

CONESCAPANHONDURAS2025paper143.pdf

 Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477703889

Submission Date

Jul 31, 2025, 9:56 PM CST

Download Date

Aug 12, 2025, 6:32 PM CST

File Name

CONESCAPANHONDURAS2025paper143.pdf

File Size

513.4 KB

5 Pages




2,319 Words

12,738 Characters

12% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Top Sources

- 12%  Internet sources
- 2%  Publications
- 0%  Submitted works (Student Papers)

Integrity Flags




0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Top Sources

- 12%  Internet sources
- 2%  Publications
- 0%  Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
www.coursehero.com		2%
2	Internet	
bumotors.ru		1%
3	Internet	
www.ijew.io		<1%
4	Internet	
es.slideshare.net		<1%
5	Internet	
www.vtomb.com		<1%
6	Internet	
revistas.itm.edu.co		<1%
7	Internet	
documents.mx		<1%
8	Internet	
repositorio.uta.edu.ec		<1%
9	Internet	
www.youtube.com		<1%
10	Internet	
idoc.pub		<1%
11	Internet	
zagan.unizar.es		<1%

12	Internet	fr.slideshare.net	<1%
13	Internet	livrosdeamor.com.br	<1%
14	Internet	oa.upm.es	<1%
15	Internet	www.tek.com	<1%
16	Internet	docplayer.es	<1%
17	Internet	www.slideshare.net	<1%
18	Internet	www.witpress.com	<1%
19	Internet	qdoc.tips	<1%
20	Internet	www.researchgate.net	<1%

Prototipo de sistema interactivo de síntesis de sonido controlado por gestos

Abstracto— Este artículo presenta un sistema interactivo desarrollado en Max 8 que genera sonido a partir de gestos manuales. Utiliza entrada MIDI para mapear datos de movimiento a parámetros sonoros, controlando un motor de síntesis aditiva. El diseño permite modificar funciones de mapeo e incorporación de módulos como modulación de amplitud.

Palabras claves — Max 8, Mapping, Síntesis aditiva, modulación, muestreo, FFT.

I. INTRODUCCIÓN

La conexión entre el cuerpo y el sonido ha atraído la atención de diseñadores y programadores por un largo periodo. Si anteriormente la interacción musical se basaba casi siempre en el contacto físico con un instrumento, las tecnologías posibilitan formas de expresión más libres, en las que el sencillo gesto puede adquirir relevancia como medio de control y generación de sonido [1]. Hoy en día, existen sistemas que transforman estos movimientos en sonidos o imágenes, proporcionando experiencias más inmersivas con menos limitaciones físicas.

Este proyecto nace en el interés en conocer y experimentar con estas nuevas formas de interacción para hacer música. La idea es situar el cuerpo del usuario en el centro del proceso creativo y utilizar sus gestos como el principal elemento de control de un sistema de generación de sonido [2]. Esto no solo abre posibilidades para imaginar nuevos instrumentos, sino también para comprender la relación entre lo que hacemos físicamente y lo que oímos de nuevas maneras. El desarrollo de la música por computadora ha facilitado este tipo de experimentación, hoy en día, muchos métodos de síntesis están disponibles en software asequible, lo que permite trabajar en tiempo real con instrumentos potentes. Desde el clásico Theremin de la década de 1920, que respondía a los movimientos de las manos modificando el tono y el volumen.

Basándonos en este enfoque, se propone un sistema multimedia interactivo capaz de generar sonido a partir de los movimientos del usuario con el Software MAX 8. El sistema consta de cuatro partes interconectadas en cascada: captación de gestos de las manos del usuario, síntesis aditiva, modulación AM y filtrado FFT. Todo esto se implementa en Max 8 para capturar hasta 10 variables relacionadas con el movimiento: la posición de los índices de ambas manos en (x,y), la distancia entre estos o gestos como juntarlos, voltear las manos o la apertura de estas.

II. FUNDAMENTACIÓN TEORICA

A. MIDI (Musical Instrument Digital Interface)

MIDI es el lenguaje de comunicación digital estándar que facilita la interacción y sincronización entre una amplia gama de dispositivos musicales electrónicos y computadoras. [3] La principal ventaja de MIDI reside en su capacidad para sincronizar mensajes de eventos musicales de manera flexible, permitiendo su interpretación y manipulación como valores numéricos. Esta característica es crucial porque transforma la información musical en datos procesables, abriendo la puerta a diversas aplicaciones en producción, interpretación y análisis musical.

B. Síntesis Aditiva

La técnica de Síntesis aditiva consiste en construir formas de onda complejas a partir de sus componentes espectrales, las cuales son múltiplos de una frecuencia fundamental; estas se caracterizan por tener su propia amplitud, frecuencia y fase. Siendo que al combinar un número suficiente de estas ondas se puede reconstruir un amplio repertorio de timbres. Siendo que la esencia de la síntesis aditiva gira en torno a su capacidad de modelar o reconstruir directamente el espectro de un sonido.

Esta reconstrucción es posible gracias a un Banco de osciladores, el cual es un generador de componentes espectrales con frecuencia y amplitud específica. Siendo que cada oscilador se encarga de generar un armónico particular del sonido final. El poder controlar amplitud y fase de cada oscilador permite esculpir con precisión el espectro armónico de un sonido.

C. Modulación de amplitud

En el contexto del procesamiento de señales en entornos como Max8, la Modulación de Amplitud (AM) se conceptualiza como la multiplicación de dos señales en el dominio del tiempo para superponer información. La **señal moduladora** (que contiene la información, como audio, control o datos de baja frecuencia) altera dinámicamente la amplitud de una **señal portadora** de mayor frecuencia.

Desde la perspectiva del procesamiento digital, esta multiplicación en el dominio del tiempo se traduce en una **convolución en el dominio de la frecuencia**. El espectro de la **señal moduladora** se replica y se desplaza para centrarse alrededor de la **frecuencia de la portadora (fc)** y su imagen negativa ($-fc$), creando bandas laterales superiores e inferiores.

4 Esto significa que el ancho de banda de la señal AM resultante es el doble del ancho de banda de la señal moduladora. Un aspecto crucial es el **índice de modulación**, un valor de $m \leq 1$ asegura que la envolvente de la señal modulada sea una réplica fiel de la moduladora, evitando la sobremodulación y la distorsión, lo cual es fundamental para una correcta demodulación.

1 D. Transformada Rápida de Fourier

La Transformada Rápida de Fourier (FFT) es un algoritmo fundamental en el procesamiento digital de señales, diseñado para convertir eficientemente datos del dominio temporal al dominio frecuencial. Su función principal es descomponer una señal compleja en sus componentes de frecuencia y fase constituyentes, revelando el contenido espectral y cómo este evoluciona a lo largo del tiempo. Este proceso permite representar una señal temporal como una suma de ondas sinusoidales con distintas frecuencias, amplitudes y fases, ofreciendo un "mapa" detallado de las frecuencias presentes y su intensidad relativa.

2 La precisión de la medición de frecuencias con la FFT está directamente ligada a la **frecuencia de muestreo**. Según el teorema de Nyquist, la frecuencia de muestreo debe ser al menos el doble de la frecuencia más alta de interés para evitar el **aliasing**. Si esta condición no se cumple, las componentes de alta frecuencia se "pliegan" incorrectamente hacia el espectro de frecuencias más bajas, distorsionando los resultados y comprometiendo la fidelidad del análisis.

16 Una de las ventajas más significativas de la FFT radica en su **eficiencia computacional**. Al transformar el cálculo de la DFT de una complejidad de $O(N^2)$ a $O(N \log N)$ (donde N es el número de muestras), la FFT ha hecho posible el procesamiento en tiempo real de grandes volúmenes de datos, como los flujos de audio. Esta eficiencia ha democratizado el análisis espectral y la manipulación de señales, siendo indispensable en diversas aplicaciones que van desde la ecualización de audio hasta el reconocimiento de voz.

III. METODOLOGIA Y HERRAMIENTAS

A. Sistema de captación de gestos

20 La interacción gestual se implementa mediante el software Handmate MIDI [5]. Esta herramienta utiliza la cámara para el seguimiento en tiempo real de la posición de manos y dedos, traduciendo estos movimientos en mensajes MIDI.

19



Fig 1. Captación de gestos con software Handmate MIDI

Las variables de movimiento captadas se categorizan en:

Control continuo

- 1) Posición de los índices izquierdo y derecho (X, Y)
- 2) Distancia entre ambos índices
- 3) Apertura de las manos
- 4) Detección de gestos: juntar índices, voltear mano izquierda o derecha.

Estas variables mapean rangos de movimiento a valores continuos (ej., de 0 a 127), se envían a través de 8 canales disponibles de comunicación entre HandMate MIDI y MAX 8 con ayuda de LoopMIDI, el cual funciona como un puerto virtual para el protocolo MIDI.

Detección de gestos discretos

Identificación de gestos únicos como juntar índices o voltear una mano específica. Estos gestos actúan como disparadores, enviando valores de nota MIDI discretos una vez por detección.

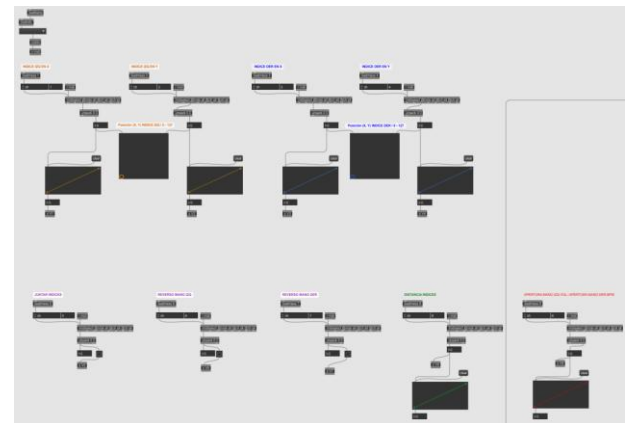


Fig 2. Recepción de gestos con MAX8 desde Handmate MIDI

B. Síntesis aditiva

Esta etapa genera el sonido asignando una frecuencia fundamental inicial, controlada por una variable de movimiento continuo seleccionada por el usuario. El proceso de síntesis sigue este flujo:

- 1) Asignación de **Forma de Onda**: Se define la forma de onda del oscilador base (ej. onda de sierra) que conformará los parciales.

- 2) Control de relación de armónicos: Un segundo gesto continuo modula la frecuencia de los armónicos/parciales en relación con la fundamental. Esto permite que sean múltiplos enteros (armónicos) o no (inarmónicos), alterando el timbre resultante.
- 3) Control de Ganancia: Un tercer gesto continuo ajusta la ganancia general (volumen) del sonido combinado.

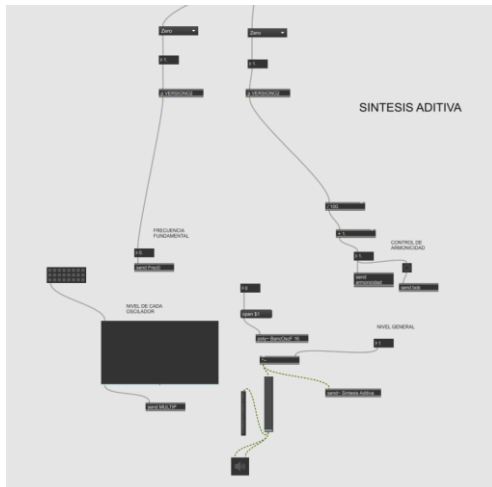


Fig 3. Módulo de síntesis aditiva MAX8

C. Modulación de amplitud

La señal generada previamente (que actúa como portadora) se somete a Modulación de Amplitud.

- 1) Control de Índice de Modulación: Un cuarto gesto continuo asigna el índice de modulación (m), con un rango predefinido de 0 a 1.5. Esto permite observar y experimentar con modulación nula ($m=0$), normal ($0 < m < 1$), al 100% ($m=1$) y sobre modulación ($m > 1$), donde la información puede distorsionarse o ser irrecuperable.
- 2) Control de Frecuencia de la Envolvente: Un quinto gesto continuo modula la frecuencia de la envolvente (frecuencia moduladora). Esto determina el ancho de banda y la separación de las bandas laterales en el espectro de la señal AM.

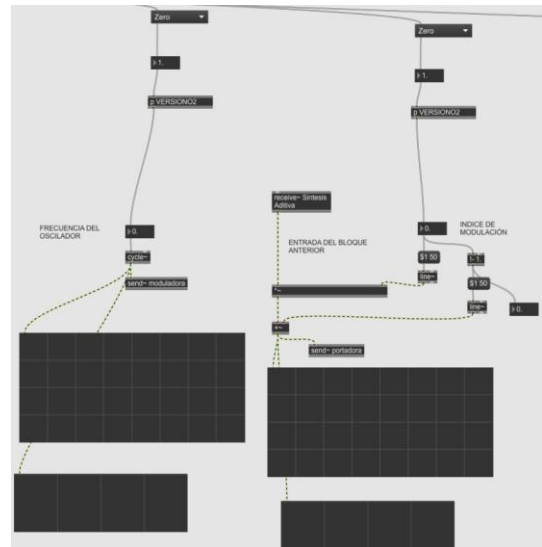


Fig 4 Módulo de modulación AM

D. Filtro de 256 bandas con FFT

Esta etapa aplica un filtro de 256 bandas basado en la Transformada Rápida de Fourier (FFT). El sistema incluye formas de filtro predefinidas.

Un sexto gesto continuo o discreto controla la interpolación entre estas formas de filtro. Por ejemplo, interpolar un filtro pasa-altas con uno pasa-bajas puede generar un filtro pasa-banda.

El sonido resultante de esta etapa final es el percibido por el usuario.

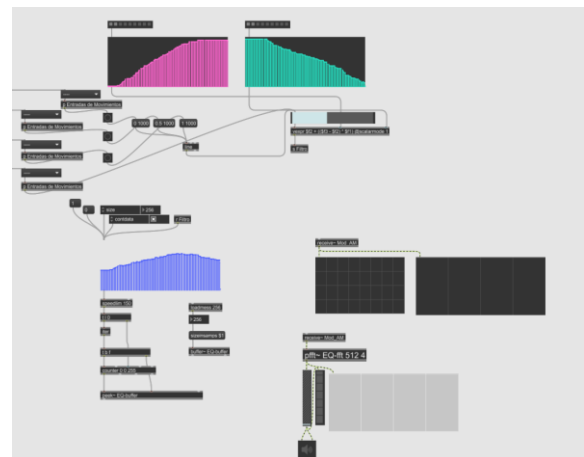


Fig 5. Módulo de filtro de 256 bandas fft

E. Funcionamiento general de todas las etapas

En su totalidad, el sistema permite al usuario controlar hasta ocho variables mediante gestos y movimientos. Esto abarca la generación de la señal (frecuencia fundamental y sus 16 armónicos/parciales) y todos los parámetros de las etapas de procesamiento. Se incluye la capacidad de seleccionar entre "presets" preestablecidos que definen la asignación de gestos a los diversos parámetros de control.

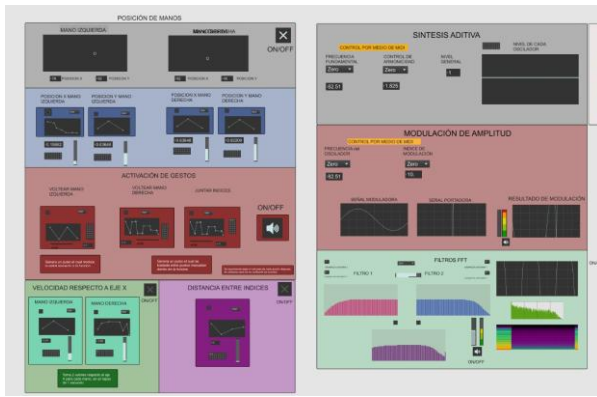


Fig 6. Visión general de interfaz de usuario

IV. RESULTADOS

A. Síntesis Aditiva

Según las variables que es posible manipular en el “Modulo” de Síntesis aditiva tenemos que, al variar la frecuencia fundamental de la señal, esta muestra cómo cambia el contenido armónico de esta señal se va ajustando a la nueva frecuencia. Como podemos observar en la imagen 4.1, observamos que las bandas horizontales de diferentes colores representan un armónico de sonido diferente, Siendo la banda más baja y brillante la frecuencia fundamental y las bandas superiores los armónicos.

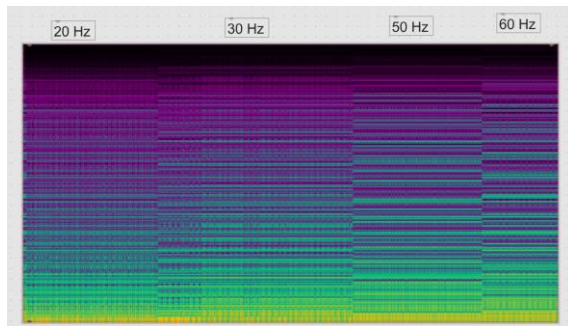


Fig 7. Variación del Espectro debido a la Frecuencia Fundamental vista en el sonograma.

Si se realiza una variación en el nivel general de los osciladores podemos observar en la imagen 4.2, como estos van aumentando su intensidad de brillo y tamaño a medida que aumenta, siendo que el nivel general afecta directamente todos los armónicos. Pero es posible ajustar los armónicos individualmente para así obtener diferentes timbres o frecuencias, siendo que como se observa en la imagen 4.3 vemos como es afectado el sonograma a medida que se aumentan o disminuyen los armónicos que lo componenn.

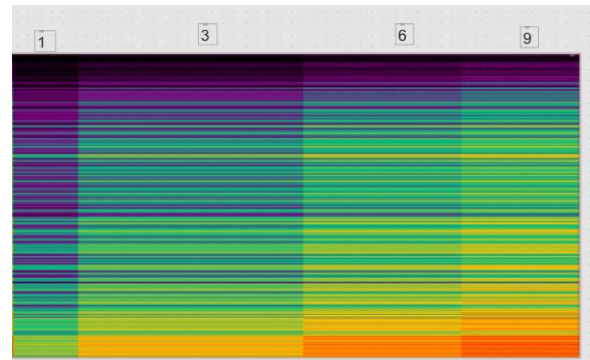


Fig 8.

Variación del Espectro debido al Nivel general, vista en el sonograma.

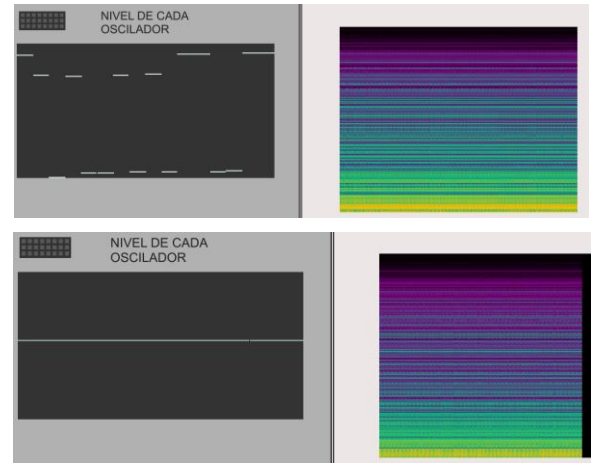


Fig 9. Variación del Espectro debido a la Variación del Nivel de los Osciladores vista en el sonograma.

B. Modulación De Amplitud

En base al control de la frecuencia del oscilador y índice de modulación podemos observar en la imagen 4.4, que, en este caso, la señal moduladora es una senoidal a la que podemos controlar su frecuencia, además de que la señal portadora que sale del “modulo” de síntesis aditiva se ve modificada o ajustada debido al índice de modulación. Siendo que al final de esto podemos observar cómo ambas señales al modularse dan como resultado la combinación de ambas en una sola señal.

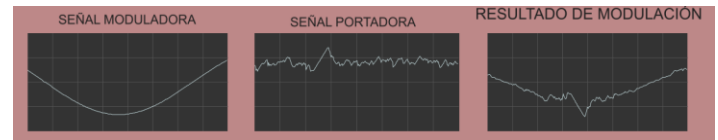


Fig 10. Modulación AM.

C. Filtros

Para los filtros, se realiza una interpolación de 2 filtros diferentes para poder ajustarse dependiendo del timbre a lograr, en el caso que se observa en la imagen 4.5. Se utilizan un filtro pasa altas (filtro 1) y un filtro pasa bajas (filtro 2). Para así realizar un ajuste entre ambos filtros en cualquier momento. Y así obtener resultados como se puede observar en la imagen 7.

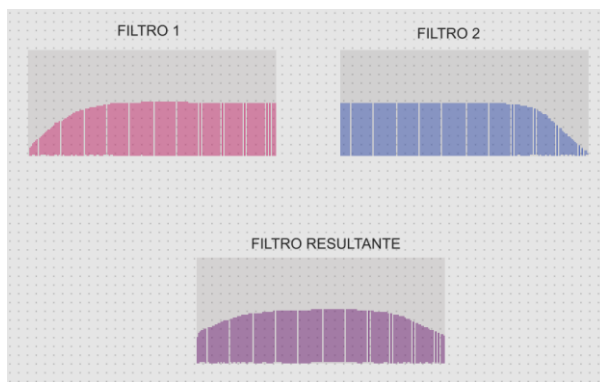


Fig 11 Filtros utilizados

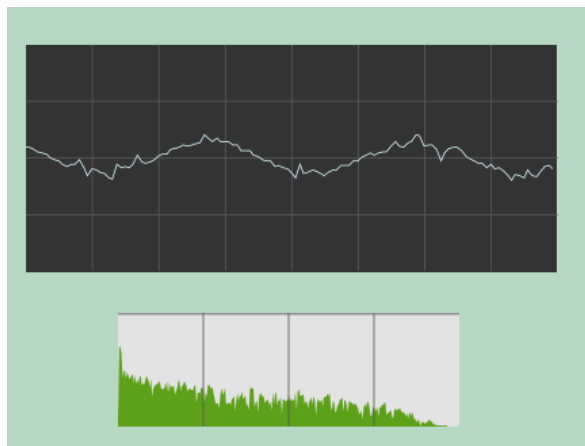


Fig 12. Señal resultante filtrada por el filtro resultante de la interpolación de ambos filtros.

CONCLUSIONES

El prototipo de sistema interactivo desarrollado en Max 8 demuestra exitosamente la capacidad de controlar la síntesis y el procesamiento de sonido en tiempo real a través de gestos manuales. La integración del software Handmate MIDI permitió una captura de movimiento fluida, traduciendo acciones como la posición de los dedos y la apertura de las manos en datos MIDI para modular parámetros sonoros. Este enfoque sitúa el cuerpo del usuario como el principal elemento de control, ofreciendo una forma de interacción musical expresiva y libre de las limitaciones de los instrumentos tradicionales.

La arquitectura modular, que comprende síntesis aditiva, modulación de amplitud y filtrado FFT, proporciona un control detallado sobre el timbre y la textura del sonido. Los resultados visualizados en los sonogramas confirman que los gestos manipulan eficazmente la frecuencia fundamental, la relación de armónicos, el índice de modulación y la interpolación de filtros. Este proyecto no solo valida la viabilidad de la interacción gestual en la creación musical, sino que también abre nuevas posibilidades para el diseño de instrumentos digitales y experiencias sonoras inmersivas.

REFERENCES

- [1] C. Quintero and D. Roa, "Aproximación musical a través de la interpretación gestual," *Tecnológicas*, vol. 25, no. 53, pp. 1–15, 2022. [Online]. Available: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992022000100202
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] SOMA Laboratory, "Flux – Interactive Sound Controller," 2024. [Online]. Available: <https://somasynths.com/flux/>
- [4] L. E. González Blanco, *Diseño y realización de un instrumento musical MIDI basado en control de inclinación*, Proyecto Final de Carrera, Escuela Técnica Superior de Ingenieros, Universidad de Sevilla, España, jul. 2013. [En línea]. Disponible en: <https://biblus.us.es/bibing/proyectos/abreproy/12132/fichero/2.+MIDI.pdf>
- [5] "What is FFT (Fast Fourier Transform)? Math Function in Oscilloscope." Tektronix. Accedido: Jul. 10, 2025. [En línea]. Disponible en: <https://www.tek.com/en/support/faqs/what-fft-fast-fourier-transform-math-function-oscilloscope-useful>
- [6] M. Lim, "Handmate MIDI Demo - MediaPipe Hand-Tracking and Gesture Recognition for Web MIDI," YouTube, 21 de diciembre de 2021. [Video]. Disponible en: http://www.youtube.com/watch?v=H97t17Q_BbM