Universidad Nacional Autónoma de México

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas

Análisis de Señales en Tiempo Frecuencia

Reporte: Identificación de notas musicales utilizando descriptores en tiempo frecuencia.

Legaria Peña Juan Uriel

8 de diciembre de 2021

1. Problema a resolver

El problema a resolver en este proyecto fue la identificación de notas musicales en una señal de audio $x[t]$ usando su espectrograma y descomposición Wavelet.

Las principal característica a suponer sobre la señal $x[t]$ es que se han grabado en ella una secuencia de notas individuales (melodía) con un instrumento, y que estas se encuentran suficientemente separadas en tiempo. En aplicaciones reales esto supone una limitación importante, ya que dentro de una pieza musical se suelen tocar varias notas a la vez. Sin embargo, trabajar con señales que cumplen con esta simplificación permite estudiar el efecto de combinar el espectrograma y la descomposición wavelet en un problema controlado de clasificación.

Concretamente lo que buscó en este proyecto fue diseñar un algoritmo que tomara la señal $x[t]$ como entrada y devolviera como salida una tablatura, indicando la secuencia de notas que se tocaron en la señal introducida así como el tiempo que duró cada una.

La hipótesis de que la fusión espectrograma - T.Wavelet podría ser provechosa surge de observar que cuando se realiza la descomposición wavelet en la señal de audio de una melodía grabada, se tiene por una parte una separación de detalles, que podrían ser sonidos particulares del instrumento o recinto donde se adquirió la señal (pulsaciones de teclas por ejemplo), y por otro lado una aproximación de la señal, que representaría la serie de tiempo aislada de la melodía. Esto sugiere que incluir en la clasificación información de los espectrogramas de las componentes Wavelet separadas, podría ayudar no solo a clasificar notas individuales usando la aproximación, sino posiblemente distinguir versiones de una misma nota tocados en distintos instrumentos o en distintos lugares usando los detalles.

En este trabajo el enfoque se ha centrado únicamente en la clasificación de notas distintas. Las siguientes secciones detallan tanto la metodología que se desarrolló, como los resultados obtenidos.

2. Metodología

2.1. Registro y análisis de señales para entrenar

Para el registro de datos se grabó una señal de entrenamiento, en la cual se tocó cada nota de la octava central del piano 4 veces (Figura 1).

Como se puede observar en dicha figura, cada muestra de las notas que se tocaron abarca un segmento temporal identificable, desde el punto en que se tocó la tecla,

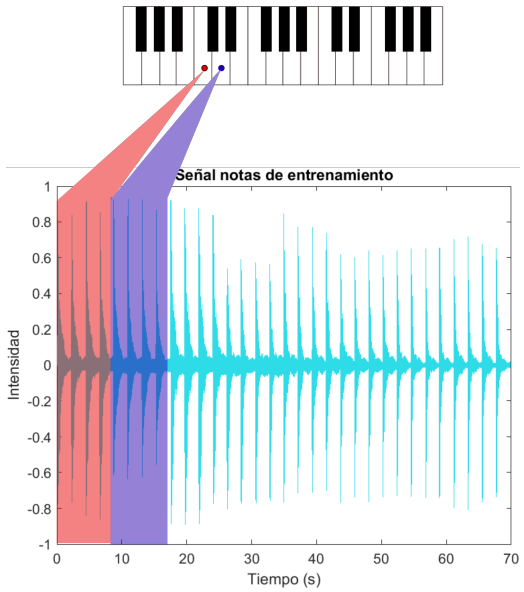


Figura 1: Señal en la que se toco 4 veces cada nota de la octava que contiene al do central en el piano.

pasando por un punto de amplitud máxima y culminando con un segmento de atenuación antes de que se tocara la siguiente nota.

Cada uno de estos segmentos se recortó para extraer características a partir de ellos. Los recortes correspondientes a las 4 veces que se tocó la nota Do, se muestran en la figura 2.

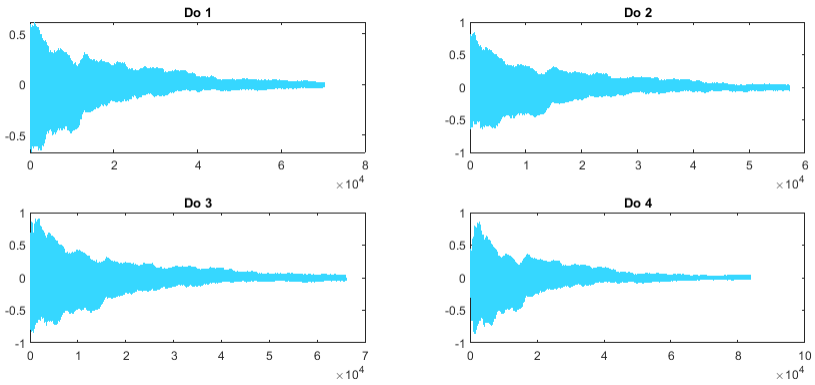
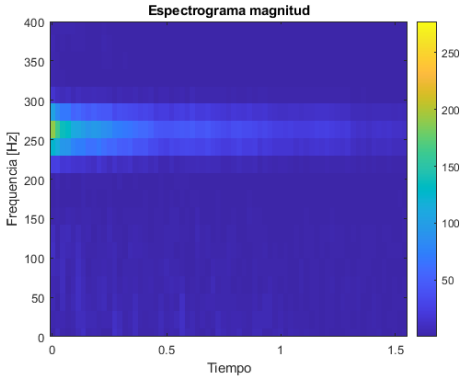
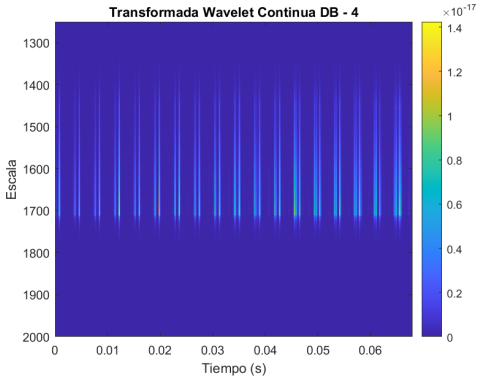


Figura 2: Recortes de la señal de entrenamiento correspondientes a las 4 veces que se tocó la nota Do.

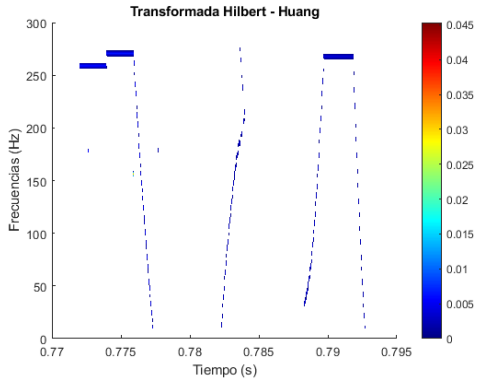
Esta nota en particular debiera tener una frecuencia de 261.63 Hz. Para corroborar que en efecto esta fuera la frecuencia predominante en las señales recortadas, se aplicaron 3 de las herramientas vistas en el curso para análisis de señales en tiempo frecuencia: el espectrograma, la transformada Wavelet continua y el espectro Hilbert - Huang. Los resultados obtenidos para el primero de los recortes son los que se muestran en la figura 3



(a) Espectrograma



(b) T. Wavelet Continua DB - 4



(c) Transformada Hilbert - Huang

Figura 3: Análisis en tiempo frecuencia del primer recorte de la nota Do.

Puede observarse que en los 3 análisis hay una activación de la frecuencia deseada en el tiempo que dura la señal. En el caso de la transformada wavelet, el análisis requiere un tanto mas de cuidado ya que aquí lo que se manejan son escalas. La escala a la que uno esperaría observar la activación sería:

$$s = 2\pi\nu = 2\pi(261.63Hz) = 1643 \tag{1}$$

Esta escala es justamente la que se muestra activada en la figura anterior.

Hasta este punto se contaba así, con 4 recortes de cada una de las notas: do, re, mi, fa, sol, la, si, do.

2.2. Obtención de características

El paso siguiente consistió en la extracción de características a cada uno de los recortes obtenidos de la sección anterior. Esta extracción consiste esencialmente en obtener diversos espectrogramas: el de la propia señal, el de 5 detalles obtenidos de la descomposición Wavelet, y el de su aproximación (en total son 7 espectrogramas). Estos espectrogramas vienen representados por matrices, donde la i - ésima columna representa un espectro de potencias obtenido para la i-ésima ventana de tiempo. Denotaremos a la i-ésima columna del j-ésimo espectrograma como $S_{ij}[f_k]$, haciendo explicito que se trata de un espectro de potencias definido en arreglo de frecuencias f_k.

A continuación se construyen vectores de características. La j-ésima componente del i-ésimo vector de características se calculará como la frecuencia promedio de $S_{ij}[f_k]$:

$$v_{ij} = \sum_k f_k \times \left(\frac{S_{ij}[f_k]}{\sum_k S_{ij}[f_k]} \right) \tag{2}$$

Estos vectores de características se obtienen para cada uno de los cortes, y se van almacenando como las filas de una matriz de datos M, así mismo se va construyendo un vector de etiquetas L, en el cual se van guardando números del 0 al 7 dependiendo de cual sea la nota que se haya tocado en el corte del que se extrajo el vector.

Cada una de las columnas de M fue normalizada restando la media y dividiendo entre su desviación estándar. Esta es una acción de preprocesamiento muy común para datos que se utilizarán en el entrenamiento de modelos de aprendizaje computacional.

En la figura 4 se muestra un gráfico de las primeras dos entradas de 1000 de los vectores obtenidos en la forma antes descrita. Estas dos componentes corresponderían a la frecuencia promedio de la señal como tal y del primer nivel de detalle. El color de los puntos índica la nota a la que (su etiqueta).

2.3. Entrenamiento del modelo

A partir de la matriz de datos M y el vector de etiquetas L, se entrenó un modelo de máquina de soporte vectorial para llevar a cabo la clasificación.

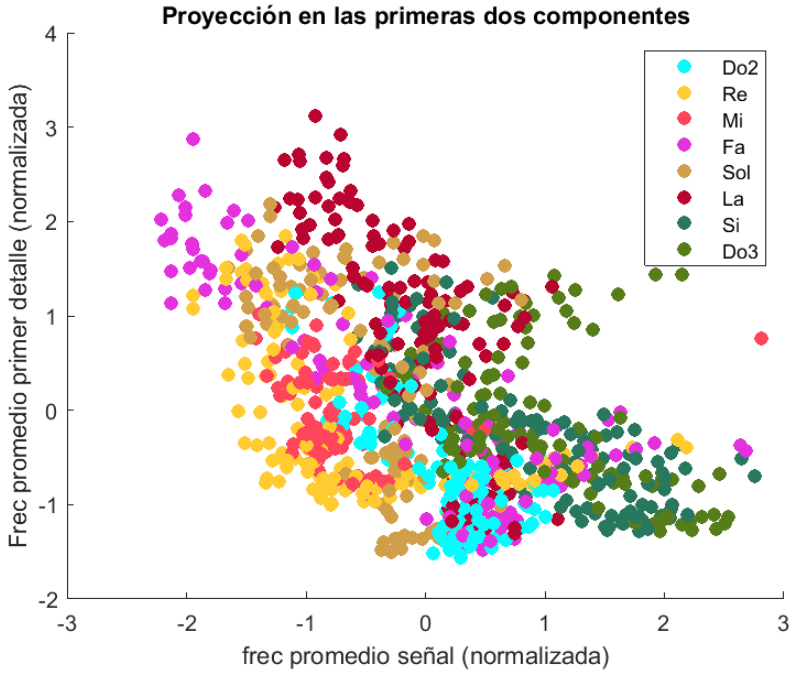


Figura 4: Gráfico de las primeras dos componentes de los vectores de características que se extrajeron.

En total, el número de vectores de entrenamiento utilizados (filas de la matriz M) fueron 5236, mientras que el número de vectores de validación fue 1309 y constituyó 20 % del total de datos.

2.4. Uso del clasificador para generar la tablatura y reconstrucción de la misma

Se creo una función `analyzeSong`, que recibiera una señal $x[t]$ y regresara una tabla, indicando la secuencia de notas que se tocaron en dicha señal. En la versión inicial del algoritmo, la cual se presenta en este trabajo, dicha función también recibe como ayuda un vector de cortes c, el cual índica los tiempos a los cuales inician y termina cada una de las notas en la señal. Usando dicho arreglo, la función corta las secciones de audio de las notas individuales, y para cada una de ellas obtiene una tabla de características en la forma que se explicó en 2.2. La tabla de características obtenida para un corte contiene varios vectores, los cuales son introducidos al clasificador para obtener la nota que corresponde a cada uno. La nota que mas se haya asignado

(moda) a los vectores de esta tabla se tomará como la nota que se tocó en el segmento correspondiente de tiempo.

Así, en la tabla de salida lo que indicaremos en cada fila son la nota que se identificó en el correspondiente corte, y los tiempos inicial y finales de este. Para los lapsos de tiempo entre cortes, indicaremos que hay un silencio poniendo un -1 en la entrada que corresponde a la nota clasificada. Por ejemplo si el fin del primer corte se dio a los 20s, y el inicio del segundo corte se dio a los 30s, añadiríamos la fila: $[-1, 20, 30]$ a la tabla.

Para corroborar que las tablas construidas con tal algoritmo fuesen correctas se diseño una función de reconstrucción llamada `playTab`. Esta función toma la tablatura y genera una reconstrucción de la señal usando señales puras (senoidales).

3. Resultados

En esta sección se resumen los resultados obtenidos con el método propuesto de detección de notas.

3.1. Validación del modelo

Las exactitudes obtenidas para la clasificación de las 8 notas de la octava central del piano se muestra en la tabla 1.

Nota	Exactitud
Do (0)	0.95
Re (1)	0.98
Mi (2)	0.96
Fa (3)	0.96
Sol (4)	0.88
La (5)	0.91
Si (6)	0.87
Do2 (7)	0.87

Cuadro 1: Exactitud obtenida para las distintas notas.

Puede observarse que las notas que presentan exactitud mas baja son Sol, Si y Do2. En el caso de Si y Do2, esto podría deberse a que dichas notas son adyacentes en el piano, i.e. no hay tecla negra de por medio, y por tanto tienden a tener frecuencias mas próximas entre sí que otros pares de notas. En el caso de Sol algunas posibles razones serían el modo en que se cortaron los segmentos o falta de datos de entrenamiento.

3.2. Clasificación y reconstrucción de canciones

Se probó el algoritmo usando la grabación de un segmento de "Mary had a little lamb". En la figura 5 se muestra la serie de tiempo de intensidad de dicha señal. Así mismo, en un cuadro de texto dentro de la misma figura se muestra la secuencia de notas que se tocaron, donde -1 denota un silencio y las 8 notas de la octava han sido codificadas con números de 0 a 7.

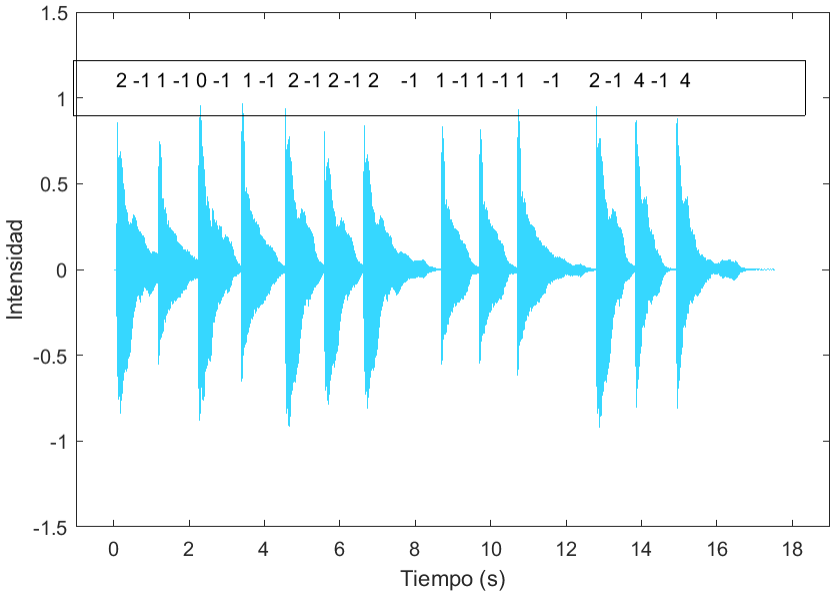


Figura 5: Señal de audio donde se grabó un segmento de "Mary had a little lamb". El cuadro de texto en la figura indica la secuencia de notas que se tocaron, donde -1 es silencio y las notas de la octava se denotan por números de 0 al 7.

En la figura 6, se muestra la secuencia de notas arrojada por el algoritmo al clasificar esta señal.

```
Secuencia de notas
disp(noteSequence);
2 -1 1 -1 0 -1 1 -1 2 -1 2 -1 2 -1 1 -1 1 -1 1 -1 2 -1 4 -1 4
```

Figura 6: Secuencia de notas arrojada por el algoritmo

Puede observarse que en efecto la secuencia de notas arrojada coincide con lo que se esperaba obtener.

Así mismo, se reconstruyó un audio a partir de la tabla de notas generada por el algoritmo, usando tonos puros (señal sinusoidal). El audio generado se enviará como anexo adjunto a este trabajo,

4. Conclusión

Se propuso un algoritmo para identificación de notas musicales en una señal de audio, utilizando los espectrogramas tanto de la señal a analizar como de sus detalles y aproximación Wavelet. El algoritmo funciona correctamente para piezas sencillas, en la que las notas se encuentran suficientemente separadas entre sí, sin embargo queda por investigar sus limitaciones en situaciones donde exista un traslape entre las notas, o para señales donde la melodía se toque en un instrumento distinto al piano.

Las exactitudes que se obtuvieron con el modelo de clasificación sugieren que el uso de características en el dominio Wavelet provee información útil para la separación de notas musicales, y podría utilizarse por ejemplo en la digitalización de señales captadas con micrófonos convencionales. Esto sin duda constituiría una herramienta útil, ya que usualmente se requiere de equipo especial (por ejemplo cables MIDI) para grabar digitalmente los sonidos producidos por instrumentos musicales en una computadora.

Referencias

[1] C. Valens. *A Really Friendly Guide to Wavelets*. 1999.

[2] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1 1992.