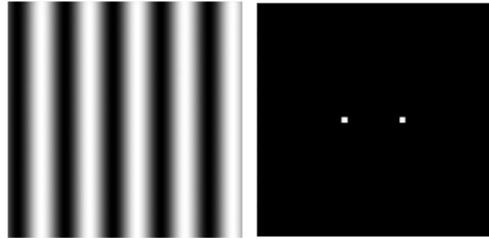




La información que trae una imagen va más allá de lo puramente visual. Hay patrones y simetrías que enriquecen a la percepción del espectador pero que no son conscientemente detectables con una simple mirada. En este trabajo nos propusimos procesar imágenes digitalizadas con el fin de extraer esa información "oculta" al ojo humano. Finalmente, quisimos llevar esa información nueva a una salida audible, para interpretar por medio de otro sentido la información resultante de la imagen. Utilizamos transformadas de Fourier en 2 dimensiones para procesar las imágenes y obtener el perfil de intensidades, promediando la imagen por bloques.

INTRODUCCIÓN

Transformadas de Fourier en 2D importantes



Patrón senoidal 'Propagación' en x

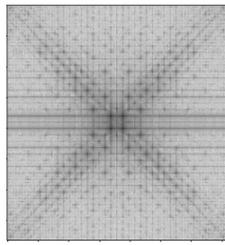
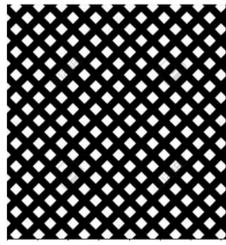
Deltas de Dirac

Las frecuencias 'altas' del espectro de frecuencias espaciales se corresponden con puntos más cercanos al origen del espacio de las frecuencias.

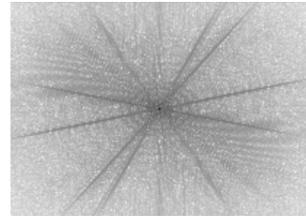


Doble rendija

Sinc

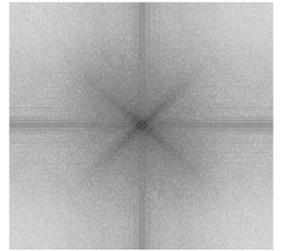


La presencia de patrones regulares o líneas sobresalientes en las imágenes, genera estructuras en las visualizaciones el espacio de Fourier.



Las líneas recta contribuyen como 'rendijas', produciendo estructuras con simetría de reflexión respecto de la diagonal.

Las contribuciones de las frecuencias más altas, asociadas con los detalles finos de la imagen, dificultan la distinción de las estructuras, aún cuando la imagen presenta un patrón claramente regular.

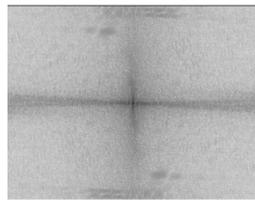


PROCESADO DE IMAGENES

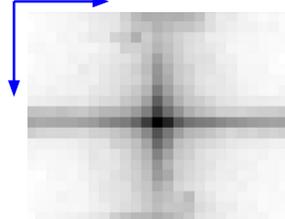
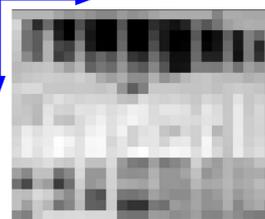
Utilizamos la información del perfil de intensidades para determinar las notas de la melodía



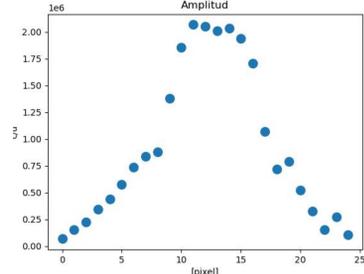
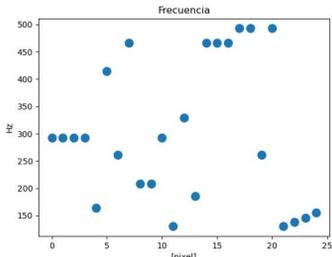
Transformada de Fourier



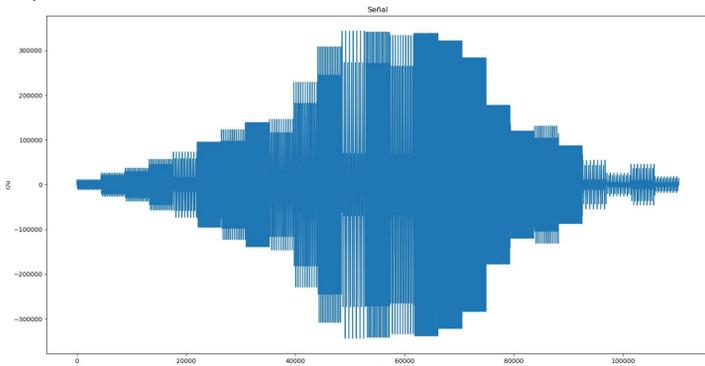
Promedio por bloques



Utilizando los valores de intensidad de los promedios para la imagen y su transformada de Fourier, se definió una sucesión de valores de amplitud y frecuencia, recorriendo las imágenes de izquierda a derecha y de arriba a abajo.

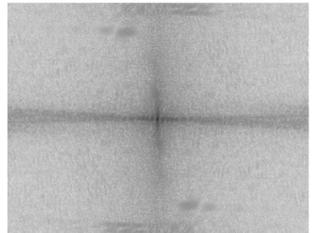
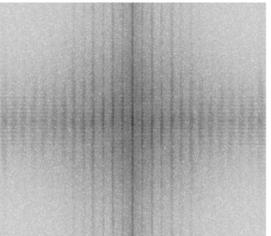
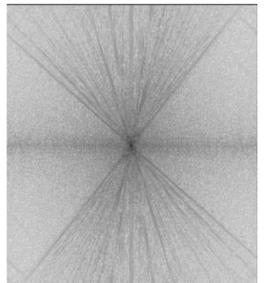


Con esas sucesiones de valores, se generó una señal de sonido compuesta por una concatenación de ondas senoidales, cada una con su frecuencia y amplitud definida por el par correspondiente de valores asociados.



RESULTADOS

Comparamos la sonoridad de 3 imágenes de edificios de ciudad universitaria.



CONCLUSIONES Y PROPUESTAS

La subjetividad de los criterios con los que interpretamos la música, hace imposible intentar 'imaginar' una imagen sólo escuchando un pieza cuya composición haya sido influenciada por algún parámetro de la imagen. Sin embargo, encontramos en la técnica un alto potencial como herramienta de creación. Como instrumento.

Se propone la implementación de distintos protocolos para generar los archivos de audio como, por ejemplo, la utilización de cada fila de la imagen como parámetro de voces simultáneas en el archivo final.

Si bien esta función no ha sido desarrollada aún, creemos que su aplicación tomando como input la información de la transformada de la imagen, puede producir patrones rítmicos más elaborados.



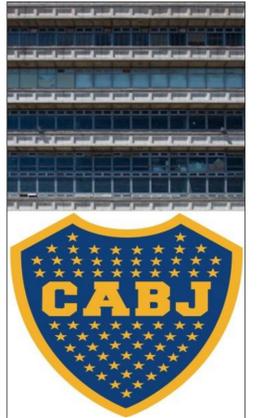
Las analogías entre imagen y sonido se repiten cotidianamente tanto en el arte como en la ciencia. En esta primera categoría, por ejemplo, en los nombres utilizados para diversas cualidades, tales como “ritmo”, “tono”, y “brillo”. En la segunda, sabemos que tanto el sonido como la luz son ondas de distinto tipo. En este trabajo se propone una manera de traducir imágenes digitales a sonidos, utilizando Python para descomponer la imagen en sus valores de RGB, y utilizarlos para generar ondas sinusoidales con parámetros dependientes de los valores extraídos. Los resultados se analizaron en función de la percepción auditiva, según cuán similares a música familiar resultaron.

INTRODUCCIÓN



Las imágenes digitalizadas están divididas en secuencias de píxeles, los cuales son tuplas de 3 valores enteros que van desde 0 hasta 255. Estos valores representan la intensidad de dicho color en el píxel en cuestión. Nosotros tomamos esa secuencia de tuplas y creamos vectores que recorren cada uno, una coordenada de esa secuencia.

Con el fin de poder analizar los patrones en las imágenes como así los cambios “bruscos” en las mismas, utilizamos la siguiente composición de dos imágenes:



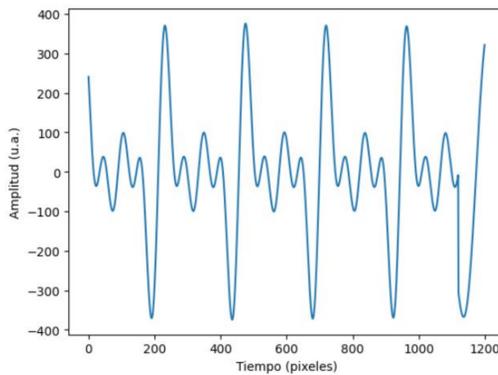
ELECCIÓN DE CRITERIOS

- **Ritmo:** Este vector indica la distancia temporal entre cambio de notas. En cada cambio de nota, el valor de “ritmo” se tradujo a un tiempo entre 1 y 1/6 de segundo, así variando la posibilidad de ritmos entre 60 y 180 bpm (6 notas/segundo = 360 notas/minuto, posibilitando contratiempos).
- **Frecuencia:** Del vector de frecuencias se tomaron los valores que serían traducidos a notas. Si el valor del vector en dado píxel se lo llama “f”, la fórmula usada resulta: $n = 440 \cdot 2^{f/12}$ donde se normaliza el rango de f para ir entre -12 y 12 (redondeado a enteros), y n es la nota resultante, “centrada” en 440 Hz. Además se eligió solamente el valor del vector en el píxel específico donde cambiará la nota, indicado por el vector de ritmos. De esta manera se mantiene una misma frecuencia hasta el siguiente cambio.

TRATAMIENTO DEL SONIDO FINAL

Timbre:

Buscando enriquecer el timbre obtenido, se sumaron senos de frecuencias múltiplos de la original. Estas frecuencias elegidas fueron 3/2, 1/2, y 2 veces la principal.



Ataque:

No se cambió el ataque de las notas respecto al predeterminado.

- **Amplitud (volumen/intensidad):** Con este valor se tomó la amplitud que tendría cada función senoidal. Si bien la relación no es lineal, de esta amplitud se desprende la intensidad del sonido. Se suavizó la señal usando la técnica de *rolling window*¹.



ELECCIÓN DE LOS VECTORES

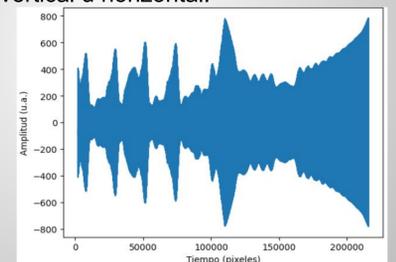
Contábamos con 6 posibilidades para separar los vectores RGB en las 3 variables para construir el sonido (Ritmo, Frecuencia y Amplitud). A continuación se muestran los audios obtenidos con cada una de las combinaciones posibles.

R	G	B
Ritmo	Frecuencia	Amplitud
Ritmo	Amplitud	Frecuencia
Amplitud	Ritmo	Frecuencia
Amplitud	Frecuencia	Ritmo
Frecuencia	Amplitud	Ritmo
Frecuencia	Ritmo	Amplitud

Observación:

La construcción del sonido se hizo mediante construir una señal temporal, barriendo los píxeles de izquierda a derecha, de arriba hacia abajo. Esto sesga fuertemente el resultado, ya que una recurrencia espacial en la imagen se oíría notoriamente distinta puesta en orientación vertical u horizontal.

Señal esquemática:



CONCLUSIONES

Ninguna conclusión pretende encontrar relaciones objetivas entre imagen y sonido. Los criterios usados se eligieron buscando generar la sensación de musicalidad. Encontramos diferencias cualitativas entre audios generados por imágenes con mayor o menor repetición de colores. Las “islas” de color se perciben como ausencia de cambios en el sonido generado. Los sonidos más similares a música parecen provenir de imágenes con repeticiones espaciales en los colores, principalmente los usados como ritmo y frecuencia, en la dirección de barrido.

Surgen perspectivas posibles a trabajos futuros, tales como el uso de distintas maneras de secuenciar los barridos de píxeles, o la incorporación de la riqueza espectral y el ataque como dependencias de los valores visuales. ¿Qué resultaría de utilizar un mecanismo similar e inverso para traducir audio a imagen?

AGRADECIMIENTOS



REFERENCIAS

- [1] Pierian Training. *Python Smoothing Data: A Comprehensive Guide* (19/05/2023) <https://pieriantraining.com/python-smoothing-data-a-comprehensive-guide/>
 [2] Anikin, A. & Johansson, N. *Implicit associations between individual properties of color and sound.* <https://doi.org/10.3758/s13414-018-01639-7>