

Control automático de ganancia con corrección en tiempo real

Universidad Tecnológica Nacional-FRBA

1st Nicolas Tobias Alamaraz
Universidad Tecnológica Nacional-FRBA
CABA, Buenos Aires, Argentina
nalmaraz@frba.utn.edu.ar

3rd Axel Nathanel Nahum
Universidad Tecnológica Nacional-FRBA
CABA, Buenos Aires, Argentina
anahum@frba.utn.edu.ar

2nd Guido Ignacio Glorioso Rego
Universidad Tecnológica Nacional-FRBA
CABA, Buenos Aires, Argentina
guido.glorioso@frba.utn.edu.ar

4th Santiago Eduardo Palozzo
Universidad Tecnológica Nacional-FRBA
CABA, Buenos Aires, Argentina
spalozzo@frba.utn.edu.ar

Abstract—Con el avance de la tecnología, los sistemas de audio en general han dejado de ser sólo exclusivos para ámbitos comerciales o industriales (por ejemplo cines o teatros). Podemos encontrarlos, por ejemplo, en equipos sound round, parlantes plug and play o televisores. Es por esto que el desarrollo de herramientas para controlar el audio cobra especial sentido.

En este caso se desarrolló y caracterizó un sistema de control de loudness. Es decir, un dispositivo que regule la ganancia del sistema de tal modo que independientemente de la entrada, siempre a la salida se obtiene el mismo nivel de sonoridad.

Para medir loudness se utilizó la norma ITU-R BS.1770. Por otro lado, para caracterizar al sistema se determinó el nivel de distorsión armónica (THD) que aporta. A tal fin se probaron diferentes configuraciones de loudness para la entrada. El THD fue medido tanto en la salida digital (I2S) como en la salida analógica (sobre el parlante). Se obtuvo que el caso de mayor distorsión es de 1.52% en la salida analógica para un set point de 60 LKFS y una frecuencia de 500Hz.

Index Terms—Loudness, Sistema de control, THD, AGC

I. INTRODUCCIÓN

A la hora de reproducir una pista de audio muchas veces ocurre que dependiendo del dispositivo que grabó el sonido, o bien debido al post-procesamiento que se le aplique, existe un gran contraste entre los niveles de potencia de un stream y el otro. Por ejemplo, si reproducimos una canción grabada en un estudio y luego reproducimos un recital grabado en vivo de forma casera probablemente el nivel de sonido que perciba el oído humano sea diferente para ambas pistas de audio. Por lo tanto, esta situación genera una constante corrección manual por parte del usuario en el nivel de volumen del reproductor para ajustar la potencia de audio cada vez que cambie de pista. Entonces, surge la necesidad de realizar esta corrección de manera automática, buscando una mayor comodidad y mejor experiencia de usuario.

La variable física que mide la potencia de audio que percibe el humano se denomina sonoridad (en inglés *loudness*) y se mide en LKFS (*Loudness, K-weighted, relative to Full Scale*). Esta variable está definida por la norma **ITU-R BS.1770** [1] junto con su algoritmo de medición.

Hoy en día es común encontrar sistemas de control de Loudness en aplicaciones más bien industriales. Por ejemplo, en emisoras de radio, estudios de grabación, equipos de transmisión de audio en vivo o en distintos software de remasterización. Es decir están abocados a un uso más bien profesional e industrial. Aquí priman marcas como Orban [2], SSL [3], DBX [4]. Sin embargo se pueden llegar a encontrar en algunos sistemas de audio hobbistas, por ejemplo, en procesadores de audio domésticos o distintas aplicaciones de audio. Algunos ejemplos pueden ser procesadores Behringer [5] y Yamaha [6] para sistemas home theater. En base a que no hay tanta oferta de estos sistemas en equipos **plug and play** de bajo costo para usuarios aficionados (no profesionales) es que se decidió desarrollar un sistema de control de loudness para un procesador de audio de propia autoría.

A. ¿Que es el loudness?

Según la **norma** antes mencionada, el nivel de **loudness** de una señal se refiere a la percepción subjetiva de la intensidad o volumen de un sonido por parte de un oyente. En otras palabras, es la medida de cuán fuerte o suave se escucha una señal de audio en función de cómo la percibe el oído humano. El **loudness** no se mide directamente en decibelios (dB), como lo haría el nivel de presión sonora, sino que tiene en cuenta la sensibilidad del oído humano a diferentes frecuencias y ajusta la medición en consecuencia.

B. ¿Cómo se mide el loudness?

Como ya mencionamos al comienzo de esta sección, la norma ITU-R BS.1770 propone un algoritmo para la medición de loudness de un stream de audio (medido en LKFS).

De esta manera, el objetivo es cuantificar el nivel de **loudness** de una señal de manera representativa para el oído humano y en base a ello ajustar el nivel de sonoridad. Este ajuste no debe ser destructivo para la señal, buscando tener la menor distorsión posible. Sin embargo, hay que tener en cuenta que dicha corrección implica una ganancia variable en el tiempo, dando lugar a posibles distorsiones en la señal.

II. METODOLOGÍA GENERAL

El sistema de control tiene un diagrama en bloques como el siguiente (**fig. 1**):

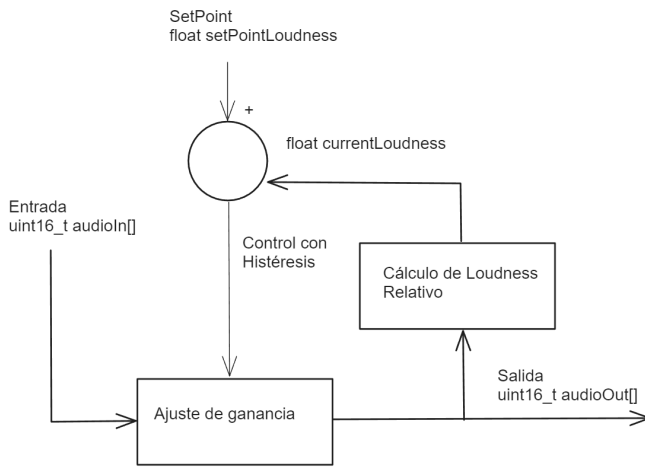


Fig. 1. Diagrama en bloques

Las etapas son:

- **Entrada:** señal de audio de entrada
- **Ajuste de ganancia:** corrección en base al loudness con un nivel de histéresis
- **Set Point:** valor de referencia de Loudness configurado por el usuario
- **Cálculo de loudness relativo:** Este es el algoritmo que plantea la norma para la medición de Loudness
- **Salida de audio:** señal de audio con la corrección pertinente

En este sistema se recibe un stream de audio vía Bluetooth [7] y se le aplica el algoritmo de medición de loudness. Luego, en función del valor obtenido se le aplica una ganancia que amplifique o atenúe la señal para que el nivel de loudness se mantenga en el valor configurado por el usuario.

A. Algoritmo de medición de Loudness

Para medir el nivel de sonoridad la norma ITU indica los siguientes pasos:

Primero, aplicarle los filtros que adapten la señal ponderando en función del comportamiento del oído humano (filtros digitales con coeficientes definidos por la norma a 48kHz). Este filtro es aplicado a cada canal.

Segundo, se calcula la potencia de la señal dividida en bloques de tiempo con un solapamiento determinado (recomienda bloque de 0.4 segundos y 75% de solapamiento). En otras palabras se toman los últimos 0.4 segundos de información y se calcula la potencia. Luego, se avanza 0.1 segundos y se vuelve a calcular la potencia de los siguientes 0.4 segundos. Se repetirá el proceso con la totalidad del stream de audio.

Tercero, se le aplica la siguiente expresión a cada bloque de potencia calculado.

$$L_{KG} = -0.691 + 10 \log_{10} \sum_i G_i \left(\frac{1}{|J_g|} \sum_{J_g} Z_{ij} \right) - 10LKFS \quad (1)$$

Donde:

- i : Representa los distintos canales. Por ejemplo canal izquierdo, derecho, sub-woofer, etc
- G_i : coeficientes de ponderación de cada canal
- Z_{ij} : Energía de la señal en un tiempo de gate para el canal i en el bloque de tiempo j
- J_g : Son los elementos que superan el umbral de disparo de 70 LKFS
- $|J_g|$: Cantidad de elementos que superan el umbral de disparo

B. Adaptación del algoritmo para utilizarse dentro de un microcontrolador en tiempo real

Para adaptar el algoritmo al microcontrolador es necesario resolver dos problemáticas principalmente. La primera es que no es un algoritmo nativo para aplicarse en tiempo real y la segunda es que el hecho de tener que almacenar grandes tiempos de audio es sumamente inviable para un microcontrolador que está muy limitado en memoria.

Para resolver la primera problemática se adoptó por implementar un buffer circular que almacene los últimos 3 segundos de audio. Estos son los que representan el período total del cálculo. Es decir, se aplicaron las ventanas de 0.4 segundos sobre este buffer de 3 segundos.

Por otro lado, al tener que resolver una integral de períodos de 0.4 segundos de manera reiterativa se optó por definir sub-ventanas de 0.1 segundos. De esta manera se redujo el costo computacional, dado que el único procesamiento que debería hacer el sistema es calcular la energía de los últimos 0.1 segundos y aprovechar todo el procesamiento previo para así ir formando los bloques de 0.4 segundos. Puede verse gráficamente este proceso con la siguiente imagen (fig. 2)

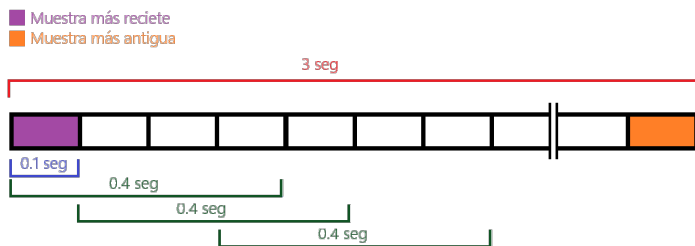


Fig. 2. Enter Caption

C. Ajuste de ganancia

Luego de analizar el desempeño de distintos algoritmos de corrección, se optó por un agregar un nivel de histéresis H al sistema de forma que una vez conseguido un equilibrio no varíe la ganancia por aumentos pequeños de señal. El valor de H quedó definido en 5 LFKS de manera empírica ya que conseguía los mejores resultados.

D. Diagrama en bloques del Hardware

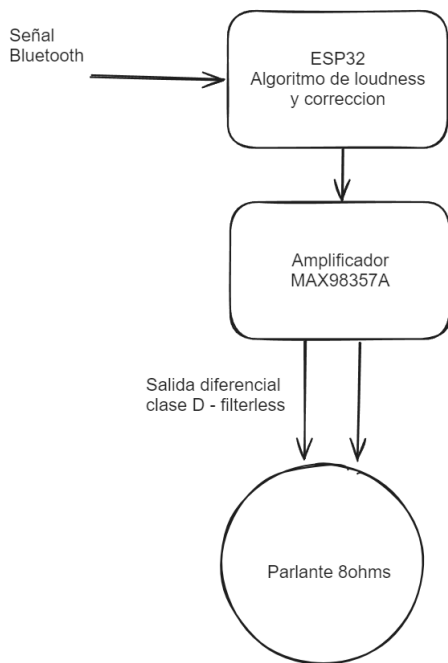


Fig. 3. Hardware

(fig. 3)

En cuanto a Hardware se tiene un microcontrolador **ESP32** [8] utilizado para desarrollar el sistema de control. Mediante este, se gestiona la recepción de la señal de audio vía Bluetooth y aplica el proceso de medición de loudness para así definir un valor de ganancia.

En cuanto a la salida, el microcontrolador entrega una señal I2S [9] que es recibida por un módulo amplificador

MAX98357A [10]. Estese encarga de recibir la señal I2S, decodificarla y amplificarla con un amplificador clase D.

E. Interfaz de usuario

Para controlar el SetPoint se diseñó una aplicación de PC que se conecta con el microcontrolador vía **UART** [11] y permite además graficar los valores de loudness en tiempo real, además de ir guardándolos en un archivo **CSV** (Comma-Separated Values).

Selección de puerto y loudness:

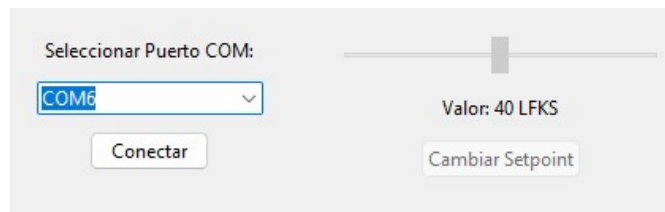


Fig. 4. Selección de puerto y loudness del sistema

En primer lugar se observa la selección del **puerto** y el **loudness** (fig. 4). Esta misma funciona de la siguiente manera:

Selección de puerto: Se debe seleccionar el puerto, una vez conectado el sistema a la PC. Al pulsar el botón conectar, ya se podrá actuar sobre el sistema.

Cambiar SetPoint: Se elije el valor de **loudness** que se requiera para el sistema. Este se enviará al microcontrolador mediante la comunicación serie **UART**.

Iniciar/parar ploteo: comienza o termina el registro de datos ploteando los últimos 30 segundos y guardando el registro completo en un archivo CSV (fig. 5).

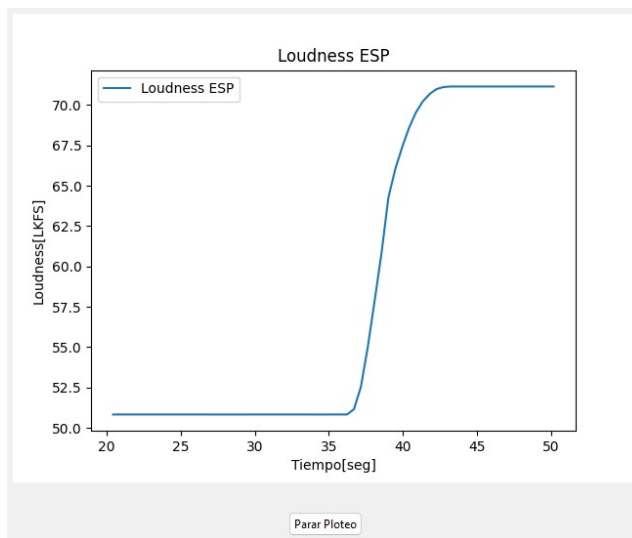


Fig. 5. Grafico de Loudness en funcion del tiempo

III. PROCESAMIENTO DE LOS DATOS

A. Magnitud a medir

Dado que hay una ganancia variable en función del tiempo, es de interés conocer el THD (Total harmonic distortion) [12] del sistema. Por lo tanto, se realizó la medición de THD para diferentes valores de loudness teniendo como señal de entrada un tono senoidal puro a distintas frecuencias.

B. Esquema de medición

El primer punto de medición del sistema se encuentra a la salida del microcontrolador. Esta señal es del tipo I2S, por lo tanto se mide con un analizador lógico. El segundo punto de medición es a la salida del amplificador. El mismo cuenta con una salida tipo *bridge* o diferencial, con lo cual es necesario medir la señal de ambos bornes de entrada del parlante. Para las mediciones realizadas se adoptó el esquema de conexión de la **figura 6**.

C. Tonos de prueba

Para caracterizar el sistema, se utilizó una señal de entrada senoidal a distintas frecuencias (500, 1000 y 10.000 Hz) y distintos valores de loudness (50, 60 y 70 LFKS). Se eligieron estos valores debido a que cubren gran parte del espectro audible [13] (que para un oído joven y sano va de 20Hz a 20kHz) y por otro lado, con los valores de referencias adoptados se cubren los niveles de Loudness más frecuentes para un usuario promedio.

D. Adaptación de salida del modulo amplificador

Un aspecto a tener en cuenta sobre las mediciones es que los amplificadores **MAX98357** son del tipo **clase D** y debido a su principio de funcionamiento en la salida se presenta una portadora de 300KHz combinada con la señal de audio. Esto implica que para recuperar la información hay que filtrar la portadora. Una vez llegado a este punto que se tiene el stream de audio limpio se le aplica el algoritmo de THD para evaluar la distorsión.

Para visualizar el comportamiento del amplificador clase D, se calculó la **FFT** [14] (Fast Fourier Transform) a la señal de salida del sistema antes y después del filtrado para observar las componentes frecuenciales que esta presenta.

E. Instrumentos y hardware utilizado

Para las mediciones se utilizaron los siguientes equipos de medición:

- Analizador lógico Duatek LA2408 [15]
- Osciloscopio Agilent MSO-X 3024A [16]

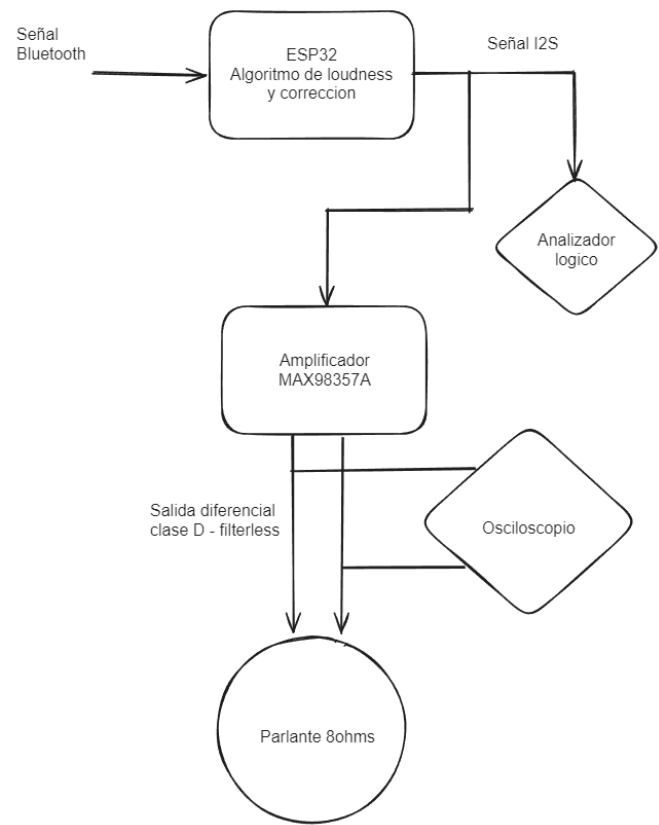


Fig. 6. Esquema de conexión del banco de medición

IV. RESULTADOS

Se adjuntan los diferentes gráficos intermedios antes de obtener el THD de las señales para la señal de prueba de 500Hz con un setpoint de 60LKFS tanto para la señal analógica como para la digital. Cabe destacar que este proceso debe repetirse para cada una de las mediciones, es decir, todas las frecuencias con todos los set points. Para visualizar todos los gráficos de cada medición puede consultarse en [2] (**Apéndice**).

A. Señal analógica a 500Hz con 60 LKFS

En la fig. 7 se observa la señal de ambos canales referidas a tierra. Este oscilograma es para mostrar que ambas señales por separado no permiten la reconstrucción de la señal original.

En la fig. 8 se muestra la diferencia entre la señal de ambos canales del osciloscopio. Puede verse que se tiene la portadora de una muy alta frecuencia y la señal que modula a dicha portadora.

Por otro lado, en la fig. 9 puede verse la FFT de la señal diferencial anterior. Se observa que en las bajas frecuencias está la senoidal de interés y en 150kHz está la señal portadora. Haciendo pruebas puede verse que esta portadora es una señal cuadrada y dependiendo de cuál es la señal de información

puede tener entre 150 y 300kHz. Cabe destacar que esto no supone un problema en cuanto a la calidad de sonido dado que está muy por encima del rango audible.

Finalmente, en la fig 10 se observa la señal diferencial luego de haberse filtrado con una frecuencia de corte de 20kHz y en la fig. 11 su correspondiente FFT. En este último gráfico está la portadora mucho más atenuada que la información.

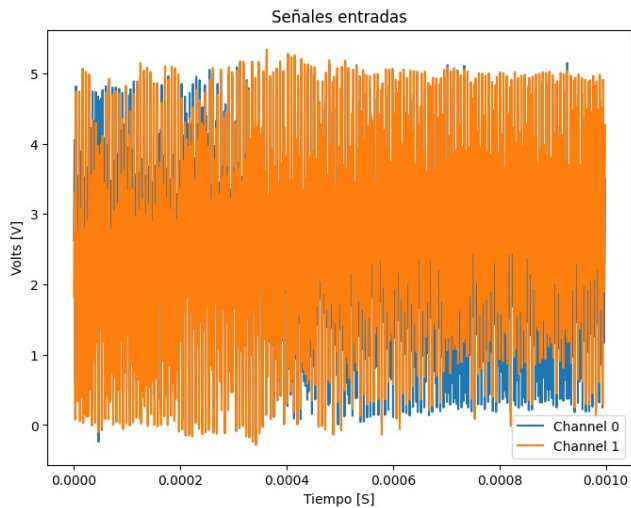


Fig. 7. Señales independientes de cada canal del osciloscopio

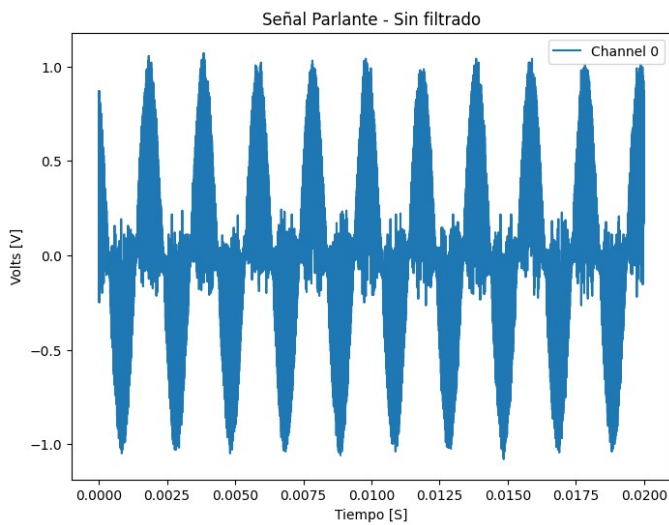


Fig. 8. Señal diferencial sin filtrar

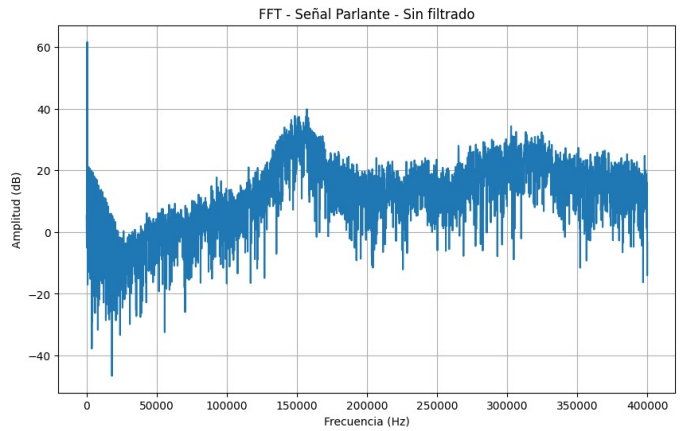


Fig. 9. FFT Señal sin filtrar

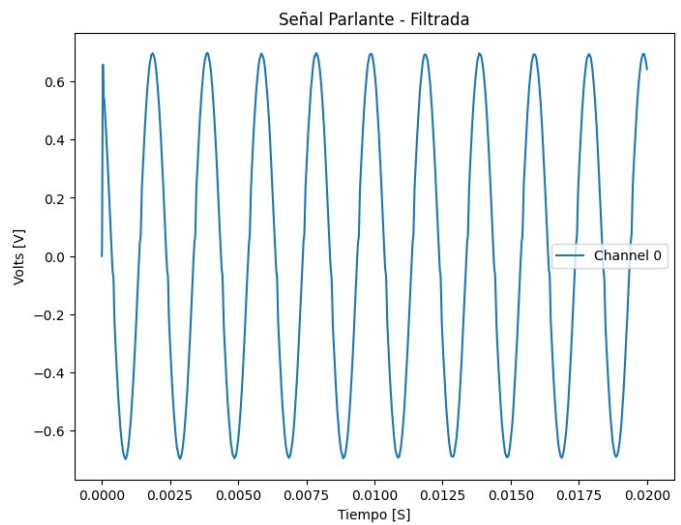


Fig. 10. Señal filtrada

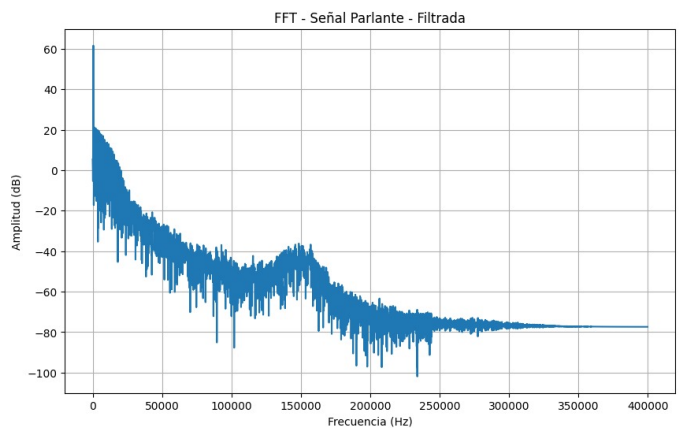


Fig. 11. FFT de la señal filtrada

B. Señal digital a 500Hz 60 LKFS

En la fig. 12 se observa la señal temporal medida con el analizador lógico a la salida del microcontrolador y en la fig. 13 se observa su correspondiente FFT.

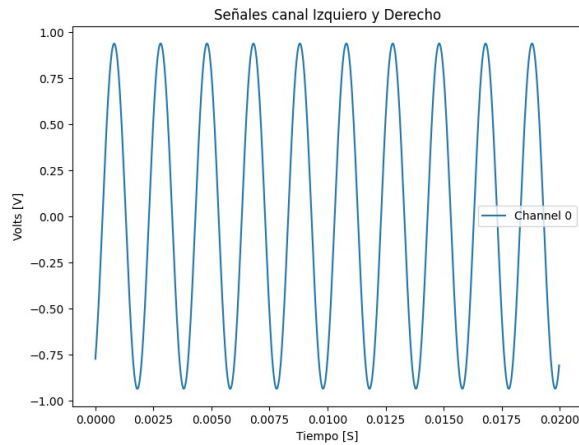


Fig. 12. Señal I2S

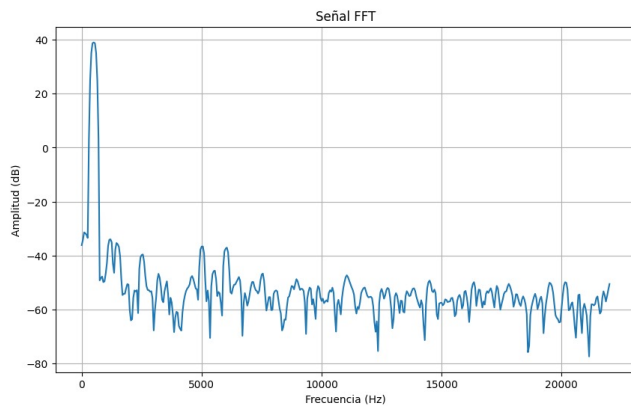


Fig. 13. FFT de la señal I2S

C. Tabla con los resultados obtenidos de THD

Analizando cada una de las señales senoidales con cada uno de los set-points se obtubieron los resultados de la tabla 1 (resultados señal I2S) y tabla 2 (resultados señal analógica).

Parámetro	70 LKFS	60 LKFS	50 LKFS
THD % - ESP32 (500Hz)	0,548%	0,235%	0,527%
THD % - ESP32 (1KHz)	0,453%	0,354%	0,202%
THD % - ESP32 (10KHz)	0,059%	0,124%	0,229%

TABLE I

Parámetro	70 LKFS	60 LKFS	50 LKFS
THD % - Parlante (500Hz)	1,25%	1,52%	1,03%
THD % - Parlante (1KHz)	0,56%	0,97%	0,9%
THD % - Parlante (10KHz)	0,33%	0,33%	0,33%

TABLE II

V. ANÁLISIS DE LOS RESULTADOS Y CONCLUSIONES

En base a las tablas de resultados (tabla 1 y tabla 2), se observa que la tasa de distorsión aumenta para señales de menor frecuencia y además aumenta cuando estamos en un rango de LKFS cercano a 60.

Por otro lado, puede verse que el mayor aporte de distorsión lo hace el módulo amplificador. Esto podemos verlo en la comparación de los THD de la señal digital y la analógica respectivamente.

Por último, el peor caso de distorsión se dio con una señal de entrada de 500Hz con un nivel de loudness de 60 LKFS obteniendo un THD de 1.52%.

VI. DESARROLLO A FUTURO

Una de las mejoras a futuro resulta ser la medición del THD en el transitorio, es decir, mientras que la señal se este estabilizando entre dos puntos de trabajo (porque cambio el setpoint). Esto se debe realizar mediante la salida digital debido a que el transitorio dura aproximadamente unos 10 segundos. Esto supone un problema para medirse con el osciloscopio debido la profundidad de memoria (no se podrá realizar mediante la salida analógica).

Por otro lado, se podría realizar un gráfico de THD en función del setpoint de loudness elegido y parametrizando diferentes frecuencias. Para conseguir esto de una manera mas precisa se deben realizar las mediciones para una mayor cantidad de setpoints y frecuencias de trabajo.

APÉNDICE

- [1] [NBviewer de Loudness](#)
- [2] [NBviewer de resultados](#)
- [3] [NBviewer de datos](#)

REFERENCES

- [1] [Norma ITU-R BS.1770](#)
- [2] [Orban](#)
- [3] [SSL](#)
- [4] [DBX](#)
- [5] [Procesadores Behringer](#)

- [6] [Procesadores Yamaha](#)
- [7] [Bluetooth-A2DP-API-Application-Example](#)
- [8] [ESP32](#)
- [9] [Inter-IC Sound \(I2S\)](#)
- [10] [Datasheet del amplificador MAX98357](#)
- [11] [Universal Asynchronous Receiver/Transmitter \(UART\)](#)
- [12] [Distorsión armónica total \(THD\)](#)
- [13] [Espectro audible](#)
- [14] [FFT](#)
- [15] [Analizador lógico Daitek LA2408](#)
- [16] [Osciloscopio Agilent MSO-X 3024A](#)