ATR09, G_{dmin})$
10. $\mu_{R10} = \min(ATR10, G_{dMax})$
11. $\mu_{R11} = \min(ATR11, G_{imin})$
12. $\mu_{R12} = \min(ATR12, G_{is})$
13. $\mu_{R13} = \min(ATR13, G_0)$
14. $\mu_{R14} = \min(ATR14, G_{ds})$
15. $\mu_{R15} = \min(ATR15, G_{dmin})$
16. $\mu_{R16} = \min(ATR16, G_{imin})$
17. $\mu_{R17} = \min(ATR17, G_{is})$
18. $\mu_{R18} = \min(ATR18, G_{is})$
19. $\mu_{R19} = \min(ATR19, G_0)$
20. $\mu_{R20} = \min(ATR20, G_{ds})$
21. $\mu_{R21} = \min(ATR21, G_{iMax})$
22. $\mu_{R22} = \min(ATR22, G_{imin})$
23. $\mu_{R23} = \min(ATR23, G_{imin})$
24. $\mu_{R24} = \min(ATR24, G_{is})$
25. $\mu_{R25} = \min(ATR25, G_0)$

AGREGACIÓN Y DESDIFUSIÓN

Una vez que se tiene la implicación podemos obtener la agregación:

$$\mu_{AT} = \text{Max}(\mu_{R01}, \mu_{R02}, \mu_{R03}, \mu_{R04}, \mu_{R05}, \mu_{R06}, \mu_{R07}, \mu_{R08}, \mu_{R09}, \mu_{R10}, \mu_{R11}, \mu_{R12}, \\ \mu_{R13}, \mu_{R14}, \mu_{R15}, \mu_{R16}, \mu_{R17}, \mu_{R18}, \mu_{R19}, \mu_{R20}, \mu_{R21}, \mu_{R22}, \mu_{R23}, \mu_{R24}, \mu_{R25})$$

Así, finalmente obtendremos el valor de salida, en este caso el incremento a la variable de la ganancia por medio del método del centroide:

$$inc_G = z^* = \frac{\int z \mu_{AT}(z) dz}{\int \mu_{AT}(z) dz} \quad (III, 12)$$

III.4 BLOQUE DE DETECCIÓN DE ACTIVIDAD DE AUDIO

PARÁMETROS

A la entrada de este bloque se tienen n muestras de la señal, en el cual se calcularán los siguientes parámetros:

- Energía en banda completa (FBE)
- Energía en banda baja (LBE)
- Energía en banda media (MBE)
- Energía en banda alta (HBE)
- Cruce por ceros (ZC)

Una vez teniendo estos parámetros, se calcularán los siguientes parámetros para que sean procesados por un controlador difuso:

- Diferencia normalizada de energía en banda baja, se define con la siguiente ecuación:

$$\Delta LBE_n = \frac{FBE - LBE}{FBE} \quad (III, 13)$$

- Diferencia normalizada de energía en banda media

$$\Delta MBE_n = \frac{FBE - MBE}{FBE} \quad (III, 14)$$

- Diferencia normalizada de energía en banda alta

$$\Delta HBE_n = \frac{FBE - HBE}{FBE} \quad (III, 15)$$

- Cruce por ceros normalizados

$$ZC_n = \frac{ZC}{n} \quad (III, 16)$$

A la salida del controlador se tendrá un valor de pertenencia de Actividad de Audio.

DISEÑO DEL CONTROL

Antes de presentar cómo funciona el controlador del bloque de detección de actividad de audio, es pertinente mostrar cuál es el comportamiento de los tres tipos de ruido acústico y de una señal de audio considerando los parámetros que se definieron previamente, esto con el objetivo de poder entender las reglas del controlador difuso.

Es necesario recordar que cuando se habla de banda completa, baja, media o alta, se entiende que las frecuencias son respectivamente:

- Banda completa: 0 – 22,050 Hz, debido al teorema de Nyquist con una frecuencia de muestreo $F_s=44,100$ Hz que es con la que se trabajaron los archivos .WAV
- Banda baja: 0 – 2,100 Hz
- Banda media: 2,000 – 16,500 Hz
- Banda alta: 16,000 – 22,050 Hz

DIFERENCIA NORMALIZADA DE ENERGÍA EN BANDA BAJA

En las siguientes figuras donde se grafica la diferencia normalizada de energía en banda baja, del tipo de ruido versus la señal de audio, se aprecia cuál es la energía sobrante al aplicar el filtro a la señal y restarle la energía calculada de la señal resultante del filtro, se puede apreciar que la señal de audio, al aplicarle el filtro paso bajas, calcular la energía en la señal obtenida y restarla de la energía en banda

completa, la energía es muy poca, esto es debido a que el ancho de banda común de una señal de audio no suele estar por encima de los 2,100 Hz y, por lo tanto, la mayor parte de la energía se encuentra en la banda baja.

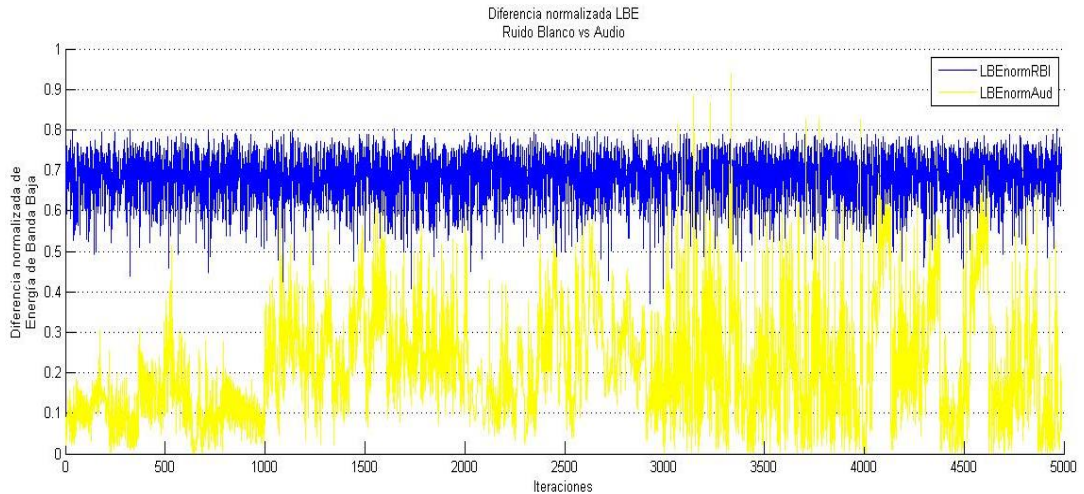


Figura III.01 Diferencia normalizada LBE Ruido Blanco vs Audio

En el caso del ruido rosa, debido a sus características, si pierde energía al aplicar el filtro de banda baja, por lo que la diferencia de energía obtenida puede ser casi igual a la de una señal de audio, así se muestra en la figura III.02

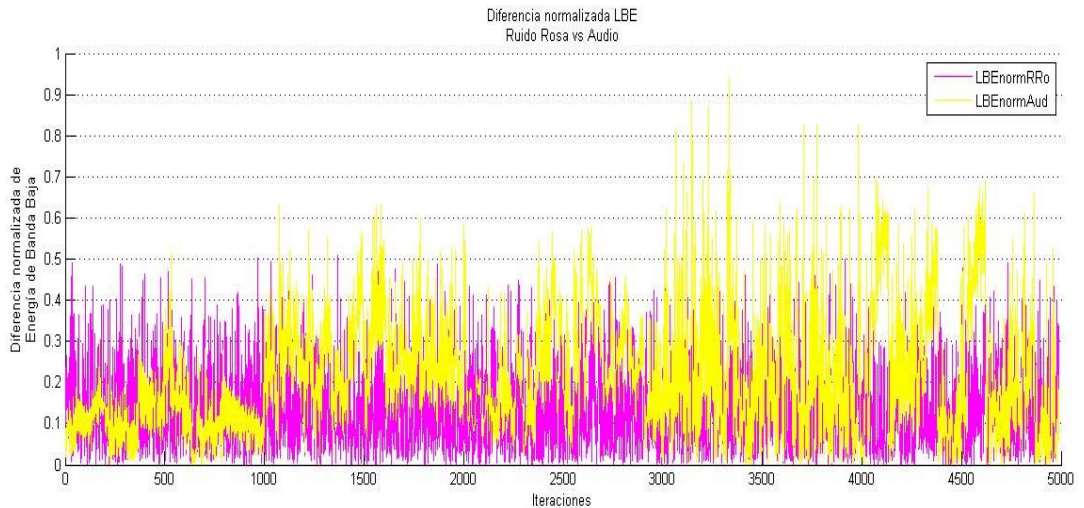


Figura III.02 Diferencia normalizada LBE Ruido Rosa vs Audio

Para la señal de ruido café, que se muestra en la figura III.03, al aplicarle el filtro se observa que pierde casi toda su energía e incluso al obtener el parámetro de la

diferencia de energía es menor que el de la señal de audio, este comportamiento también nos ayudará a discriminar las diferencias entre el ruido café y la señal de audio, que suelen ser parecidas considerando sus espectros en frecuencia.

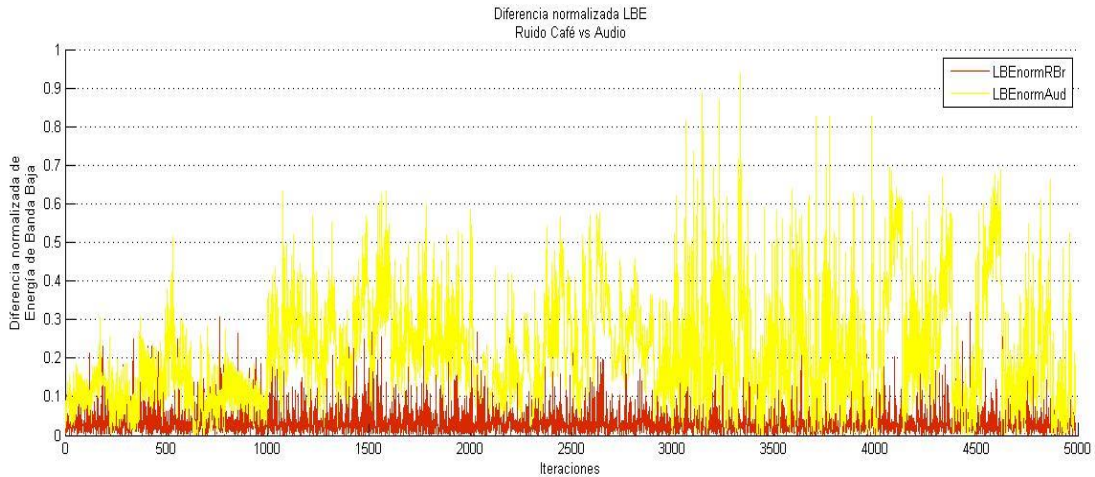


Figura III.03 Diferencia normalizada LBE Ruido Café vs Audio

DIFERENCIA NORMALIZADA DE ENERGÍA EN BANDA MEDIA

En las siguientes figuras (III.04, III.05 y III.06) se aprecia la diferencia de energía aplicando el filtro paso banda, en el caso del ruido blanco, pierde la mayor parte de información, por lo que refleja poca energía, a diferencia del audio, donde sí se percibe, pero es mayor al del ruido blanco.

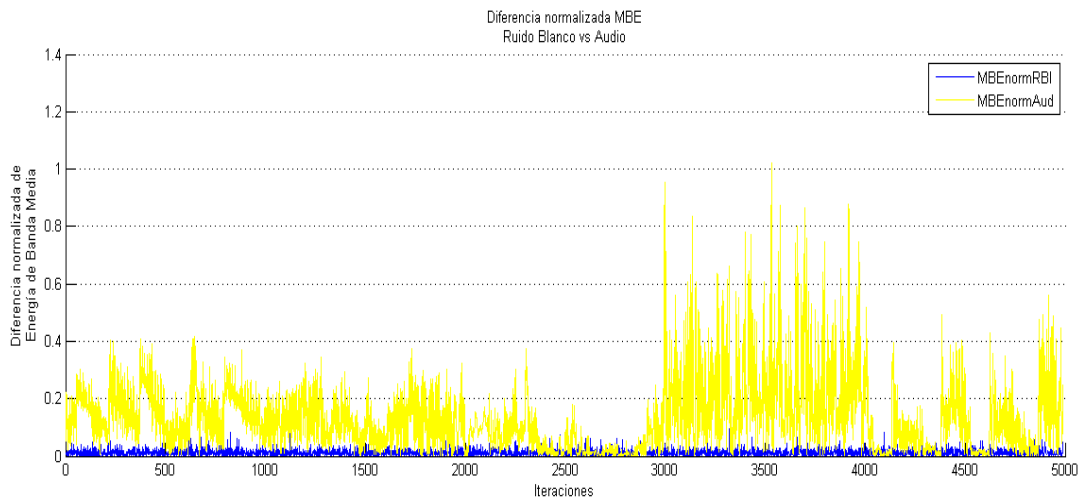


Figura III.04 Diferencia normalizada MBE Ruido Blanco vs Audio

Para el ruido rosa, la diferencia de energía en comparación con el audio, llega a ser parecido al audio, aunque dependiendo del género musical llega a variar la amplitud de la diferencia.

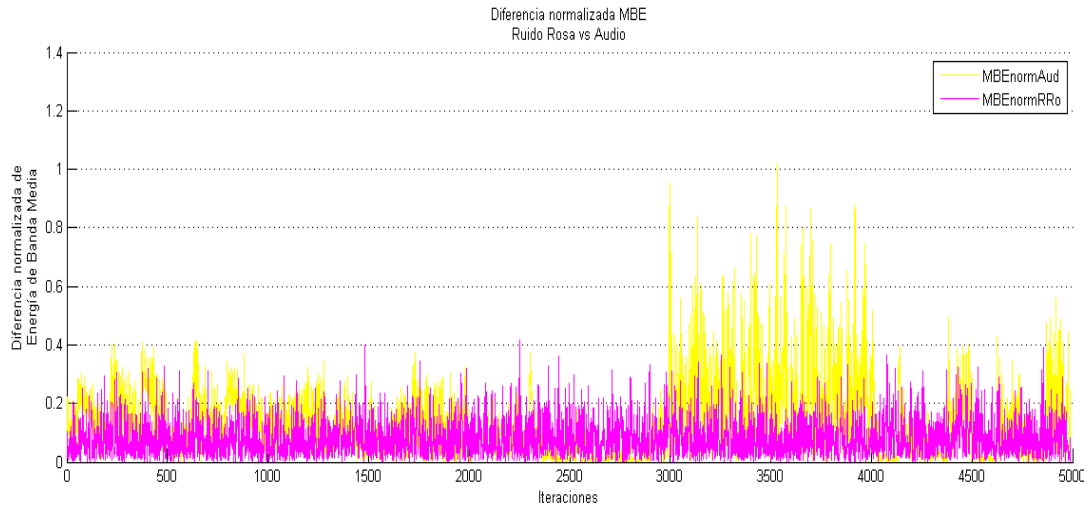


Figura III.05 Diferencia normalizada MBE Ruido Rosa vs Audio

En el caso del ruido café, la diferencia de energía es mayor a la del audio, y se debe al comportamiento del ruido café tomando en cuenta su magnitud con respecto a la frecuencia.

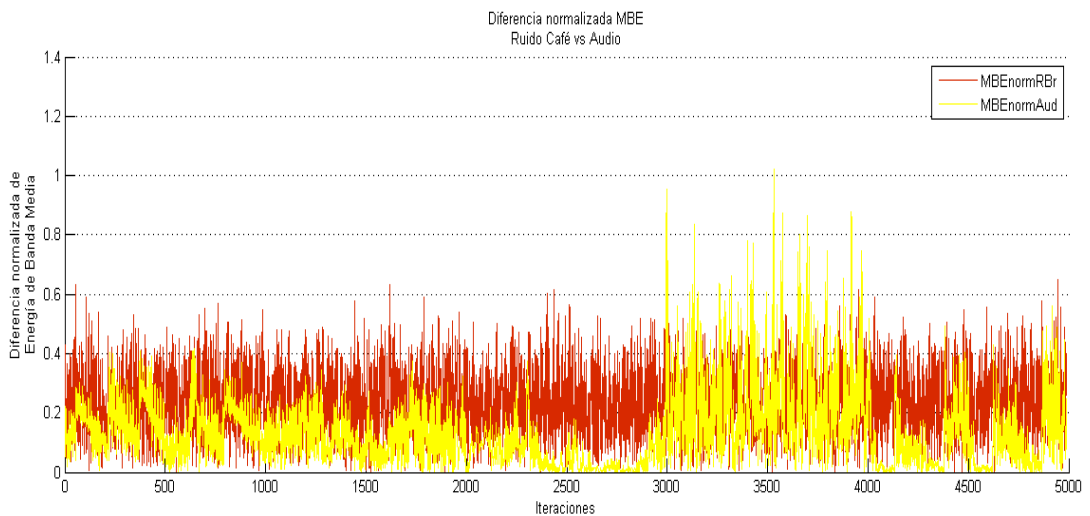


Figura III.06 Diferencia normalizada MBE Ruido Café vs Audio

DIFERENCIA NORMALIZADA DE ENERGÍA EN BANDA ALTA

Finalmente para el caso de la diferencia normalizada de energía en banda alta, donde se nota más la diferencia es en la señal de ruido blanco, (figura III.07), ya que al aplicar el filtro, la señal pierde la mitad de su información, mientras que a la señal de audio, no le afecta, como se mencionaba antes la mayor parte de su información se encuentra en frecuencias más bajas.

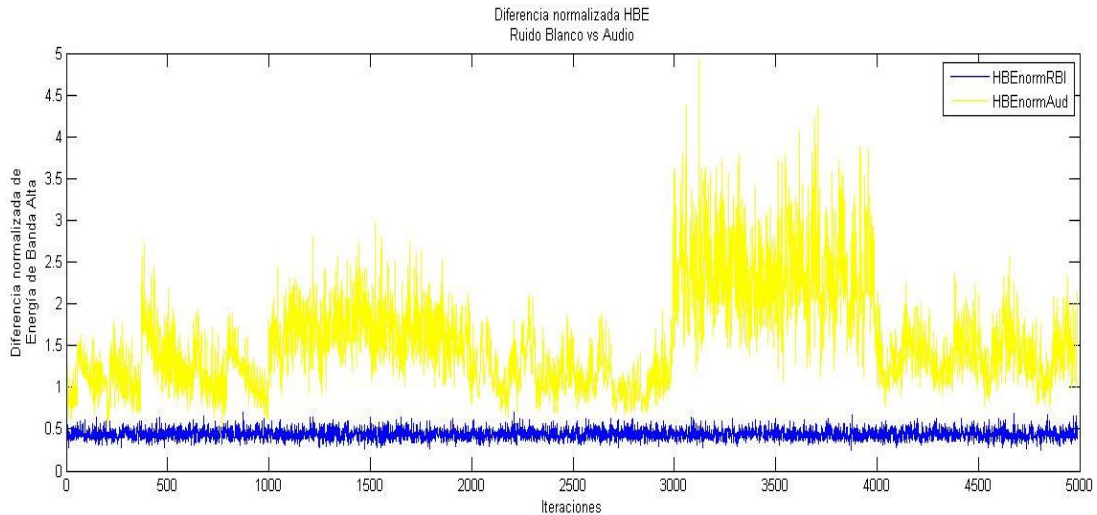


Figura III.07 Diferencia normalizada HBE Ruido Blanco vs Audio

Lo mismo sucede para el caso del ruido rosa, (figura III.08) donde se aprecia que, de igual manera al ruido blanco, pierde aproximadamente la mitad de la energía.

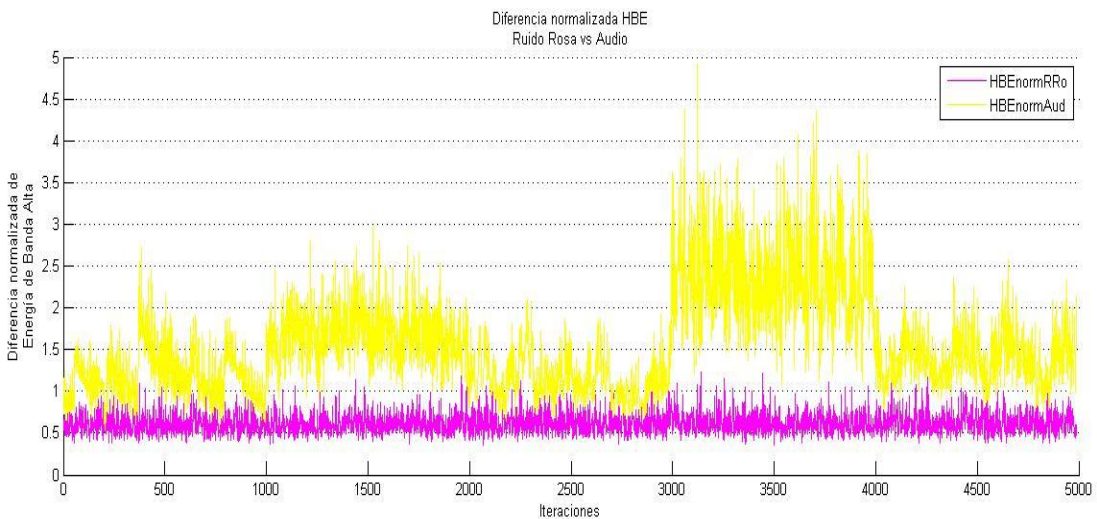


Figura III.08 Diferencia normalizada HBE Ruido Rosa vs Audio

Finalmente, para el ruido café (figura III.09), al someter ambas señales al filtro paso altas, se observa que casi no pierden energía ya que por su comportamiento característico en frecuencia, no llegan a tener información en frecuencias correspondientes del filtro.

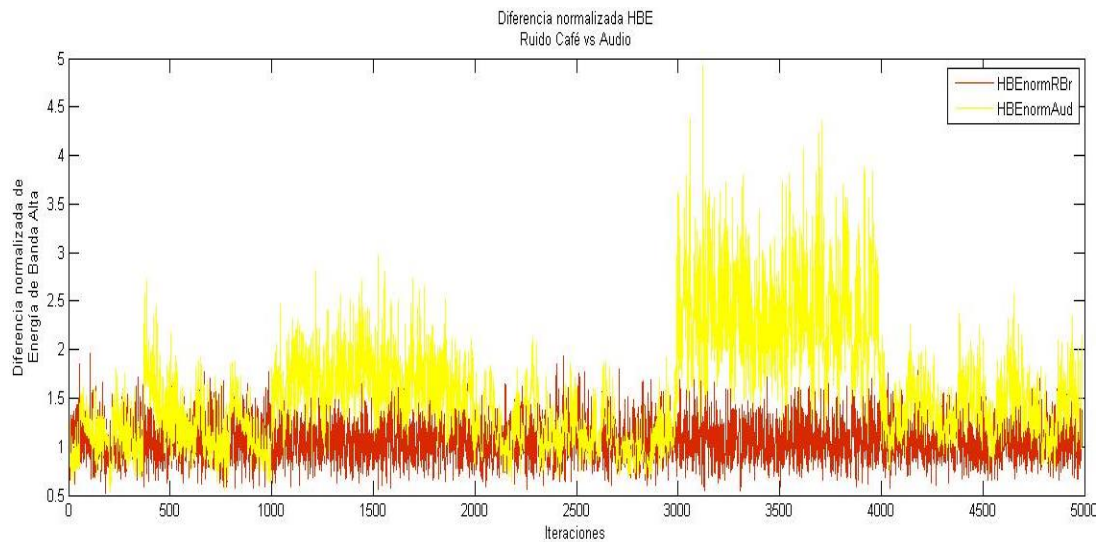


Figura III.09 Diferencia normalizada HBE Ruido Café vs Audio

CRUCE POR CEROS NORMALIZADO

En el caso del cruce por ceros, es interesante analizar este parámetro, debido a que es una ayuda para distinguir qué frecuencias pueden estar presentes en la señal, sin tener que aplicarle la Transformada Rápida de Fourier (FFT).

Comparando el cruce por ceros normalizado del ruido blanco contra la señal de audio en la figura III.10, es fácil observar que el cruce por ceros del ruido blanco supera a la señal de audio, esta característica nos permite diferenciar rápidamente entre ambas señales, lo que en nuestro controlador difuso será de mucha ayuda durante el proceso de hacer las reglas.

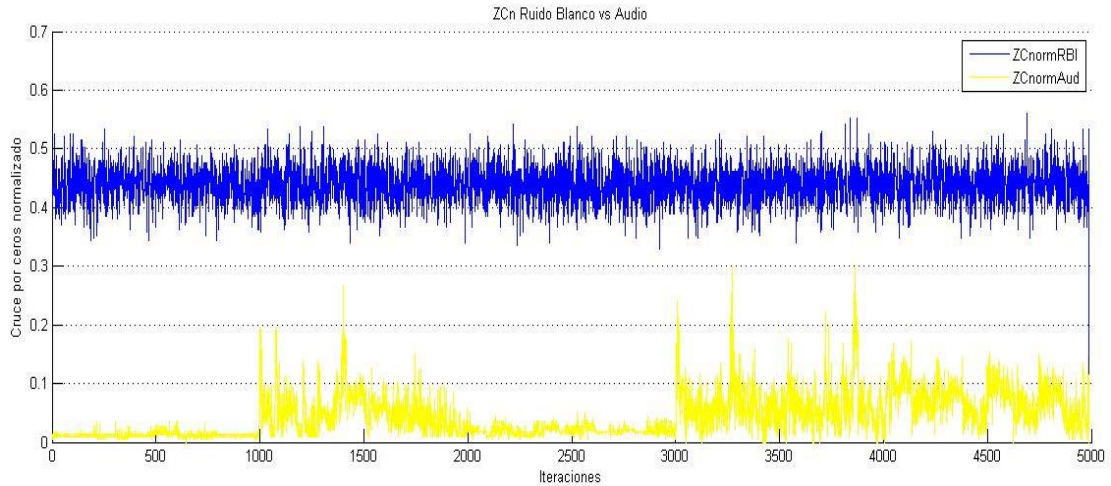


Figura III.10 Cruce por ceros normalizado Ruido Blanco vs Audio

Para el caso del ruido rosa y la señal de audio (figura III.11), el cruce por ceros es mayor en la mayoría de los casos en comparación del ruido rosa, con algunas excepciones donde, dependiendo del género musical, llega a tener algunas coincidencias en cuanto a su magnitud.

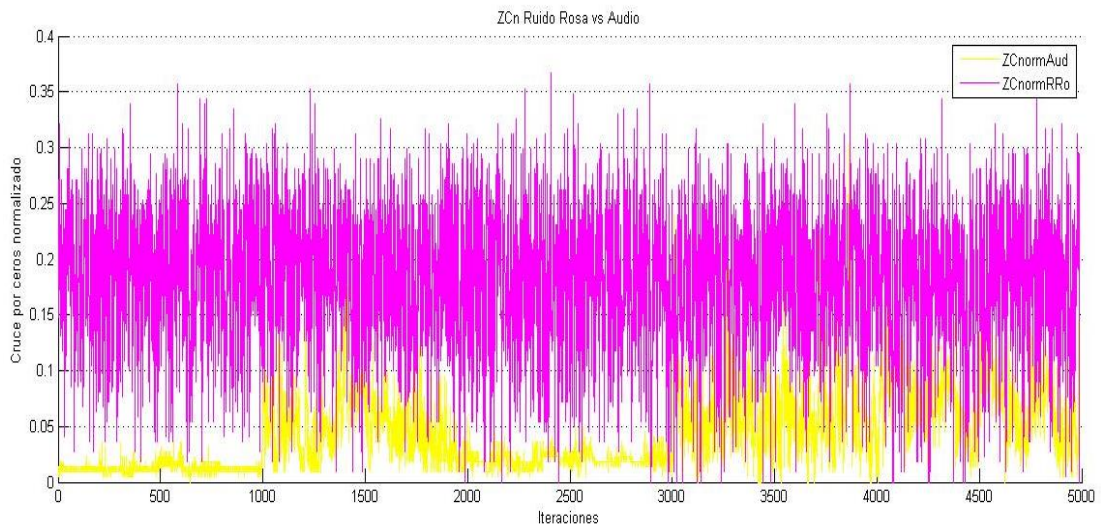


Figura III.11 Cruce por ceros normalizado Ruido Rosa vs Audio

Finalmente, para el cruce por ceros del ruido café y el audio, ambas señales son parecidas al aplicarle este parámetro, tal como se aprecia en la figura III.12, es por eso

que el comportamiento de las señales en su espectro de frecuencia, sus características son muy parecidas.

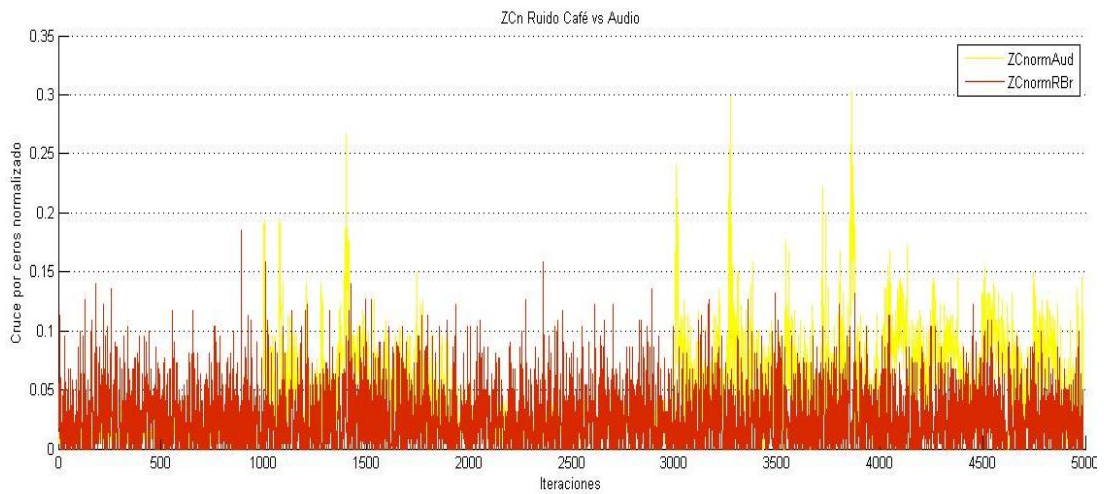


Figura III.12 Cruce por ceros normalizado Ruido Café vs Audio

Para este bloque de control difuso de inferencia, se utilizó la herramienta del software MATLAB Anfis Editor GUI, el cual permite optimizar por medio de redes neuronales el diseño de un control de inferencia difusa basado en el método de Sugeno, el procedimiento consiste en declarar cuantas entradas habrá, para nuestro caso cuatro: $difLBE_n$, $difMBE_n$, $difHBE_n$ y ZC_n ; y permite por cada entrada una regla, se hacen las funciones de pertenencia para cada entrada y se propone una forma para la función, se declaran las salidas y así mismo se asignan éstas a cada regla, además, la herramienta permite configurar cómo se desea la salida del controlador en un modo entrenamiento y, finalmente, la herramienta optimiza las funciones de pertenencia y propone una función de implicación para ser usada después en el proceso de agregación y desdifusión.

DIFUSIÓN

A continuación se muestran las funciones de pertenencia de los parámetros de las entradas:

- Diferencia grande de energía en banda baja normalizada (LBE_n): se define con la siguiente función:

$$\mu_{High}(difLBE_n) = \begin{cases} 1, & LBE_n \leq 0 \\ -\frac{10,000}{6,606}LBE_n + 1, & 0 \leq LBE_n < 0.6606 \\ 0, & COC \end{cases} \quad (III, 17)$$

- Diferencia grande de banda media de energía (MBE_n): se define con la función de pertenencia:

$$\mu_{High}(difMBE_n) = \begin{cases} 1, & MBE_n \leq 0 \\ -\frac{10,000}{1,932}LBE_n + 1, & 0 \leq MBE_n \leq 1.932 \\ 0, & COC \end{cases} \quad (III, 18)$$

- Diferencia grande de banda alta de energía (HBE_n): su función de pertenencia es:

$$\mu_{High}(difHBE_n) = \begin{cases} 1, & MBE_n \leq 0 \\ -\frac{10,000}{4,598}LBE_n + 1, & 0 \leq MBE_n \leq 4.598 \\ 0, & COC \end{cases} \quad (III, 19)$$

- Cruce por ceros normalizado grande: este parámetro nos permite predecir que las muestras tienen frecuencias altas, por lo que es probable que haya ruido, su función se define:

$$\mu_{High}(ZC_n) = \begin{cases} 1, & ZC_n > 0 \\ -\frac{10,000}{5,249} * ZC_n + 1, & 0 \leq ZC_n < 0.5249 \\ 0, & COC \end{cases} \quad (III, 20)$$

- Cruce por ceros normalizado medio: se trata de una señal de audio que tiene agregado ruido, su función de pertenencia se define:

$$\mu_{Medium}(ZC_n) = \{0, -\infty < ZC_n < \infty \quad (III, 21)$$

- Cruce por ceros normalizado bajo: gracias a este parámetro permite saber que las frecuencias de las muestras son bajas, por lo que en la señal hay actividad de audio, su función se define:

$$\mu_{Low}(ZC_n) = \{0, -\infty < ZC_n < \infty \quad (III, 22)$$

ANTECEDENTE TOTAL

A continuación se muestran se presentan las reglas que se usarán en el control difuso:

R1: IF difLBE_n is High OR ZC_n is High THEN Audio Activity is Z₁

R2: IF difHBE_n is High THEN Audio Activity is Z₂

R3: IF difMBE_n is NOT High AND ZC_n is NOT Low THEN Audio Activity is Z₃

R4: IF ZC_n is Medium THEN Audio Activity is Z₄

Ya que tenemos definidas las reglas, se obtienen los antecedentes totales:

$$W_1 = ATR1 = \text{Max}(\text{difLBE}_{n\text{High}}, ZC_{n\text{High}})$$

$$W_2 = ATR2 = \text{difHBE}_{n\text{High}}$$

$$W_3 = ATR3 = \text{min}(1 - \text{difMBE}_{n\text{High}}, 1 - ZC_{n\text{High}})$$

$$W_4 = ATR4 = ZC_{n\text{High}}$$

IMPLICACIÓN

Una vez que se tienen los antecedentes totales, procederemos a la implicación, en la cual el método de Sugeno nos da las siguientes funciones para cada una de las reglas:

$$Z_1 = 4.845 * \text{difLBE}_n - 0.4043 * \text{difMBE}_n + 1.656 * \text{difHBE}_n - 9.253 * ZC_n - 4.933 \quad (III, 23)$$

$$Z_2 = -11.3 * \text{difLBE}_n - 7.93 * \text{difMBE}_n + 1.868 * \text{difHBE}_n + 2.533 * ZC_n + 5.434 \quad (III, 24)$$

$$Z_3 = -12.1 * \text{difLBE}_n + 5.932 * \text{difMBE}_n - 4.793 * \text{difHBE}_n + 6.661 * ZC_n + 11.11 \quad (III, 25)$$

$$Z_4 = 5.406 * \text{difLBE}_n + 4.651 * \text{difMBE}_n - 1.327 * \text{difHBE}_n - 0.3003 \quad (III, 26)$$

AGREGACIÓN Y DESDIFUSIÓN

Finalmente obtendremos nuestro valor de la salida que representa la pertenencia a la actividad de audio:

$$AAD_i = Z^* = \frac{w_1 * z_1 + w_2 * z_2 + w_3 * z_3 + w_4 * z_4}{w_1 + w_2 + w_3 + w_4} \quad (III, 27)$$

IV. RESULTADOS

Una vez que ya se explicó el desarrollo del algoritmo, a continuación se presentan los resultados obtenidos con el programa implementado en MATLAB. En primer lugar, se presentará la evaluación para cada bloque independiente y, posteriormente, el acoplamiento de ambos bloques.

IV.1 BLOQUE DE DETECCIÓN DE ACTIVIDAD DE AUDIO

La evaluación se hizo a partir de los tres tipos de ruido: blanco, rosa y café y una colección de pistas de audio las cuales son las siguientes:

1. Archivo de audio de Campana Tibetana
2. Archivo de audio de artista Leo Brouwer (música instrumental)
3. Archivo de audio de artista Pentatonix (música vocal)
4. Archivo de audio de artista Brooklyn Duo (música instrumental)
5. Archivo de audio de artista Black Eyed Peas
6. Música de audio de artista Slash

El programa toma muestras de cada colección aleatoriamente y se les aplicó el Detector de Actividad de Audio. En la figura IV.1.1 se muestra, en color verde, el valor deseado de pertenencia y, en color azul, el valor de pertenencia obtenido. Se hicieron quinientas iteraciones por archivo de audio y, a continuación, se presenta el valor promedio obtenido:

1. AAD Ruido Blanco: 0.0071
2. AAD Ruido Rosa: 0.1800
3. AAD Ruido Café: 0.6934
4. AAD Archivo de audio de Campana Tibetana: 0.8681
5. AAD Archivo de audio de artista Leo Brouwer (música instrumental): 0.7818
6. AAD Archivo de audio de artista Pentatonix (música vocal): 0.9977
7. AAD Archivo de audio de artista Brooklyn Duo (música instrumental): 0.7912
8. AAD Archivo de audio de artista Black Eyed Peas: 0.9894
9. AAD Archivo de audio de artista Slash: 0.8811

El bloque detecta de manera correcta una nula actividad de audio para el ruido blanco y rosa y, únicamente, para el ruido café el AAD aumenta pero es menor que el AAD de los archivos de audio.

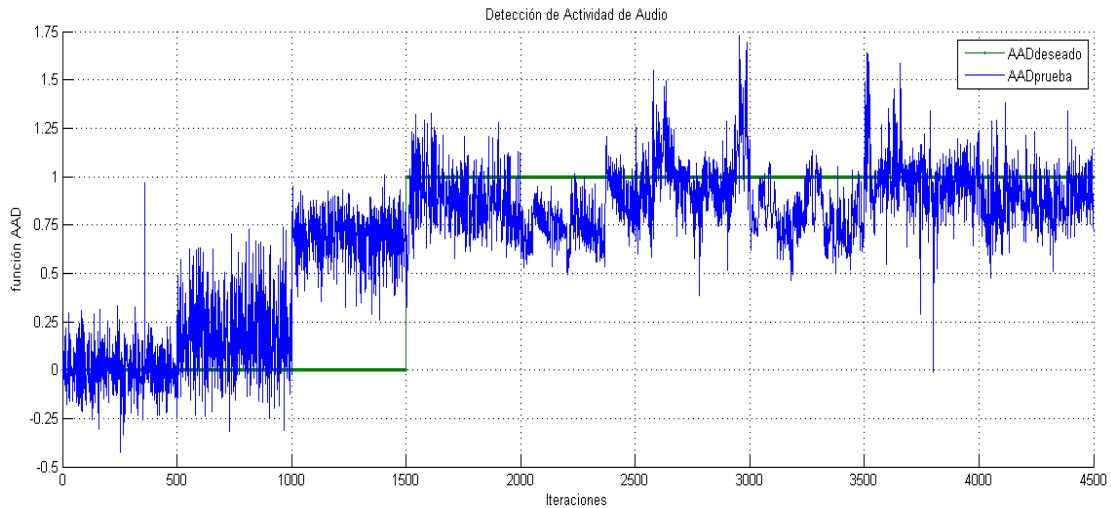


Figura IV.01 Función de pertenencia del Detector de Actividad de Audio

IV.2 BLOQUE DE CONTROL DE GANANCIA AUTOMÁTICA

Para el bloque de Control de Ganancia Automática, se hizo la prueba con dos archivos de audio sin ruido agregado. Éstos contienen pistas de música comercial y, otro, una grabación del sonido de una campana tibetana. Se fijó un nivel RMS deseado a la salida y se probaron tres casos para cada pista: se amplificó la señal en la entrada buscando que en la salida el valor RMS fuera menor que la entrada; reduciendo la señal de entrada, buscando amplificarla y, por último, dando un valor deseado a la salida sin manipular la señal de entrada.

A continuación, se muestra el comportamiento de los diferentes casos propuestos: La figura IV.02 corresponde al caso 1 en donde sin manipular la señal de entrada y el valor RMS deseado igual a 0.5 muestra el nivel RMS de las señales de entrada y de la

salida obtenida, se aprecia como al inicio se amplifica, pero debido a que la señal de entrada empieza a disminuir, la ganancia llega a su límite superior, por lo que ya no puede mantener el nivel RMS esperado.

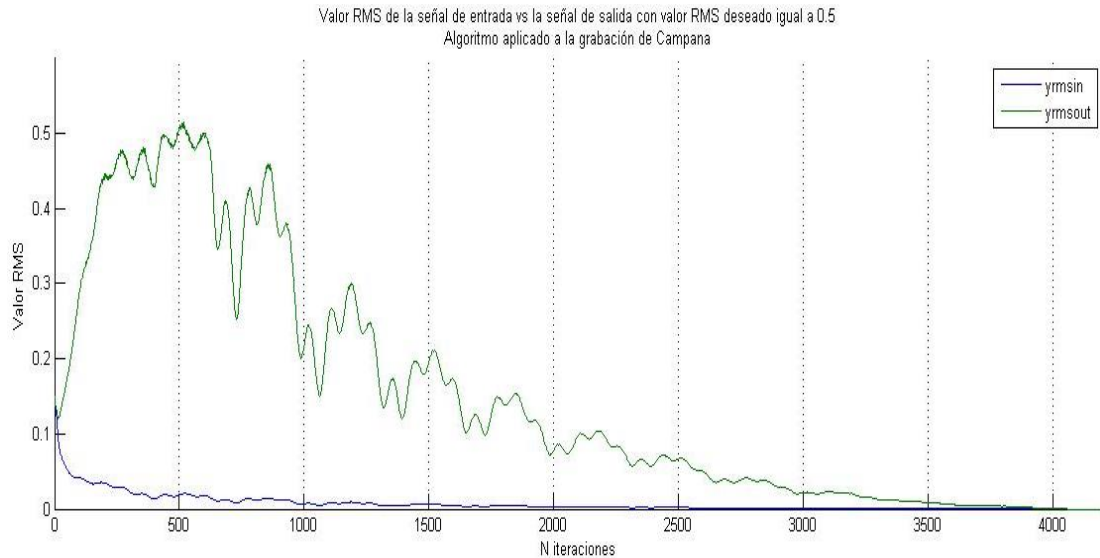


Figura IV.02 Valor RMS Caso 1

Para el caso 2, donde la señal de entrada se multiplica por 3, se desea en la salida un valor RMS deseado igual a 0.3, por lo que el objetivo del controlador será disminuir la ganancia para alcanzar el nivel deseado, se observa en la figura IV.03 la comparación del nivel RMS de la entrada *versus* la señal a la salida:

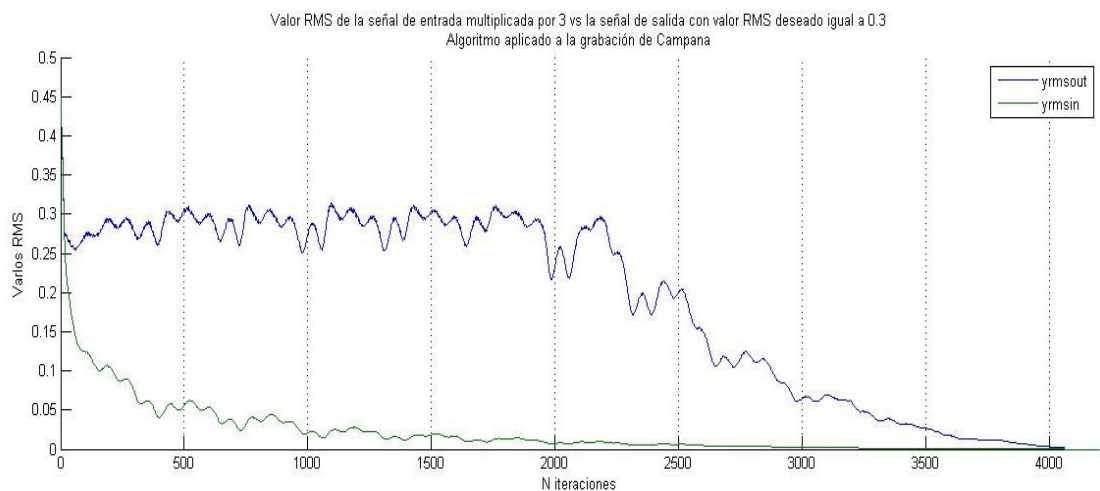


Figura III.03 Valor RMS Caso 2

Por último, para el mismo archivo, en el caso 3 se multiplicó la señal de entrada por 0.25, y a la salida se propuso un valor deseado de 0.2, se aprecia en la figura IV.04 que únicamente por unos instantes se logra mantener el nivel deseado debido al comportamiento de la señal de entrada:

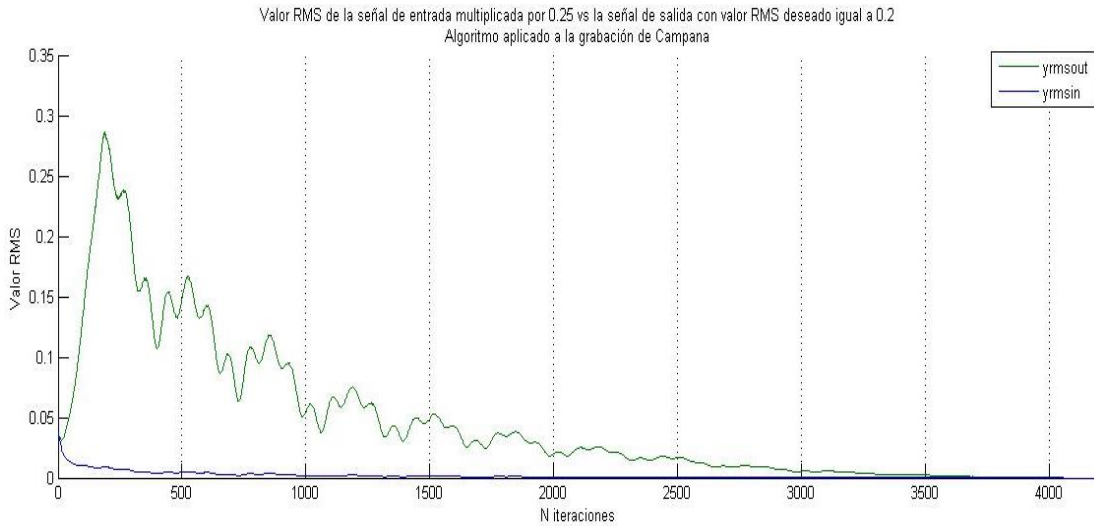


Figura IV.04 Valor RMS Caso 3

Ahora se presentarán los resultados para la señal de error, para el caso 1, se aprecia en la figura IV.05 como inicialmente la señal de error al ser casi de 0.8 (80%) significa que la señal de salida es mucho menor que el valor deseado, y a medida que el controlador permite darle ganancia, alrededor de la iteración 500 el error es igual a 0, y éste nuevamente empieza a crecer debido al comportamiento de la señal de entrada.

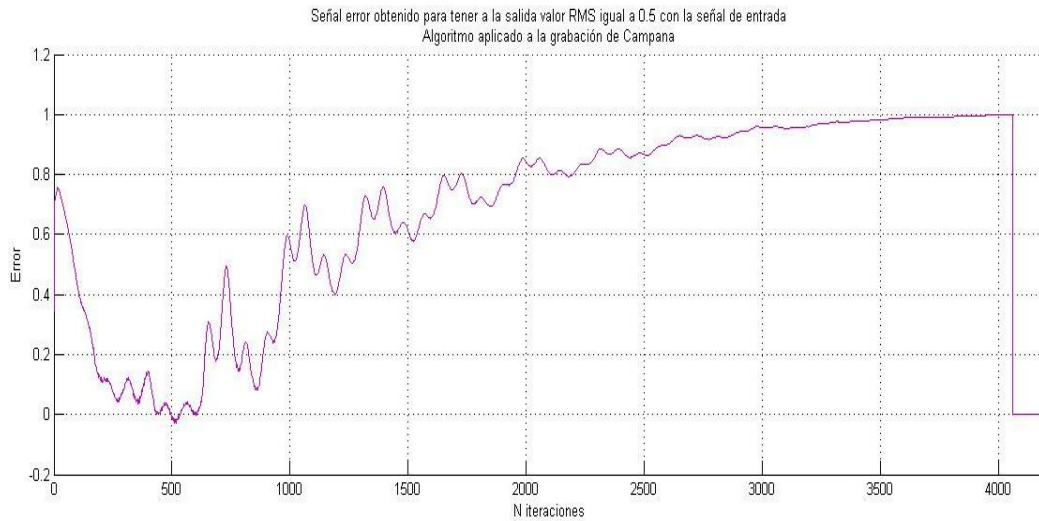


Figura IV.05 Señal error Caso 1

En el caso 2, se observa en la figura IV.06 que el error permaneció cercano a cero por más tiempo, confirmando que se alcanzó el nivel RMS deseado, y únicamente cuando la ganancia ya no es suficiente para mantener el nivel deseado, crece la señal de error debido al comportamiento de la señal de entrada.

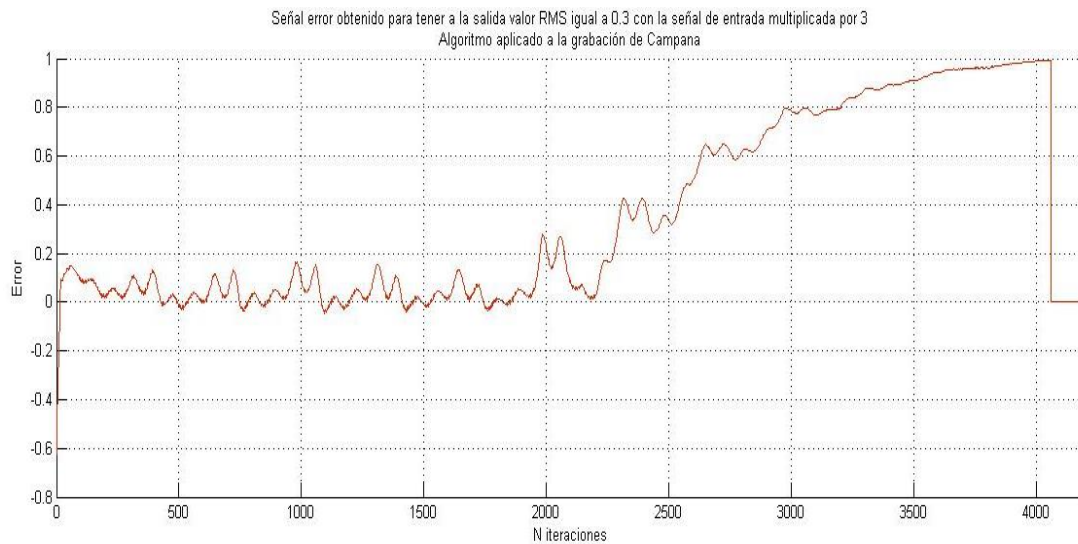


Figura IV.06 Señal error Caso 2

Para el caso 3, se muestra en la figura IV.07 que el error no llega a cero debido a que la ganancia no es suficiente para dar a la salida el nivel RMS deseado.

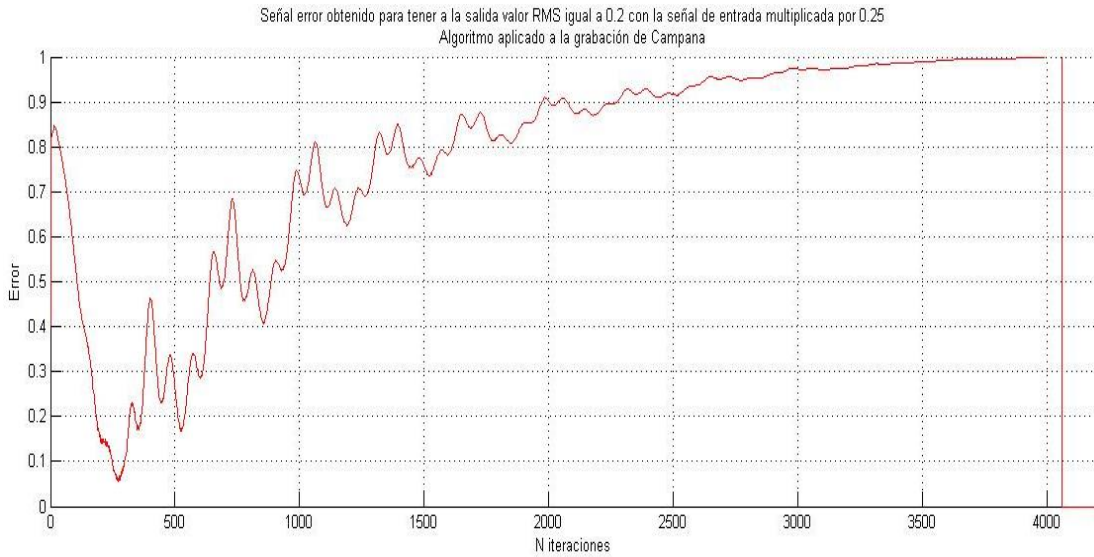


Figura IV.07 Señal error Caso 3

La señal derivada del error significa el cambio del nivel de audio (volumen) en la señal de salida, para el caso de la aplicación del algoritmo al archivo de la campana Tibetana, al ser un sonido que va disminuyendo en volumen, la figura muestra como en la parte que llega a amplificar varía mucho el nivel hasta finalmente se estabiliza en cero debido a que la señal de salida ya no tiene el nivel deseado y no varía por el límite superior de la ganancia, así será el comportamiento en los casos 1, 2 y 3, tales como lo muestran las figuras IV.08, IV.09 y IV.10:

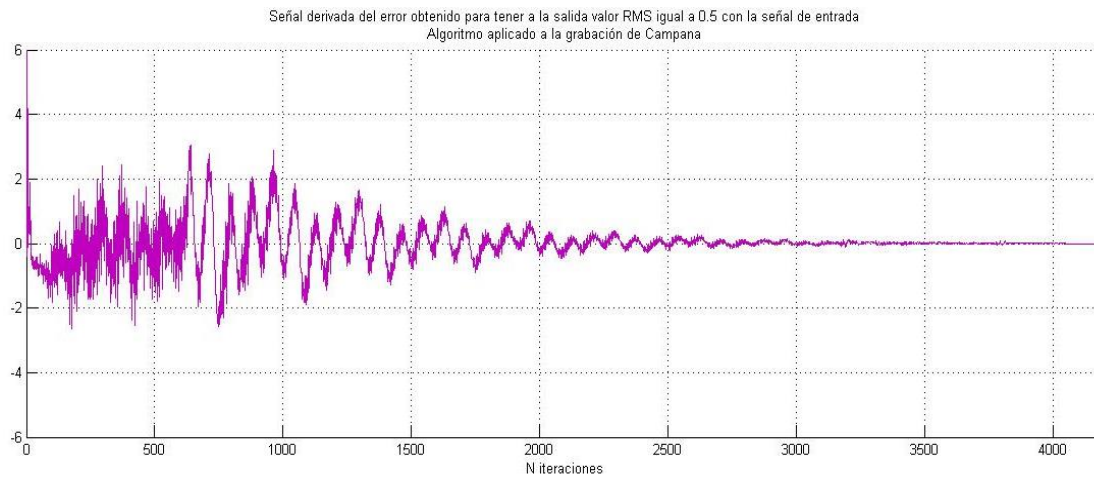


Figura IV.08 Señal derivada del error Caso 1

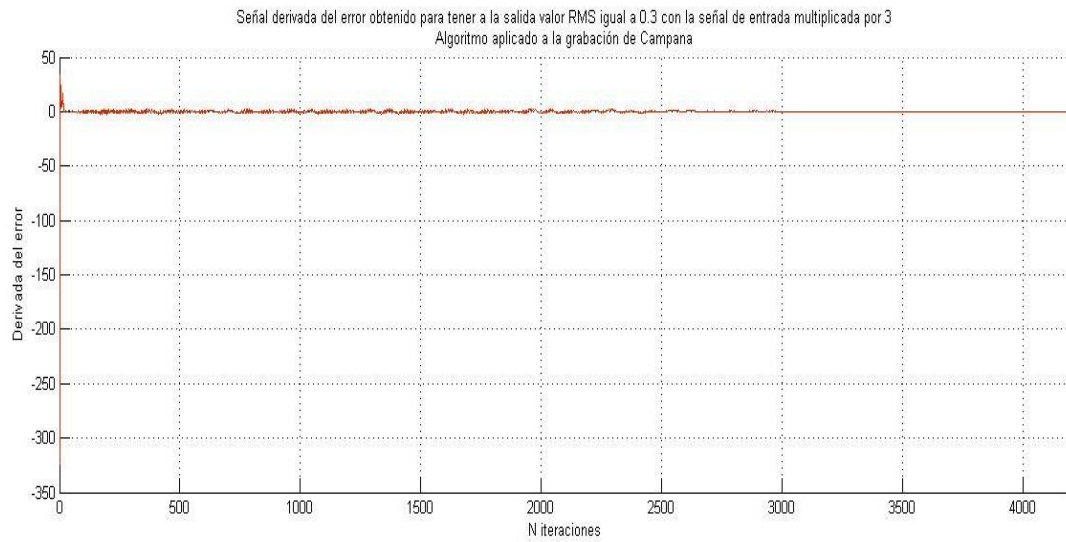


Figura IV.09 Señal derivada del error Caso 2

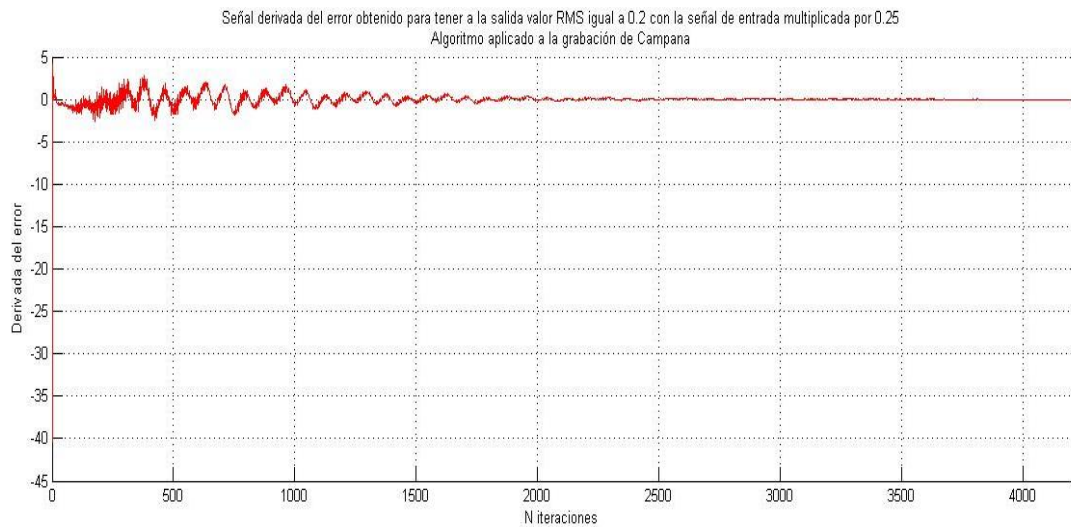


Figura IV.10 Señal derivada del error Caso 3

Sobre el incremento de la ganancia, en la figura IV.11 se observa que, al inicio, la magnitud de los incrementos disminuyen para las primeras iteraciones, porque se ha alcanzado el nivel de salida deseado, pero por el comportamiento de la señal de entrada, el incremento aumenta hasta que se estabiliza en 0.05 aproximadamente debido a la combinación con el error en el bloque de control de ganancia.

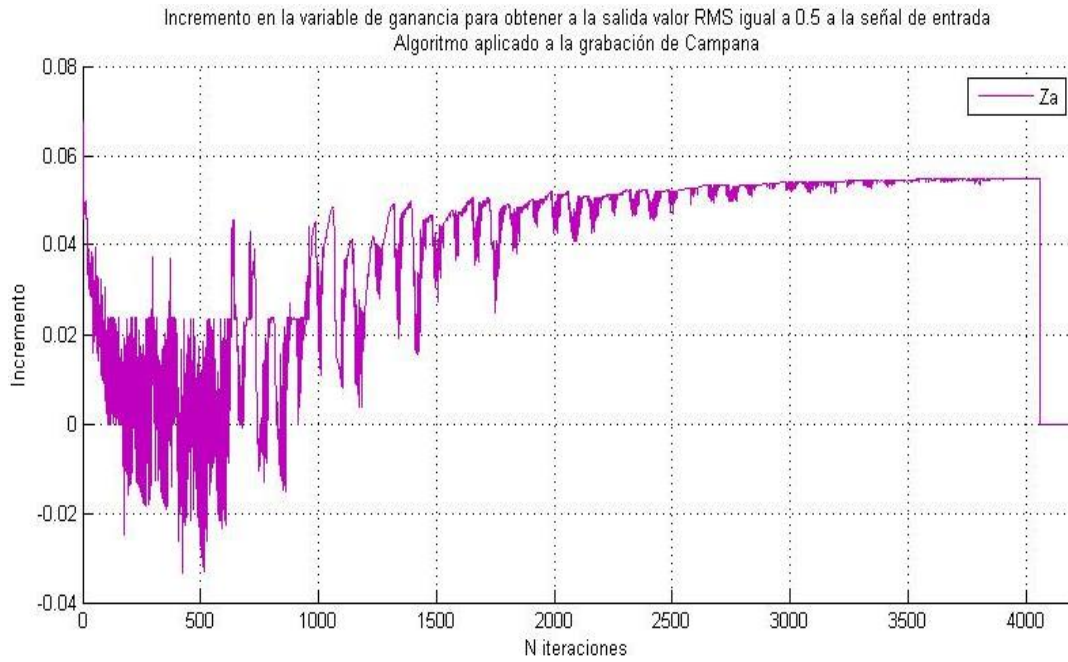


Figura IV.11 Incremento de ganancia Caso 1

En la figura IV.12 se observa cuál es el valor de la ganancia que multiplica a la señal de entrada para obtener el valor RMS deseado, el cual alcanza el valor del límite superior.

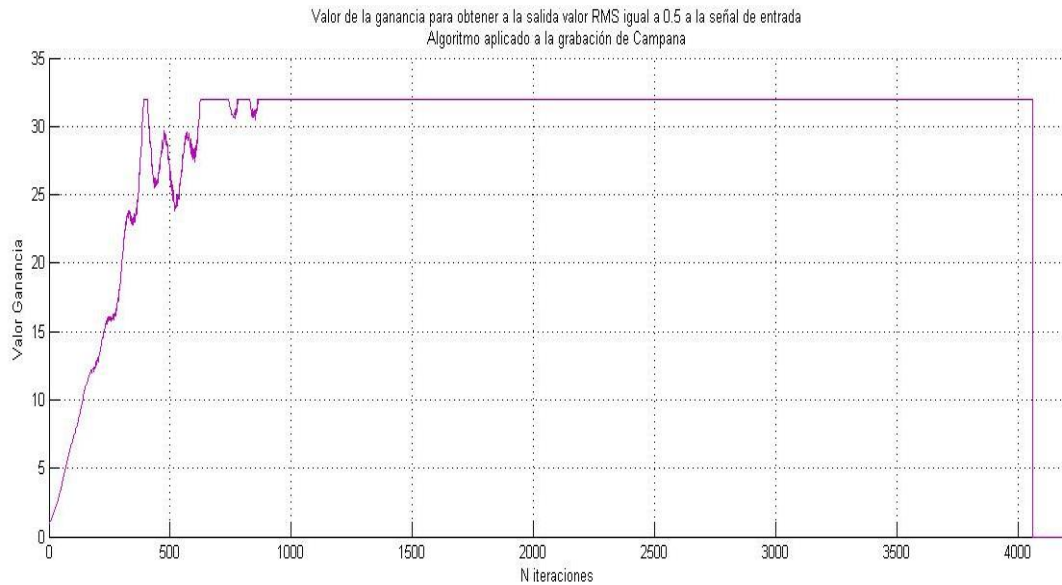


Figura IV.12 Valor de la ganancia Caso 1

A diferencia del caso 1, para mantener el nivel RMS deseado en el caso 2, el incremento aumenta y disminuye, hasta que a señal de entrada empieza a disminuir, se puede ver el comportamiento en la figura IV.13

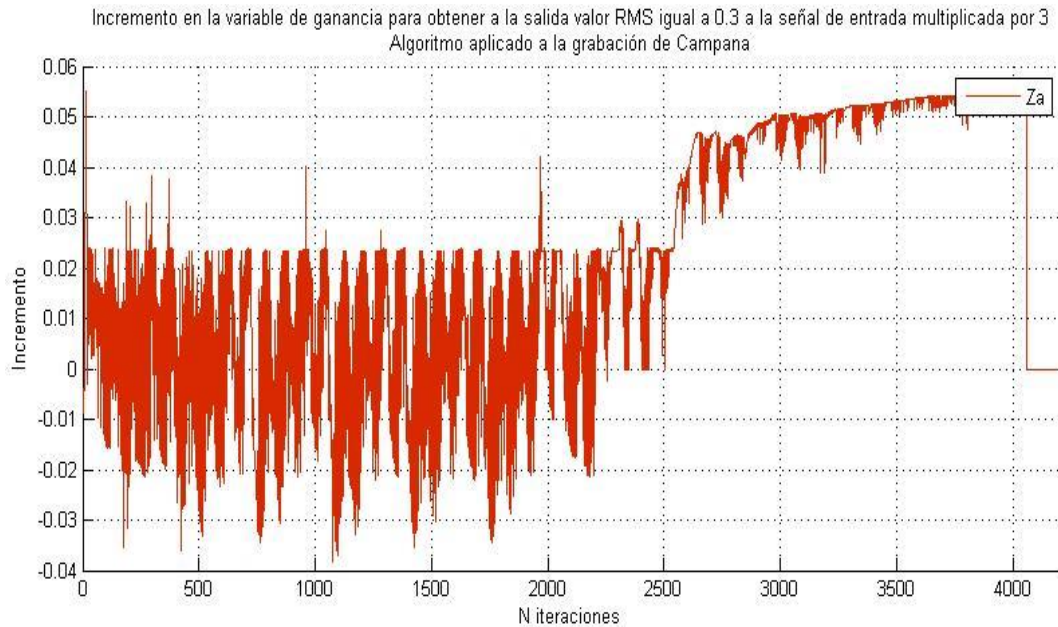


Figura IV.13 Incremento de Ganancia Caso 2

En la figura IV.14 se distingue el comportamiento que tiene la ganancia hasta que llega a su valor máximo.

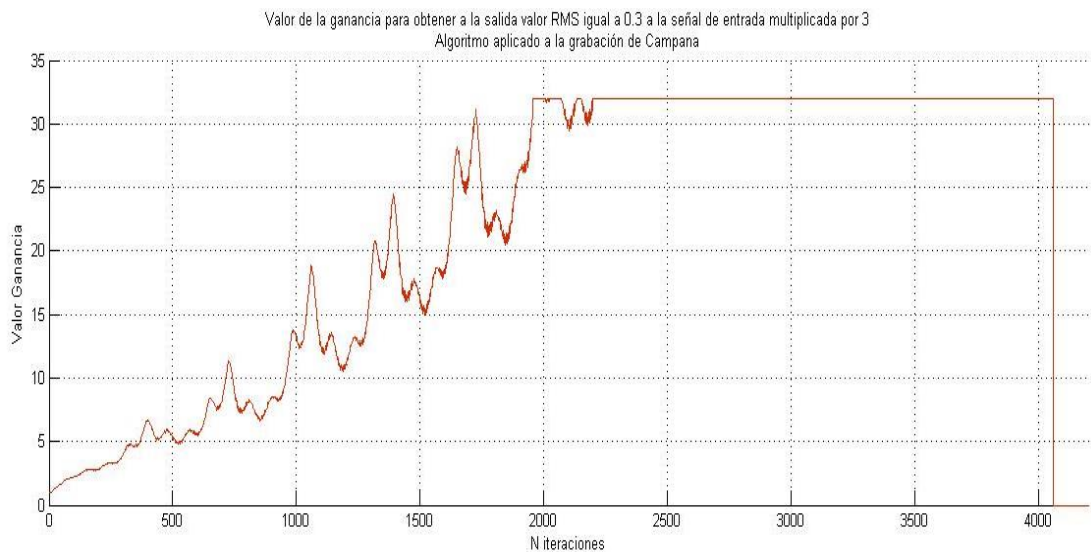


Figura IV.14 Valor de la Ganancia Caso 2

Finalmente, para el caso 3, los incrementos son en su mayoría positivos, con el objetivo de alcanzar el nivel RMS deseado, tal y como se muestra en la figura IV.15:

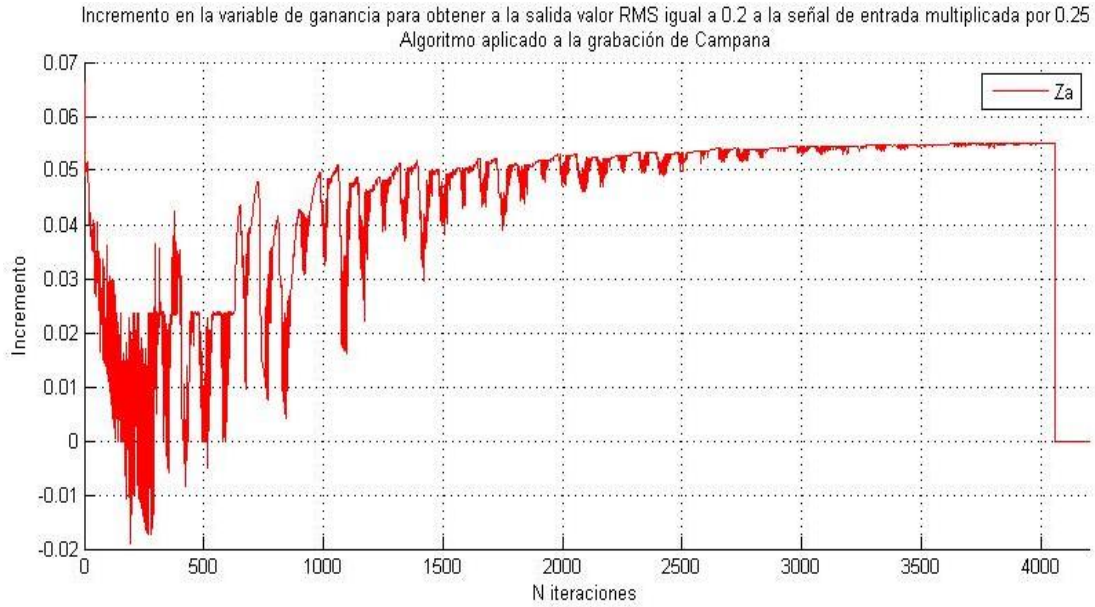


Figura IV.15 Incremento de la Ganancia Caso 3

En la figura IV.16 se alcanza el valor máximo de ganancia desde las primeras iteraciones debido a que la señal de entrada es muy pequeña y debe ser multiplicada por un valor muy grande para así poder tener a la salida el valor RMS deseado.

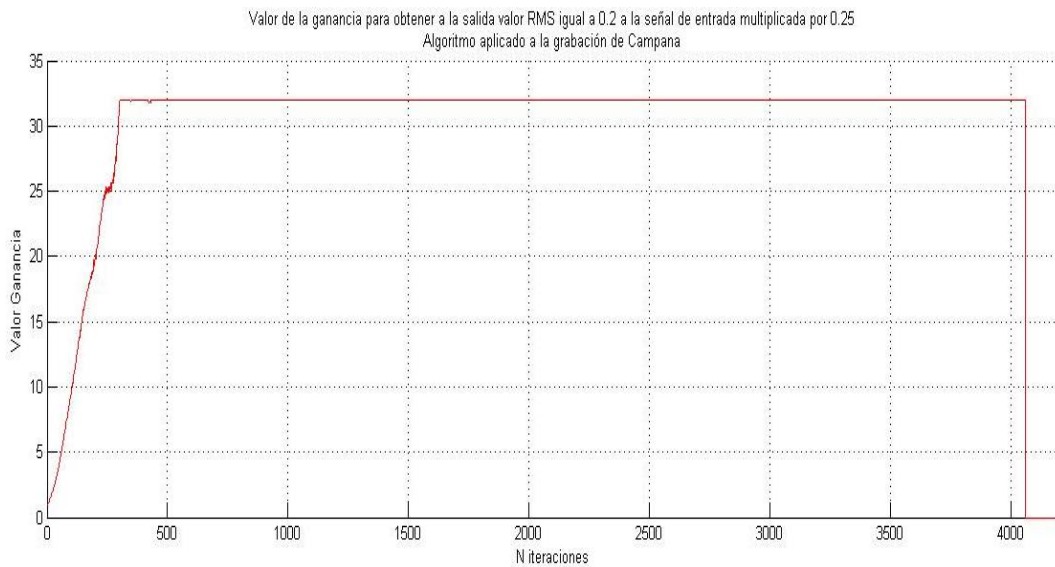


Figura IV.16 Valor de la Ganancia Caso 3

Continuando ahora con la aplicación del control difuso a un archivo WAV de música comercial, los casos 4, 5 y 6 son con las mismas condiciones que los casos 1,2 y 3, por lo que a continuación se muestran las gráficas de los valores RMS de entrada contra la salida, en las figuras IV.17, IV.18 y IV.19:

Se observa que el nivel RMS en la salida varía mucho, y es debido al comportamiento de la entrada, pero para los tres casos busca mantener el nivel RMS deseado.

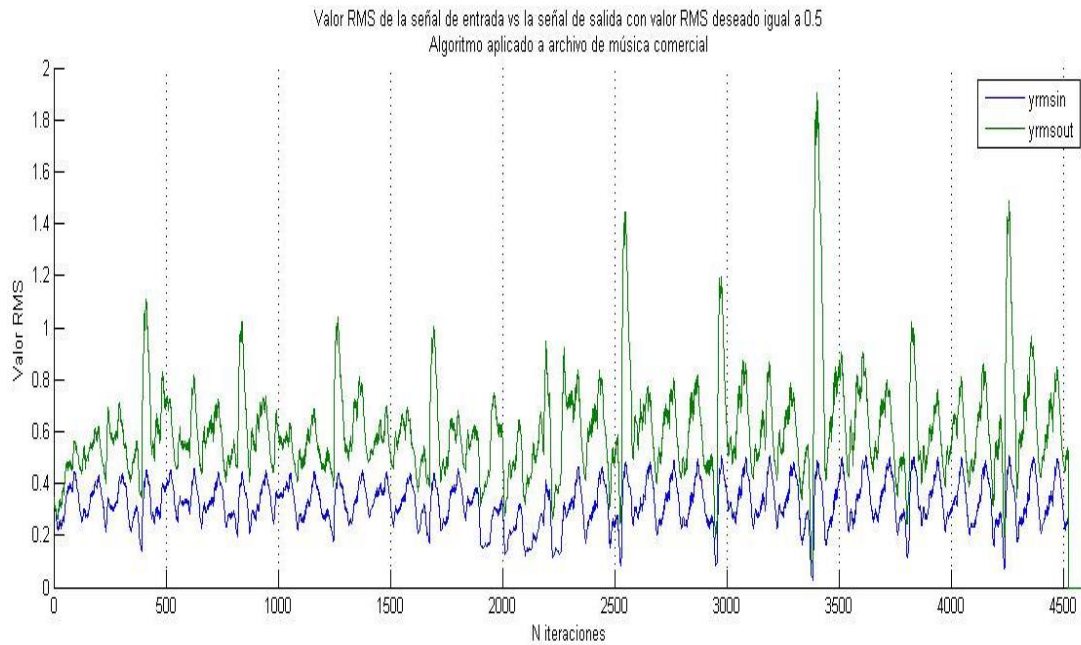


Figura IV.17 Valor RMS Caso 4

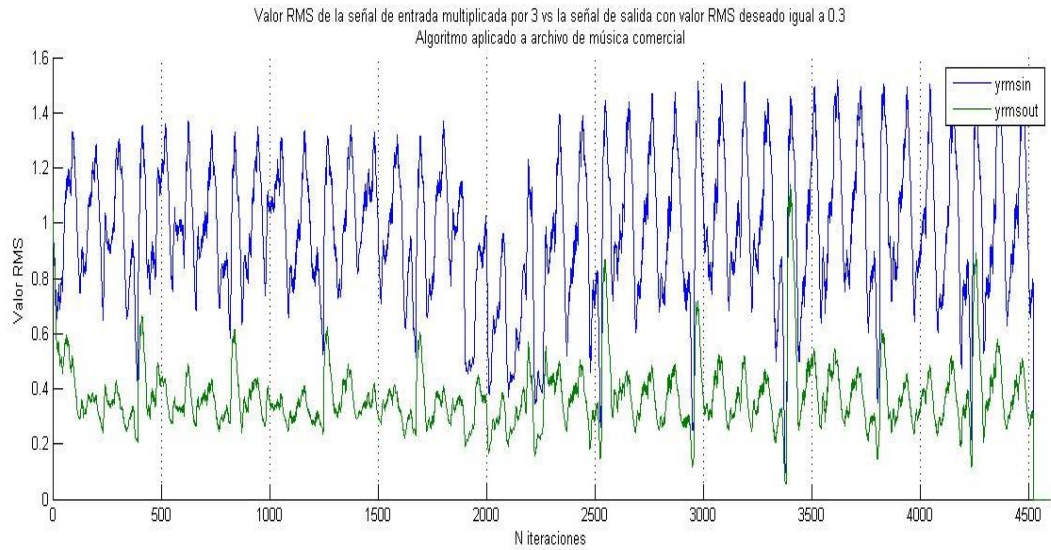


Fig. IV.18 Valor RMS Caso 5

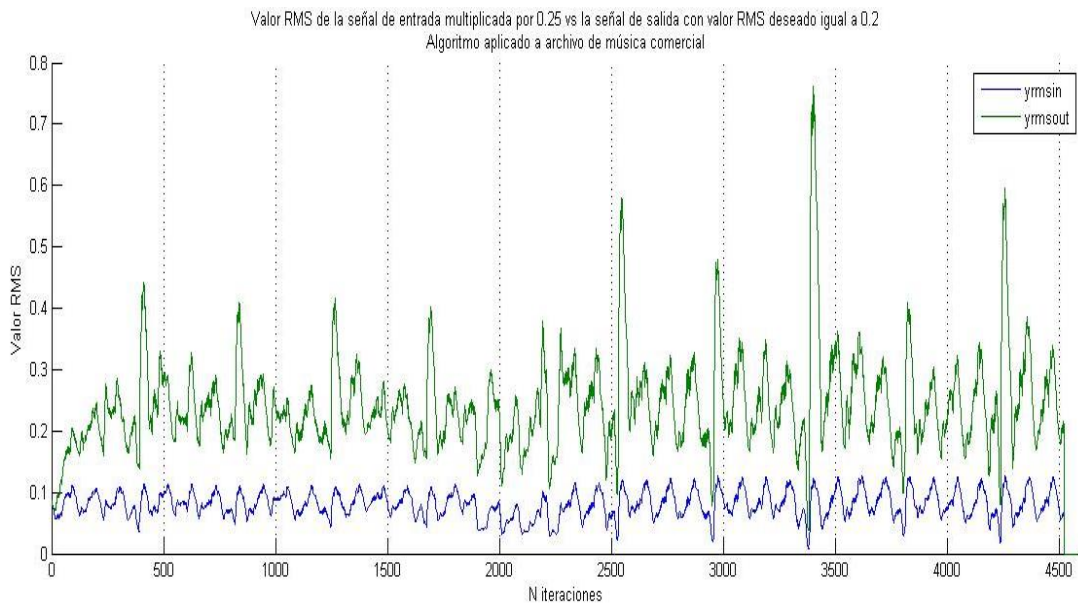


Figura IV.19 Valor RMS Caso 6

Debido al comportamiento de la señal de entrada, cuando llega a haber un cambio significativo en el nivel RMS, es cuando en conjunto con la señal derivada del error, permiten un cambio en el incremento de ganancia. Además, a diferencia de los casos 1, 2 y 3 no es una señal cuyo nivel RMS vaya decreciendo, sino que las variaciones que

va teniendo permiten al controlador buscar el incremento necesario para mantener el nivel RMS deseado.

Para los casos 4, 5 y 6, la señal de error relativo tiende a tener la misma forma, y se observa en las figuras IV.20, IV.21 y IV.22, es decir, tiene los mismos picos aun cuando la señal de entrada esta multiplicada por diferentes constantes, esto es por el comportamiento de la señal de entrada.

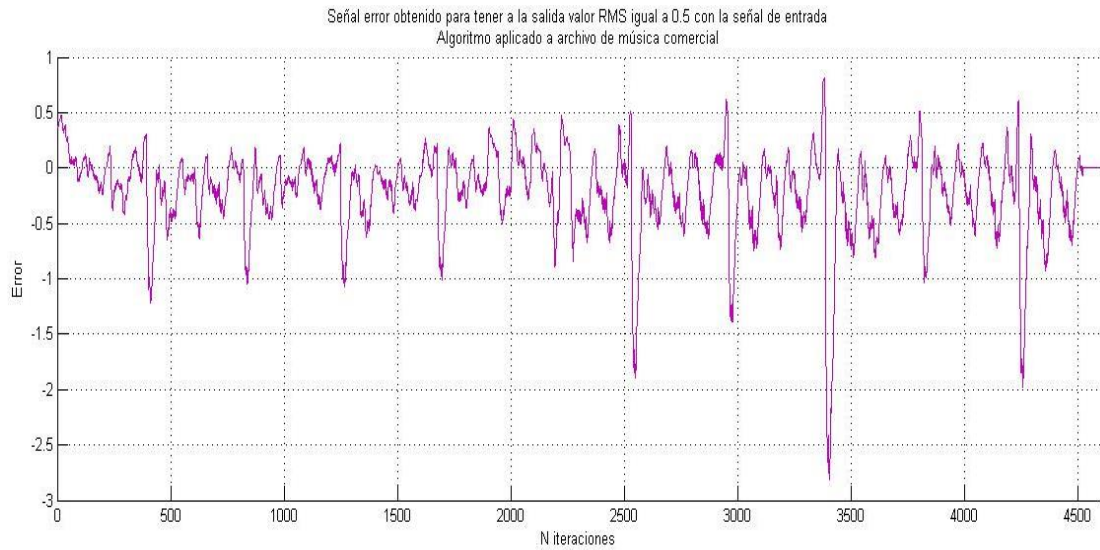


Figura IV.20 Señal Error Caso 4

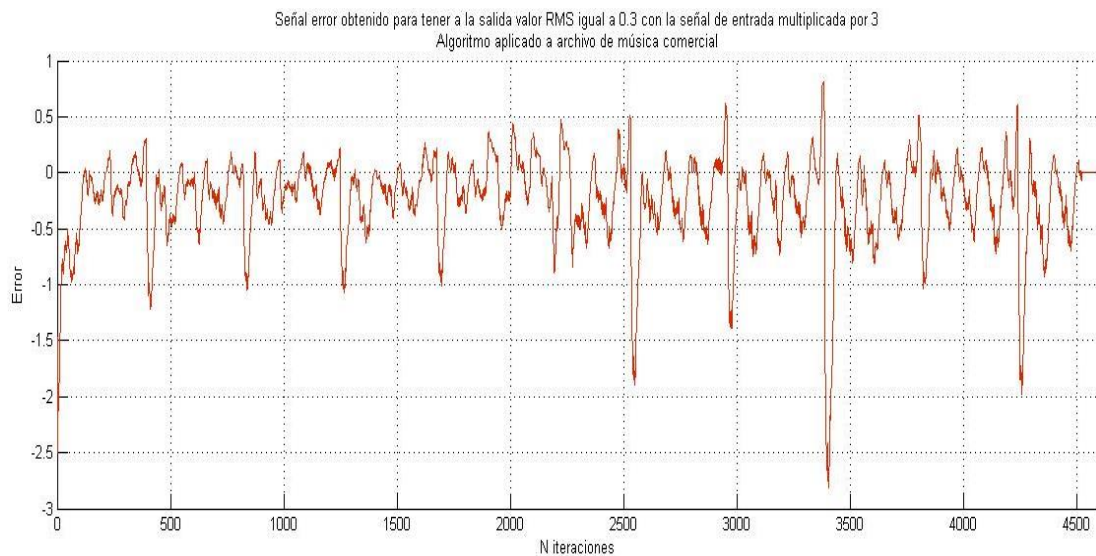


Figura IV.21 Señal Error Caso 5

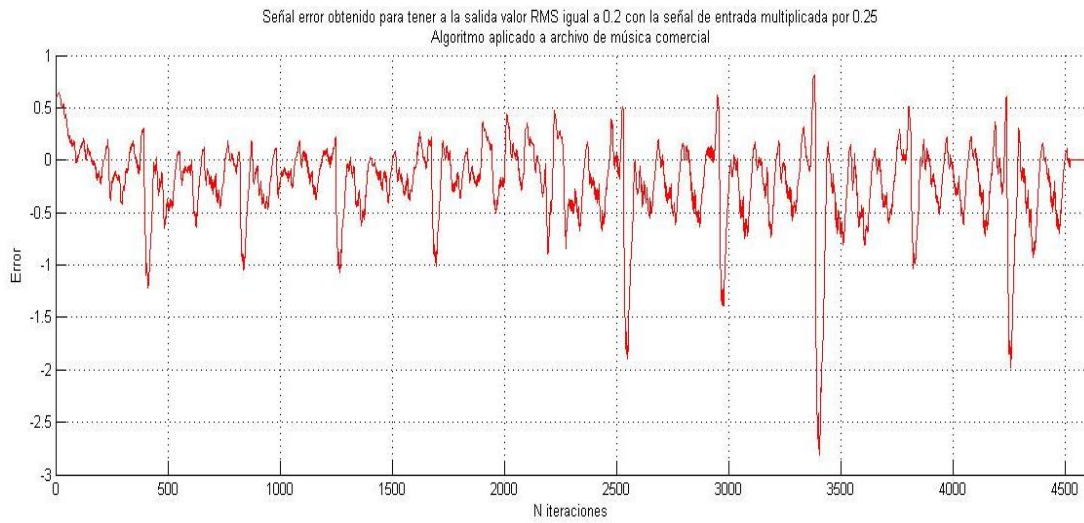


Figura IV.22 Señal Error Caso 6

Lo mismo pasa con la señal derivada del error y se debe a que tienen los mismos cambios de nivel, los cuales son proporcionales a la constante que multiplica a la señal de entrada, pero suceden en el mismo lapso de tiempo, este comportamiento se observa en las figuras IV.23, IV.24 y IV.25:

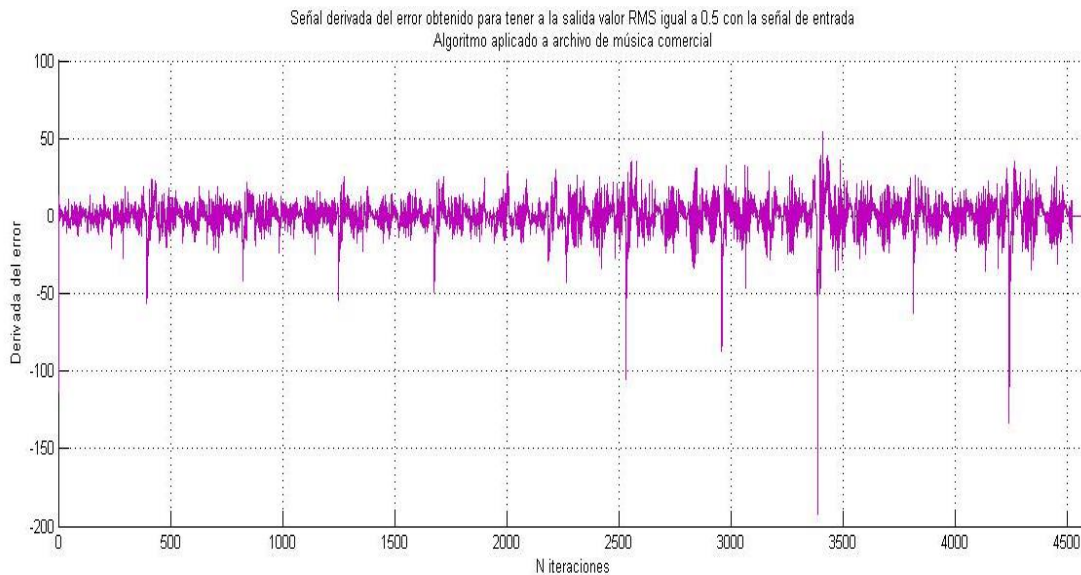


Figura IV.23 Señal derivada del error Caso 4

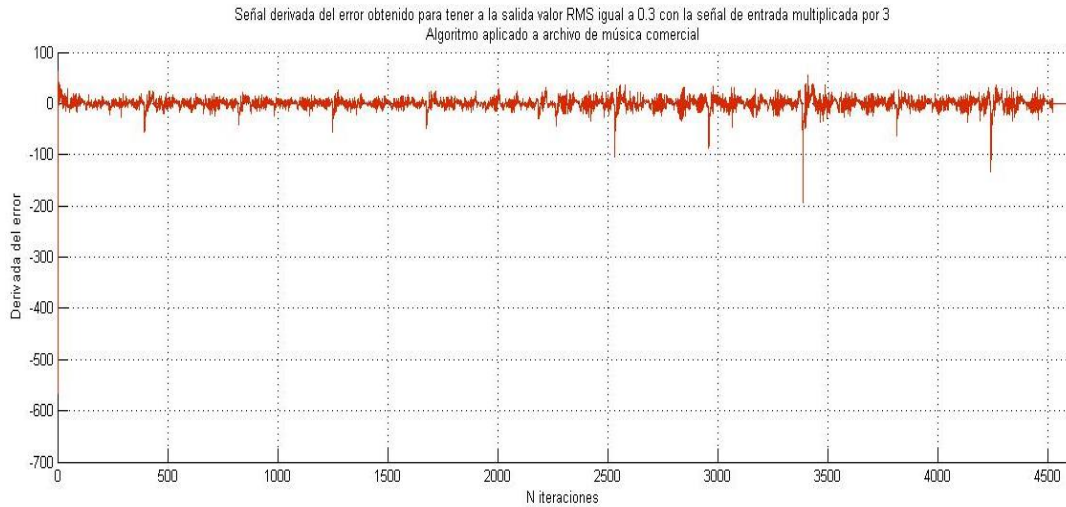


Figura IV.24 Señal derivada del error Caso 5

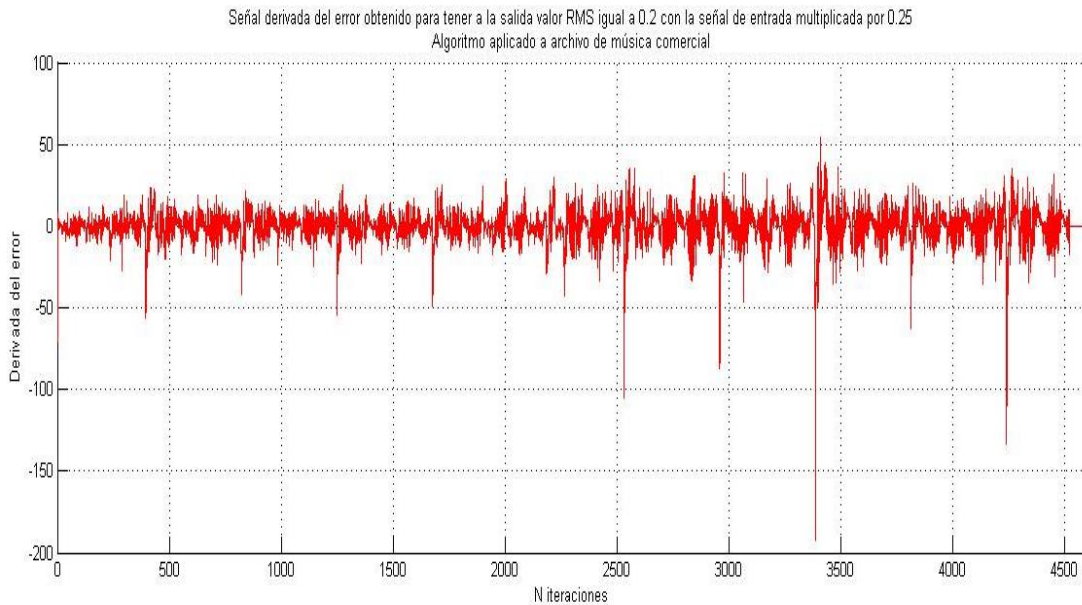


Figura IV.25 Señal derivada del error Caso 6

En las siguientes figuras (de la IV.26 a la IV.31) se aprecian los incrementos de ganancia para cada caso y también el valor de la ganancia, la diferencia entra cada caso es que para los casos 4 y 6 se multiplica por constantes mayores a 1 debido a que la señal se debe amplificar, mientras que en el caso 5 la constante es menor a 1 ya que la señal de entrada en valor RMS supera al valor deseado.

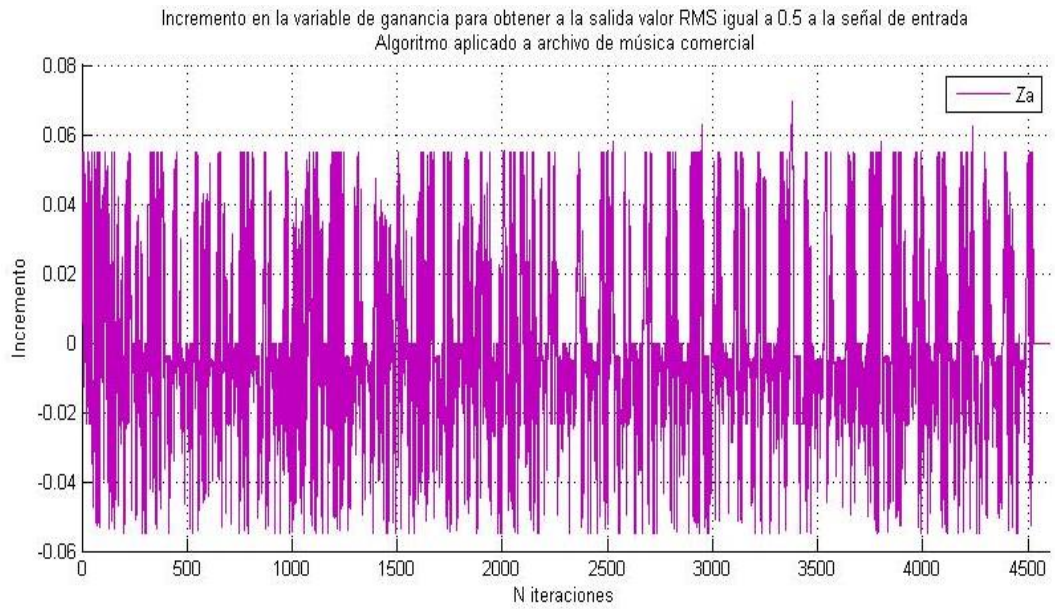


Figura IV.26 Incremento de la Ganancia RMS Caso 4

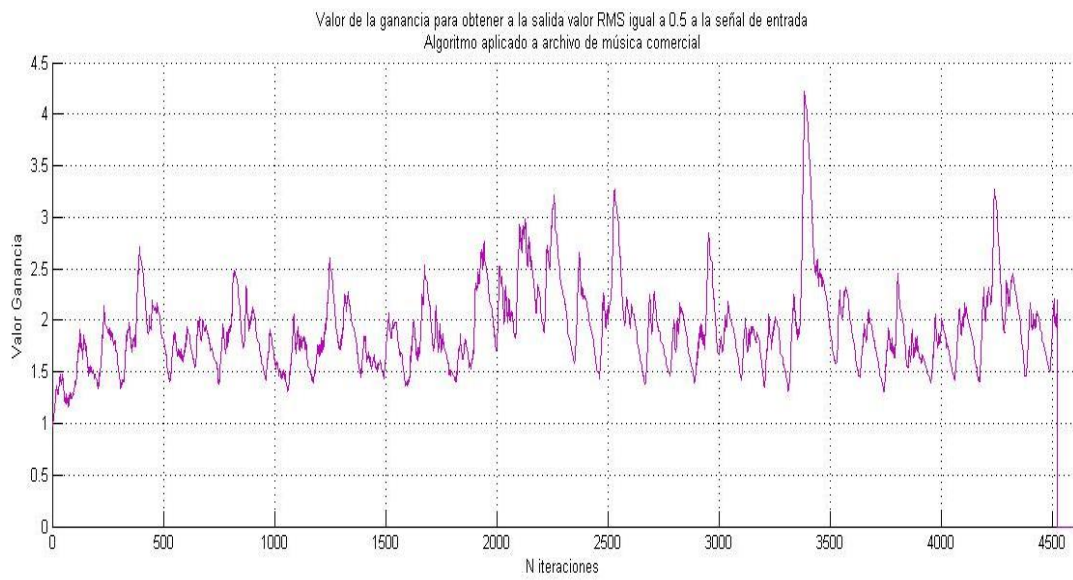


Figura IV.27 Valor de la Ganancia Caso 4

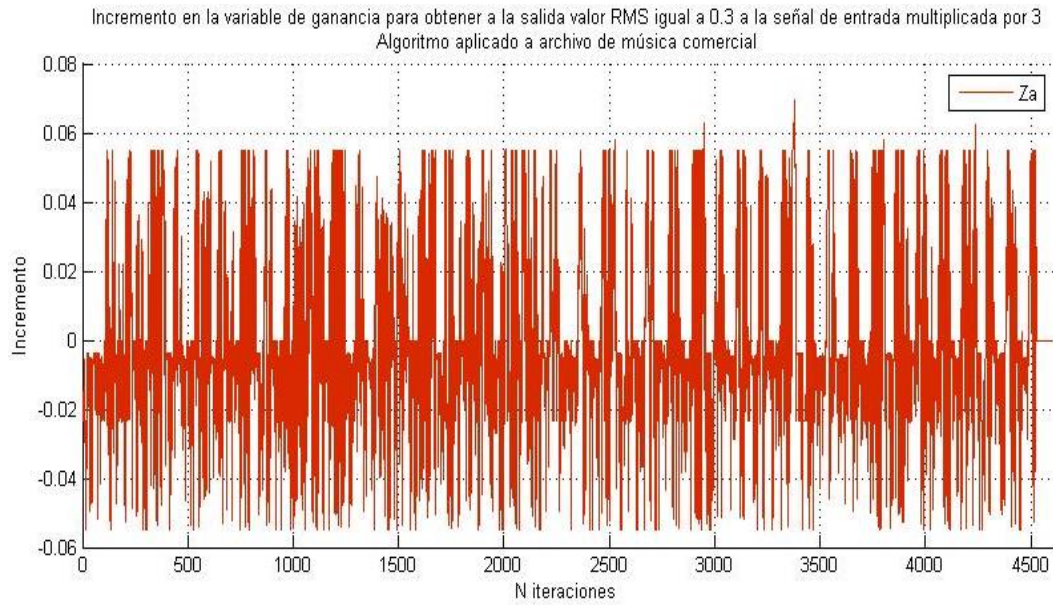


Figura IV.28 Incremento de la Ganancia Caso 5

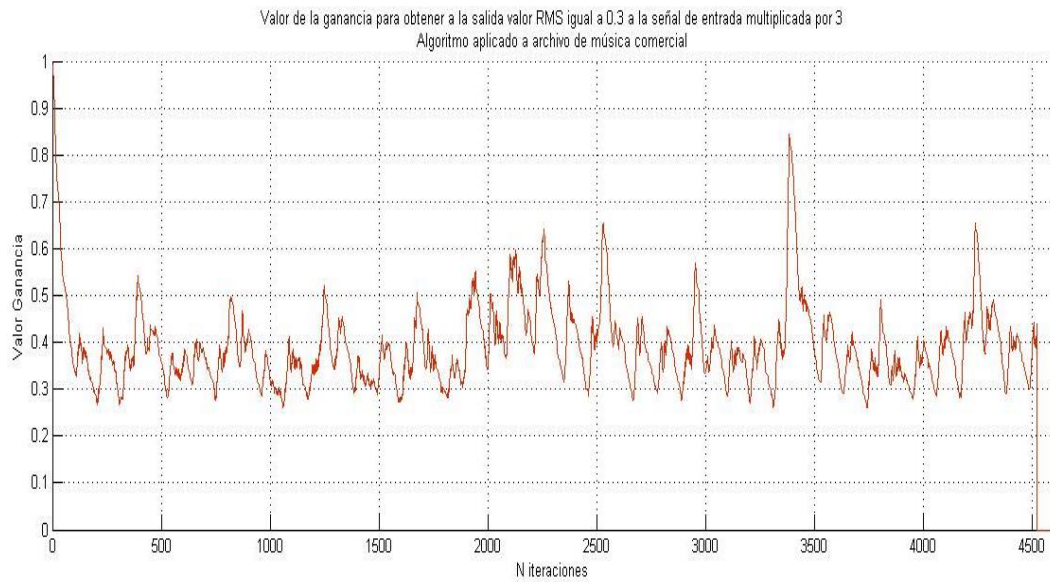


Figura IV.29 Valor de la Ganancia Caso 5

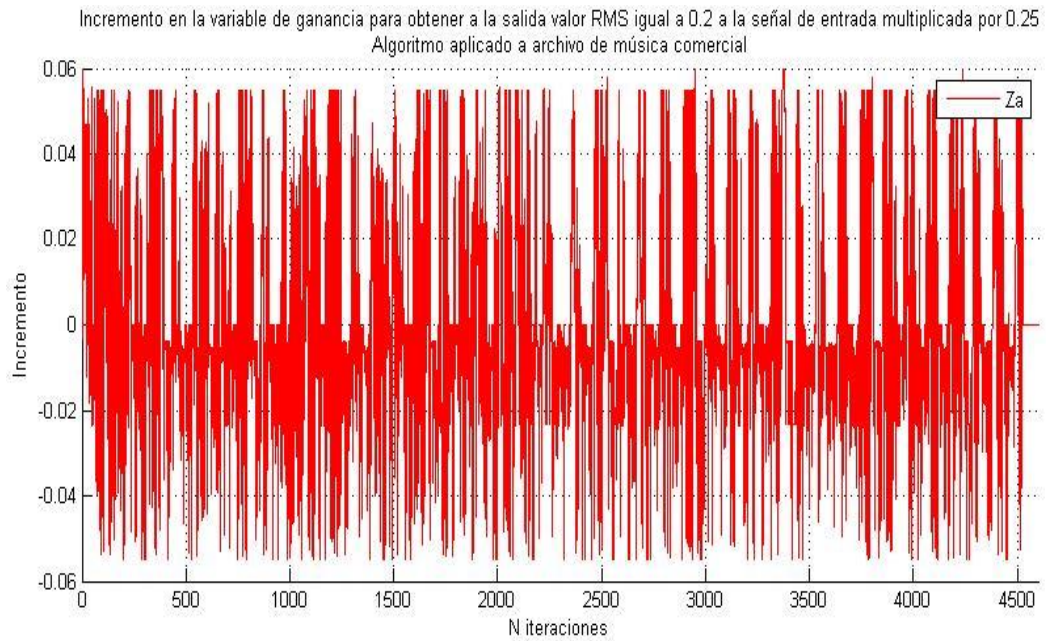


Figura IV.30 Incremento de la Ganancia Caso 6

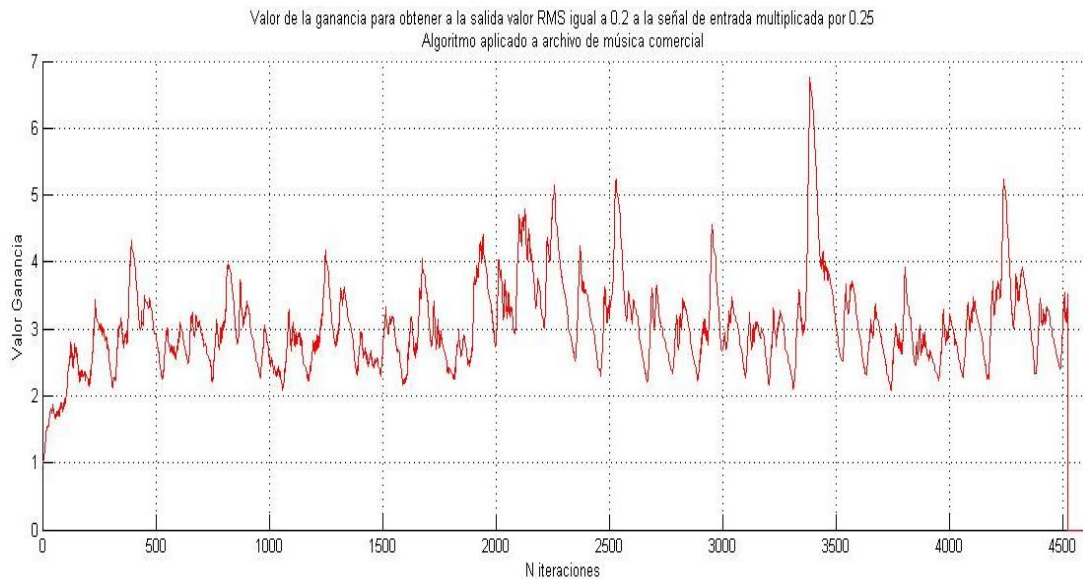


Figura IV.31 Valor de la Ganancia Caso 6

IV.3 ACOPLAMIENTO DE AMBOS BLOQUES

Los bloques AAD y AGC están acoplados en cascada, donde el amplificador toma como entrada la salida del bloque de detección de actividad de audio, para determinar el

nivel RMS deseado, para evaluar los resultados se hicieron tres pruebas: La primera con el objetivo de verificar que funciona el detector de actividad de audio en el acople de ambos bloques, se usó el mismo archivo que contiene los tres tipos de ruido y la señal de audio.

Se observa en la figura IV.32 que el bloque identifica fácilmente el ruido blanco y ruido rosa, mientras que para el ruido café su actividad de audio, en promedio, es de aproximadamente 0.75 y, para el archivo de audio, su valor promedio es mayor que para el ruido café pero los valores son parecidos al del ruido café.

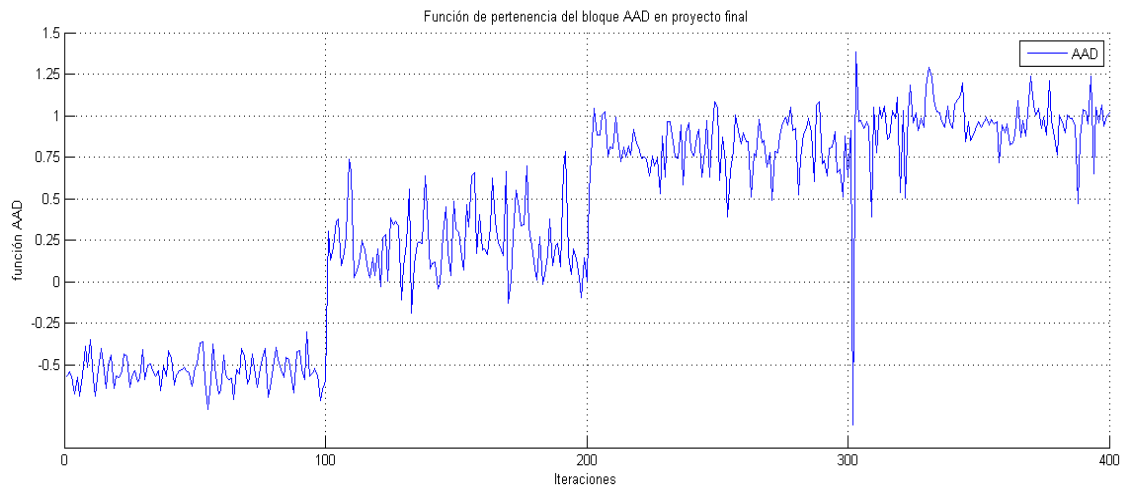


Figura IV.32 Función de pertenencia de detección de actividad de audio del bloque AAD del proyecto final

La segunda prueba tuvo como objetivo verificar el correcto comportamiento del amplificador de ganancia automática, y se usó el mismo archivo usado para el bloque por separado.

Se aplicó el algoritmo en los archivos de audio de campana tibetana y en el de música comercial, para la figura IV.33 muestra el valor RMS de a señal de entrada y la salida aplicado en el archivo de la campana tibetana, la señal no se manipuló y el valor RMS deseado fue de 0.5, en comparación de la figura IV.02 se aprecia que son iguales, por lo que el bloque funciona correctamente para el archivo probado.

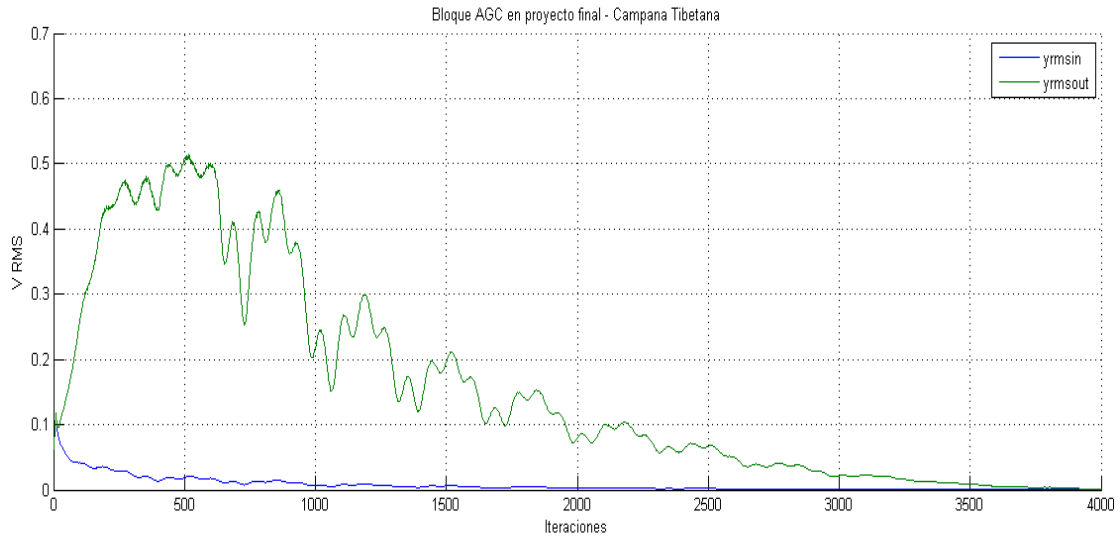


Figura IV.33 Valor RMS de entrada y salida para la el archivo de campana tibetana del bloque AGC del proyecto final

De mismo modo si se compara la figura IV.3.3 con la figura IV.2.16 se aprecia el mismo comportamiento a la salida, lo que nos reporta un comportamiento correcto

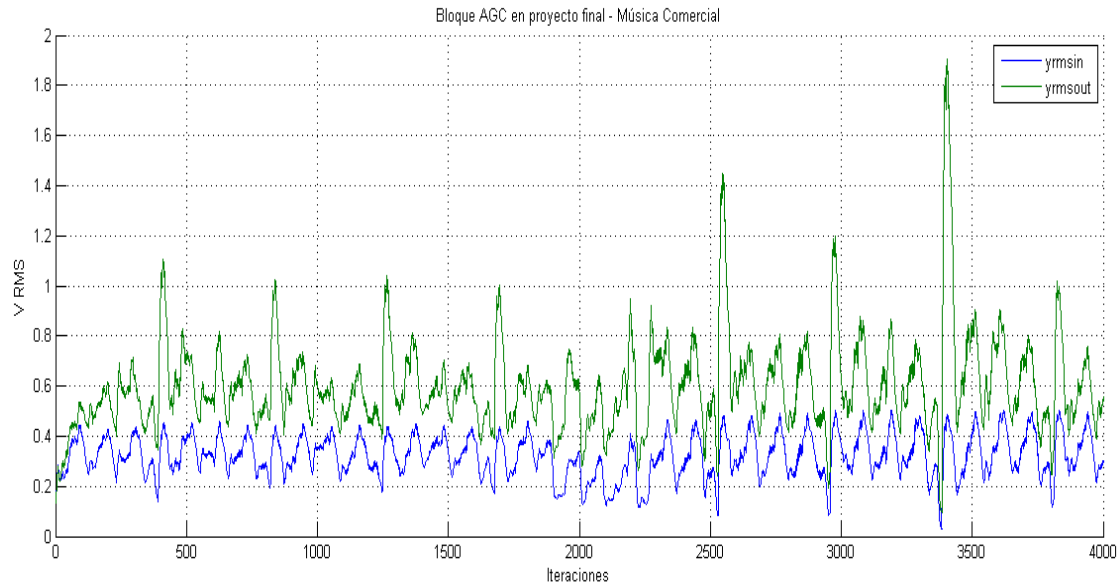


Figura IV.34 Valor RMS de entrada y salida para el archivo de música comercial del bloque AGC del proyecto final

Finalmente, para la prueba final, se hizo una colección de datos para ser analizados, la cual contiene muestras de siete archivos de audio y los tres tipos de ruido. El algoritmo toma 221,000 muestras de cada archivo, para hacer mil iteraciones por pista y, cada una, se realiza con doscientas veintiún muestras. Cada una de estas colecciones se multiplica por un número aleatorio entre 0 y 2 para que la amplitud cambie, finalmente, se conforman todas las pistas en un solo archivo, también se incluyen pistas de audio con ruido agregado, con el objetivo de verificar el funcionamiento del detector de actividad de audio y del amplificador de ganancia automática para que, con el valor RMS deseado y el valor de detección de actividad de audio, amplifique las muestras detectadas como audio y atenúe las detectadas como ruido, o audio más ruido.

El archivo está formado por doce partes entre ellas los tres tipos de ruido y diferentes tipos de audio como son música vocal, instrumental y algunos géneros musicales de música comercial, todos en formato WAV. A continuación se muestra el orden del archivo que se sometió al proyecto y la constante por la que se multiplica a la entrada:

1. Archivo de audio de artista Black Eyed Peas (C=1.8896)
2. Ruido Blanco (C=0.2224)
3. Archivo de audio de artista Coldplay (C=0.7385)
4. Ruido Rosa (C=1.5605)
5. Archivo de audio de artista Pentatonix (música vocal) (C=0.9785) con ruido café agregado (C=0.7795)
6. Ruido Café (C=0.7795)
7. Archivo de audio de Campana Tibetana (C=1)
8. Ruido Rosa (C=1.5605)
9. Archivo de audio de artista Brooklyn Duo (música instrumental) (C=0.9817)
10. Música de audio de artista Slash (C=0.6754) con ruido rosa agregado (C=1.5605)

11. Ruido Café (C=0.7795)

12. Archivo de audio de artista Leo Brouwer (música instrumental) (C=1.8001)

A continuación se muestra en la figura IV.35 la función de Detección de Actividad de Audio, en ella se observa que el bloque funciona bien para detectar los ruidos blanco y rosa, y el ruido café le es difícil identificarlo, al igual que algunos tipos de audio, donde debido a los instrumentos que se utilizan, generan frecuencias altas que hace que el bloque AAD los confunda con poca detección de actividad de audio.

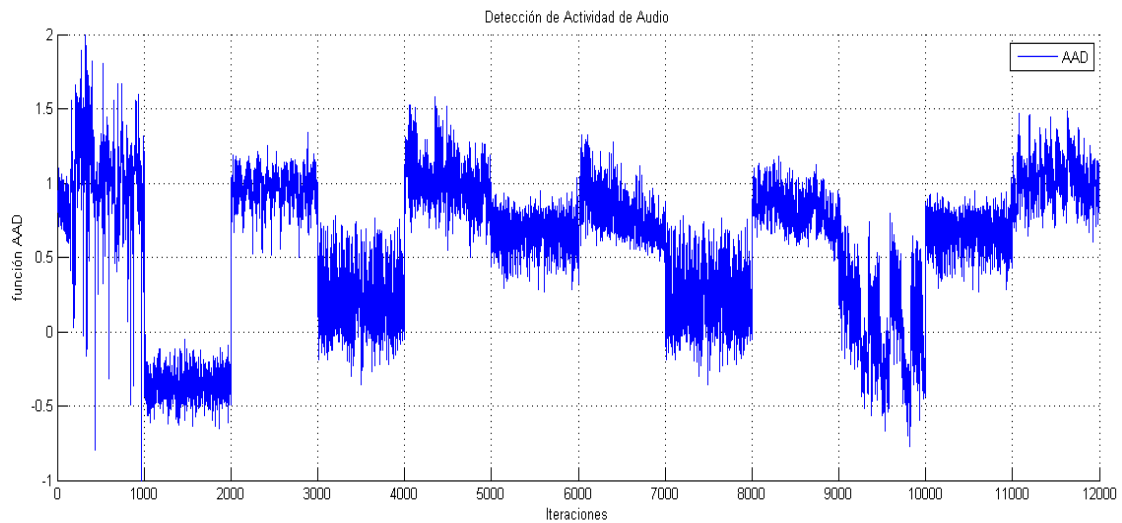


Figura IV.35 Prueba de proyecto completo: Función de Detección de Actividad de Audio

En la tabla 1 se aprecia el número de muestras que pertenecen al rango AAD esperado para cada caso:

Caso	Rango AAD esperado del caso	Número de muestras que están dentro del rango	Porcentaje del rango esperado
1	Mayor a 0.75	871	87.10%
2	Menor que 0.5	1000	100%
3	Mayor a 0.75	942	94.20%
4	Menor que 0.5	891	89.10%
5	0.5 – 0.75	36	3.66%
6	Menor que 0.5	85	8.50%

7	Mayor a 0.75	543	54.30%
8	Menor que 0.5	891	89.10%
9	Mayor a 0.75	717	71.70%
10	0.5 – 0.75	84	8.40%
11	Menor que 0.5	85	8.50%
12	Mayor a 0.75	974	97.40%

Tabla 1 Número de muestras y porcentaje del rango AAD esperado

En la tabla 2 se aprecia el valor promedio del AAD para cada uno de los 12 casos, así como también la desviación estándar para poder tener los valores máximos y mínimos que se pudieran permitir en la muestra, finalmente se agrega la columna con el valor de la mediana:

Caso	Valor Promedio	Desviación estándar	Valores máximos y mínimos	Mediana
1	1.0273	0.3092	0.7181 – 1.3365	1.0165
2	-0.3576	0.0935	-0.4511 – -0.2641	-0.3553
3	0.9665	0.1221	0.8444 – 1.0886	0.9844
4	0.2143	0.2066	0.0077 – 0.4209	0.1972
5	1.0297	0.1594	0.8703 – 1.1891	1.0318
6	0.6727	0.1175	0.5552 – 0.7902	0.6867
7	0.7914	0.1546	0.6368 – 0.9460	0.762
8	0.2143	0.2066	0.0077 – 0.4209	0.1972
9	0.8291	0.1244	0.7047 – 0.9535	0.8232
10	0.0447	0.3224	-0.2777 – 0.3671	0.048
11	0.6727	0.1175	0.5552 – 0.7902	0.6867
12	1.0268	0.1547	0.8721 – 1.1815	1.0213

Tabla 2 Valor Promedio, desviación estándar, valores máximos mínimos y mediana del AAD

Una vez que tenemos estos valores, se procede a contar las muestras que están dentro de rango del valor promedio más menos el valor de la desviación estándar, y de esta manera determinar la eficiencia.

Caso	Número de muestras dentro del rango máximo y mínimo	Eficiencia
1	578	57.80%
2	698	69.80%
3	709	70.90%
4	675	67.50%
5	702	70.20%
6	689	68.90%
7	694	69.40%
8	675	67.50%
9	668	66.80%
10	614	61.40%
11	689	68.90%
12	669	66.90%

Tabla 3 Eficiencia del acoplamiento entre bloques AAD y control automático de ganancia

Derivado del comportamiento del bloque AAD, y debido a que éste depende el valor de la ganancia cuando el AAD es menor a 0.75, la ganancia que se ve en la figura IV.3.5, no llega a amplificarse, además, la ganancia se reduce cuando no hay detección de actividad de audio y si el AAD está entre 0.5 y 0.75 se multiplica por el valor RMS.

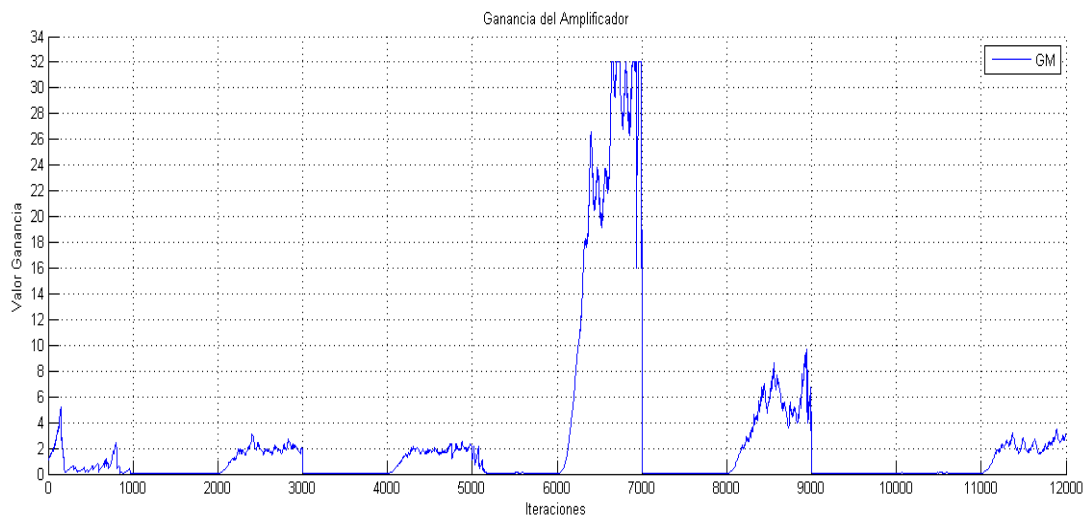


Fig. IV.36 Prueba de proyecto completo: Valor de la ganancia

Finalmente se muestra en la figura IV.3.6 los valores RMS de la señal de entrada y de la salida, se aprecia como cuando hay detección de ruido, el algoritmo se encarga de

suprimirlo ya que no es deseado a la salida, también es posible observar cómo en los casos de audio se logra amplificar y mantener el nivel RMS de la señal hasta el valor deseado.

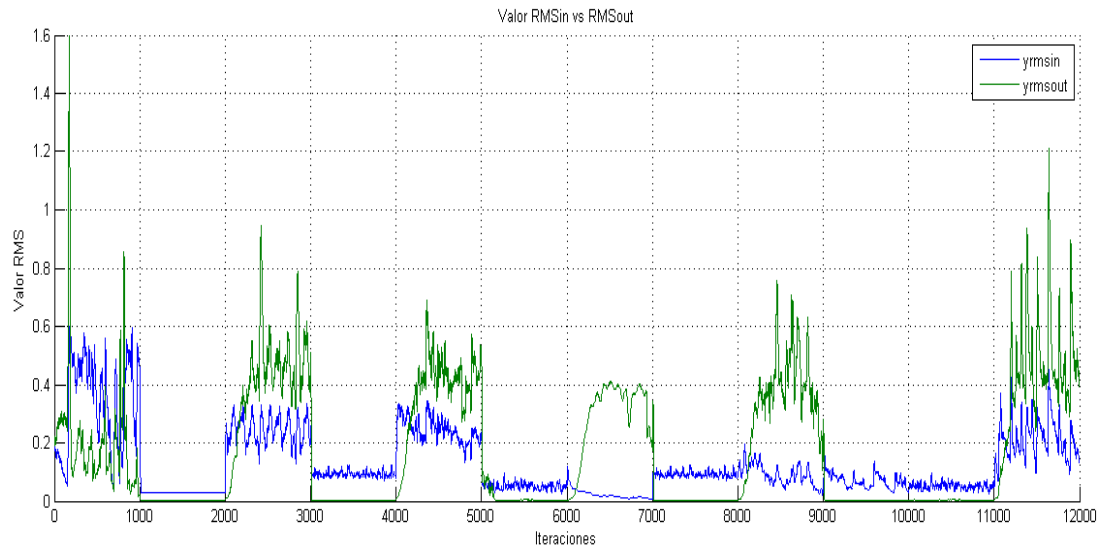


Fig. IV.37 Prueba de proyecto completo: Valor RMS de entrada y salida

V. CONCLUSIONES

El algoritmo del amplificador de audio, diseñado para mantener el nivel de volumen deseado por el usuario, cuya ganancia es controlada tomando en cuenta la actividad de audio que hay en la muestra, fue desarrollado a partir de dos bloques, cada uno con un controlador de inferencia difusa, el primero dedicado a la detección de actividad de audio, y el segundo al control de ganancia automática utilizando un control incremental a partir de los valores cuadrático medio (RMS) y de actividad de audio en la señal analizada. El algoritmo logra diferenciar de manera correcta en archivos de audio ruido blanco y rosa de pistas de música comercial, vocal e instrumental, no así para el ruido café, que por sus características parecidas a una señal de audio, su actividad de audio es parecida a algunos géneros musicales.

Una vez que se analizaron los resultados obtenidos del detector de audio se calcularon las dos medidas de las eficiencias para las de este algoritmo: para la primera, se calculó siguiendo el algoritmo que se utilizaba para definir la ganancia según el AAD, valores menores a 0.5 se clasifican como ruido, entre 0.5 y 0.75 como señal con ruido agregado y mayor a 0.75 como una señal sin ruido, se observa que para los ruidos blanco y rosa se identifica con facilidad como señal sin actividad de audio con una eficiencia del 100 y 89.1% respectivamente, lo que tiene como consecuencia, disminuir la ganancia para atenuar la señal a la salida. No así con el ruido café, con una eficiencia del 8.5% y con valores de AAD cercanos a 0.75, cuando el valor esperado es menor a 0.5, lo llega a clasificar como señal con ruido agregado.

Con respecto a señales de audio, se observó que, dependiendo del género musical, se tuvieron también diversos resultados, mientras que las pistas que en su instrumentación integran frecuencias más altas, como lo son las percusiones, el valor de AAD puede disminuir, produciendo el efecto de disminuir la ganancia para atenuar

la señal, mientras que en la música vocal e instrumental no se tuvo ese comportamiento.

Al analizar el promedio y desviación estándar para cada tipo de audio sometido al algoritmo, así, se obtuvo el rango de concentración de los valores de AAD para cada bloque, de esta manera también se evaluó la concentración de valores que tiene la salida del bloque de detección de actividad de audio para distintos tipos de audio. La eficiencia sobre la concentración de valores AAD varió entre un 57.8% el porcentaje menor para audio comercial, hasta 70.9% para ruido blanco, esto se puede ser debido a que la señal de audio, al ser una señal aleatoria, varía su comportamiento rápidamente, por lo que para cada colección de muestras analizadas se obtiene un valor AAD diferente, así mismo, se afecta de manera significativa al controlador de ganancia, ya que también produce una inestabilidad en el valor RMS de la salida, esto manifestándose en la pista de audio de la salida con variaciones de volumen.

El proyecto logró el objetivo de identificar actividad de audio en una señal dada, sin tener que hacer un análisis en frecuencia, así como también mantener un nivel de volumen deseado por el usuario. Sin embargo, hay puntos del proyecto que se pueden tener una mejora en un trabajo a futuro, entre ellas destacan:

- En la parte del bloque del Amplificador de Ganancia Automática, se apreció que el tiempo de respuesta para estabilizar el nivel deseado le toma varias iteraciones, además de que los cambios abruptos en la señal de entrada tiene como consecuencia que no se pueda mantener el nivel deseado, esto debido al comportamiento de las señales que procesa. Una mejora sería diseñar una nueva función incremental así como también funciones de pertenencia a la salida, de esta manera, para cada caso existiría un incremento o decremento diferente.

- Con respecto al bloque de Detección de Actividad de Audio, el valor AAD tuvo diferentes resultados cuando se somete con diferentes géneros musicales, así como también la discriminación del ruido café fue difícil; este efecto podría evitarse agregando a la entrada del controlador difuso una nueva variable para que también intervenga en las reglas de correspondencia, considerando observar el comportamiento que tiene cada género musical, así como también identificar diferencias con el ruido café. Y también podría ser de impacto positivo en el bloque de detección al usar filtros digitales del tipo IIR, para el cálculo de la energía en las diversas bandas.
- Agregar una etapa intermedia entre los bloques de detección de actividad de audio y control de ganancia automática, la función de este bloque sería asignar un valor RMS conveniente para el bloque de ganancia automática, tomando en cuenta el comportamiento que ha tenido el valor AAD así como también los incrementos que ha tenido la ganancia, el valor RMS deseado y el de la salida o la señal de error, pudiendo tener en esta etapa un controlador difuso en el cual se tendría la interacción de las variables mencionadas por medio de diversas reglas de correspondencia, de esta manera, el objetivo de esta etapa sería evitar incrementos o decrementos abruptos al nivel RMS de la salida.

VI. REFERENCIAS

- [1] Sector de Radiocomunicaciones de la UIT, «Algoritmos para medir la sonoridad de los programas radiofónicos y el nivel de cresta de audio real,» Unión Internacional de Telecomunicaciones, Ginebra, 2015.
- [2] F. Beritelli, S. Casale y A. Cavallaro, «A Robust Voice Activity Detection for Wireless Communications Using Soft Computing,» *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, pp. 1818-1829, 1998.
- [3] J. Jantzen, *Foundations of Fuzzy Control*, England: John Wiley & Sons, Ltd, 2007.
- [4] D. Davis y C. Davis, *Sound System Engineering*, Indianapolis, Indiana: Howard W. Sams & Co., Inc., 1982.
- [5] M. A. Ibarra Carrillo, *Procesamiento Digital de Señales*, Facultad de Ingeniería; Telecomunicaciones, 2012.
- [6] L. Escobar Salguero, *Diseño de filtros digitales*, Ciudad Universitaria, CDMX: Departamento de Publicaciones de la Facultad de Ingeniería, 2006.
- [7] J. G. Proakis y D. G. Monolakis, *DIGITAL SIGNAL PROCESSING Principles, Algorithms, and Applications*, New Jersey: Prentice Hall, 1996.
- [8] M. F. J. Meca, «Ruido Intrínseco en Dispositivos Electrónicos,» UAH, Madrid, 2007.
- [9] C. González Morcillo, «Lógica Difusa, una introducción práctica. Técnicas de Softcomputing,» UCLM, Castilla-La Mancha, 2011.
- [10] T. J. Ross, *Fuzzy Logic with Engineering Applications*, New Mexico: John Wiley & Sons, Ltd, 2004.
- [11] D. Jorge Matich, *Redes Neuronales: Conceptos Básicos y Aplicaciones*, Rosario: Universidad Tecnológica Nacional – Facultad Regional Rosario, Departamento de Ingeniería Química, GIAIQ, 2001.
- [12] F. Lara Rosano, «FUNDAMENTOS DE REDES NEURONALES ARTIFICIALES,» UNAM, Facultad de Ingeniería, UNAM, 1996.
- [13] J. P. Alegre-Pérez, B. Calvo-López y S. Calvo-Pueyo, *Automatic Gain Control: Techniques and architectures fo RF recivers*, New York, NY: Springer, 2011.
- [14] G. Kang y M. Lidd, «Automatic gain control,» *IEEE ICASSP*, pp. 1-4, 1984.
- [15] R. Harjani, «A low-power CMOS VGA for 50 Mb/s disk drive read channels,» *IEEE T Circuits II*, vol. 42, nº 6, pp. 370-376, 1995.

- [16] R. K. Singh, R. Dhiman y R. Chandel, «Design and analysis of a novel automatic gain control preamplifier circuit for hearing aid device,» *IEEE CONNECT*, pp. 1-6, 2015.
- [17] J. Silva-Martínez, S. Solís-Bustos, J. Salcedo-Suñer, R. Rojas-Hernández y M. Schellenberg, «A CMOS hearing aid device,» *Analog Integr Circ S*, vol. 21, pp. 163-172, 1999.
- [18] S.-P. M. D.-S. A. Moro-Frías D., «Diseño y simulación de un circuito de control automático de ganancia para sistemas sensores,» *Superficies y Vacío*, vol. 25, nº 1, pp. 36-42, 2012.
- [19] D. V. Mercy, «A review of automatic gain control theory,» *IET Radio and Electronic Engineer*, pp. 579-590, 1981.
- [20] Y. Z. L. J. Chengyu Wen, «Design and implementation of fast convergence and larga dynamic digital automatic gain control,» *IEEE ICCT*, pp. 767-770, 2011.
- [21] G. R. Steber, «Digital Signal Processing In Automatic Gain Control Systems,» *IEEE IECON*, pp. 381-384, 1988.
- [22] J. Wang, X. Lu y L. Zhang, «Modeling of a multiple digital automatic gain control system,» *Tsinghua Science and Technology*, vol. 13, nº 6, pp. 807 - 811, 2008.
- [23] A. Benyassine, E. Shlomot y H.-Y. Su, «ITU-T Recommendation G.729 Annex B: A Silence Compression Scheme for Use with G.729 Optimized for V.70 Digital Simultaneous Voice and Data Applications,» *IEEE ommunications Magazine*, pp. 64-73, 1977.
- [24] B. Francesco, S. Casale y S. Serrano, «Peformance Evaluation and Comparison of G.729/AMR/Fuzzy Voice Activity Detectors,» *IEEE SIGNAL PROCESSING LETTERS*, pp. 85-88, 2002.