UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO – SALAMANCA

DIVISIÓN DE INGENIERÍAS

"Detección de ritmos musicales en la ejecución de instrumentos de percusión"

TESIS

PARA OBTENER EL GRADO DE:

MAESTRO EN ADMINISTRACIÓN DE TECNOLOGÍAS

PRESENTA:

Hugo Armando Aguilera García

DIRECTOR DE TESIS:

Dra. Rocío Alfonsina Lizárraga Morales

YURIRIA, GTO Diciembre, 2022

Agradecimientos

Personales

Al resplandeciente sonido omnipresente y eterno, fundamento de todo lo visible e invisible, el cual se encuentra en constante movimiento rítmico.

A mi familia por su apoyo, mis amigos por su escucha. Por su actitud crítica incansable a los doctores del Departamento de Estudios Multidisciplinarios en lo general. En particular a la Dra. Rocío por su confianza y respaldo. A Mario por su pensamiento creativo en el proyecto.

Institucionales

Al Departamento de Estudios Multidisciplinarios de la Universidad de Guanajuato por la oportunidad y al CONACYT por el financiamiento con número 781149.

Resumen

En el actual proyecto de tesis se presenta una propuesta de un algoritmo para generar un descriptor de alto nivel del ritmo en la música digital. Un conjunto de grabaciones de una batería se analiza para demostrar los elementos de la propuesta. El algoritmo lee directamente de los archivos de audio la información de la forma de onda del sonido. La señal de audio se analiza en marcos de tiempo con la Transformada de Fourier de Tiempo-Corto y se construye una función de detección basada en la energía. Cinco algoritmos de aprendizaje máquina se entrenan y validan para identificar el patrón en la función de detección alrededor del inicio. Ubicando los inicios se calculan las distancias relativas entre ellos. Lo anterior permite, calcular los descriptores de alto nivel del ritmo, los Índices de Variabilidad en Pares. Estas representaciones del ritmo hacen posible hacer comparaciones entre los patrones del ritmo en frases de piezas musicales.

Índice General

Índice de Figuras

Índice de Tablas

Capitulo 1. Introducción

> “Ritmo es el latido del corazón. Es el primer tambor, una historia en sonido que revela nuestra imaginación y celebra nuestro poder. El ritmo es el común fundamento multi cultural de la familia humana”. Tony Vacca

La música es el arte de combinar sonidos para agradar al oído, en cumplimiento con un conjunto de leyes que lo rigen. Los músicos hacen uso de sonidos para componer sus piezas artísticas. Con el propósito de ser agradable al oído, la música debe cumplir con las cualidades de melodía, armonía, textura, forma, expresión, movimiento y ritmo. Toda obra musical cuenta con una sucesión regular de sonidos, alternando entre fuertes y débiles denominada, pulso o pulsación. La melodía es la sucesión de sonidos de una forma determinada, de acuerdo con los estilos o épocas musicales. La armonía hace referencia a la unión de varios sonidos diferentes de manera simultánea. El ritmo en una pieza musical es la combinación y sucesión de los sonidos y silencios a través del tiempo. El ritmo, cualidad de la música es uno de los aspectos más sobresalientes e intrigantes de una canción y de la música en general. Junto con la armonía y la melodía, el ritmo constituye uno de los elementos estructurales básicos de toda la música.

El diccionario de la real academia española define el sonido, como “la sensación producida en el órgano del oído por el movimiento vibratorio de los cuerpos, transmitido por un medio elástico, como el aire”. El sonido posee las cualidades físicas de altura o tono, intensidad o volumen, duración y timbre. La altura o tono del sonido lo identifica como agudo o grave, está dada por la frecuencia de la vibración sonora, se mide en Hercios (Hz). Un sonido grave tiene frecuencias bajas, mientras, el agudo altas. Las frecuencias audibles al oído humano están dadas entre 20 y 20000Hz, a este rango se le denomina espectro audible. La intensidad o volumen clasifica los sonidos en fuertes o suaves. La separación está determinada por la amplitud de la vibración sonora, medida en decibeles (dB). Por duración del sonido entendemos el tiempo en el cual la vibración sonora está presente, asociado con el ritmo de la música y el timbre hace posible reconocer la fuente productora del sonido.

Los sonidos combinados en una pieza musical son producidos por un instrumento. Las personas que toman estos instrumentos se conocen como ejecutantes, intérpretes o músicos. Una amplia variedad de instrumentos está disponible para los músicos en la composición de una obra musical. Normalmente en una canción se combinan varios instrumentos. Los instrumentos se clasifican en familias: cuerdas, vientos y percusión. Un ejemplo de instrumento de cada familia seria violín, flauta y bombo respectivamente. Cada instrumento produce señales una experiencia sonora particular al escucha.

La importancia de su estudio

Hace un par de siglos atrás, la posibilidad de escuchar una pieza musical era posible solamente al tener los ejecutantes con sus instrumentos interpretando al alcance del oído. La experiencia auditiva de la música para fines de entretenimiento o análisis estaba limitada a tener músicos enfrente del espectador o estudioso. Fue en la década de los 1870 donde comenzaron a aparecer los primeros dispositivos con la capacidad de grabar y reproducir sonidos como el fonógrafo y el gramófono. Varias décadas después en 1940 el disco de vinilo permite grabar una cantidad de tiempo mayor y con mejor calidad y se construyó el magnetófono de bobina abierta. Otras tecnologías posteriores son el casete, el VHS y el mini DV. En el siglo 20, nuevas tecnologías surgieron: El disco compacto, almacenando de manera digital el audio, el formato MP3 usado en computadoras, reproductores de mp3 y teléfonos celulares como relatan Gronow y Saino [1].

Desde la aparición del primer dispositivo grabador comercial en el año 1977, fue posible que millones de canciones hayan sido digitalizadas. Desde ese momento se han creado colecciones de canciones personales de algunos centenares o miles, hasta algunas de decenas de millones en las grandes compañías de la industria musical. La gran cantidad de canciones disponibles representan un reto mayúsculo en las tareas de búsqueda, recuperación y organización de los contenidos musicales. Además, el análisis de estas piezas individuales y las colecciones se presenta como una oportunidad para los estudiosos de la música: musicólogos, estudiantes, compositores e interpretantes, otros beneficiados son los lingüistas, neurocientíficos y por supuesto los consumidores finales de contenidos musicales.

Enfrentando el reto de organizar y analizar la música digitalizada se han propuesto algoritmos y metodologías que se les conocen con el nombre de sistemas de recuperación de información musical, en inglés se le denomina Music Information Retrieval (MIR). Las investigaciones en

la recuperación de información musical han hecho participes a expertos de diversas áreas como percepción musical, cognición, musicología, ingeniería, ciencias computacionales, entre otras. Estos expertos colaboran en una actividad multidisciplinaria, resultando en estos algoritmos y soluciones metodológicas propuestas para solucionar el problema de búsqueda musical, haciendo uso de métodos basados en el contenido como lo indica Casey *et al.* [2].

Casey *et al*. [2] afirma que las técnicas de MIR se proponen apoyar tareas relacionadas con la organización y recuperación de información musical: identificar la música, monitorear los derechos de autor, manejo de versiones, reconocer melodía, trabajos idénticos, ejecutantes, compositores, humor, estilo y género, instrumentos, piezas que suenan parecido, ofrecer recomendaciones, permitir hacer alineaciones y detección de plagios. Las tareas anteriores se basan en la extracción de un conjunto de características de alto nivel como el timbre, melodía, tono, armonía, tecla, lírica, estructura y ritmo. En este trabajo nos enfocamos en el ritmo.

Un conjunto de técnicas forma sistemas MIR. Estos permiten consultar datos de la música o piezas musicales (archivos de audio). Casey *et al*. [2] lista tres enfoques para el logro de esta tarea en los sistemas MIR: el basado en metadatos, el que usan descriptores de contenido musical de alto nivel y el que acceden a las características de audio de bajo nivel. En el primero acompañando a un archivo de audio hay un conjunto de datos en formato de texto que lo describen, informando sobre el intérprete, el nombre de la canción, la duración, la fecha de publicación o en algunos sistemas el humor de la pieza, el género, el estilo, la emoción u otros. Este enfoque es el usado por las grandes compañías de entrega por demanda, como Spotify, Apple Music, Amazon Music, Google Play y otros.

En el segundo enfoque, los sistemas MIR generan o almacenan datos relativos a la pieza musical que la describen en un alto nivel. Los descriptores indican el timbre, la melodía, el ritmo, tono, armonía, estructura, letra o descriptores no pertenecientes a la música occidental. El tercer enfoque, usa información directamente del audio digital. La extracción de esta información de bajo nivel hace uso de tres estrategias: la segmentación basada en marcos de tiempo, la segmentación basa en la sincronización de pulsación y las basadas en medidas estadísticas. La caracterización de los audios en bajo nivel no nos dice algo sobre la música, pero nos ayuda a construir otras características de un mayor nivel.

Solo son unos pocos los motores de búsqueda de contenido musical basados en descriptores de alto o bajo nivel que están disponibles, dos ejemplos de ellos son: naiyo.com y shazam.com. La extracción de descriptores de bajo y alto nivel de la música están en estudio actualmente. A continuación, discutiremos las ideas que permitan proponer un sistema MIR para acceder a contenidos musicales basado en un descriptor de alto nivel del ritmo.

Las representaciones del ritmo

El ritmo como uno de los elementos esenciales de la música es motivo de estudio, sin embargo, su representación no ha sido unificada. Según Sethares y Bañuelos [3] las notaciones rítmicas presentan el tiempo a través de metáforas espaciales. Los tres enfoques de notaciones son: simbólicas, literales y abstractas. Las primeras, hacen énfasis en la información de alto nivel del sonido, las segundas permiten recrear el sonido y las terceras son manifestaciones artísticas visuales. El conjunto de notaciones simbólicas está formado por: notación lirica, la notación musical, notación de collar, notación numérica, notación funcional, tablatura usada en las percusiones, notación de Schillinger, Interface Digital de Instrumento Musical en inglés Musical Instrument Digital Interface (MIDI). En el conjunto de notaciones Literales encontramos: Formas de ondas, espectrogramas, representación granular. Mientras en las representaciones visuales o artísticas están: La pintura de los dos danzantes de Matisse, ritmo en líneas negras de Mondrian y lo que Goethe llama como "Música congelada".

Desde hace un poco más de dos décadas una representación del ritmo ha sido usada como una herramienta para analizar el ritmo musical, el índice de variabilidad en pares. Toussaint [4] define a la versión normalizada del índice de variabilidad en pares (nPVI) como la media de las variaciones de un conjunto de distancias obtenidas de la secuencia ordenada de pares de eventos. Esta medida se ha usado para analizar y comparar el ritmo entre culturas por Toussaint [4] y Grabe y Low [5]. Además, Daniele y Patel [6] y Patel y Daniele [7] lo han usado en el área de las neurociencias para comparar el ritmo en el habla y la música. También, a Raju, Asu y Ross [8] les ha permitido comparar el ritmo entre las partituras y las interpretaciones de una pieza musical.

La extracción del ritmo en la música digital

El ritmo es una de las características fundamentales de la música. Es un descriptor de alto-nivel de la música. Es extraído a través de una de las tareas fundamentales de un sistema MIR, la detección de inicios en las notas musicales. El reto principal en el problema de detección de inicios de notas musicales es construir un algoritmo que pueda detectar entradas en varios tipos de señales. Cada instrumento musical tiene sus propias características, debido a sus materiales de construcción, produciendo señales muy particulares. La caracterización de instrumentos no se ha detenido solamente en instrumentos de la música occidental como lo hace Stasiak *et al.* [9] , trabajos recientes han sido elaborados en instrumentos de música no-occidental, como instrumentos de percusión de música Carnática por Kumar, Sebastian y Murthy, [10], la música Gamelán por Wulandari, Tjahyanto y Suprapto [11], y ensambles de percusiones de la ópera de Beijing por Tian *et al.* [12].

Diferentes propuestas a la detección de entradas de notas musicales pueden ser encontradas. Aunque todas ellas comparten una serie de pasos en común. Bello et al. [13] delinea un algoritmo general para la detección de las entradas de las notas musicales que consiste en los siguientes pasos:

1. Captura de la señal de audio.
2. Pre-procesamiento.
3. Reducción de la señal.
4. Generación de una función de detección.
5. Un proceso de selección y comparación de picos.

El pre-procesamiento implica la transformación de la señal original para acentuar o atenuar varios aspectos de la señal de acuerdo con su relativa importancia en la tarea en cuestión. El concepto de reducción refiere al proceso de transformar la señal de audio en una sub-muestra usada por la función de detección que permita manifestar la ocurrencia de transiciones en la señal original. La función de detección es un proceso clave para la detección de entradas, creando una función en la que los inicios de las notas coincidan con los máximos o mínimos. Mientras, el proceso de selección y comparación de picos identifica máximos locales, con algún nivel de variabilidad en el tamaño y forma y enmascarada por algo de ruido en la señal.

En la literatura de detección de inicios encontramos una amplia variedad de enfoques, destacan por su continuo uso aquellos donde se toman las características espectrales de la señal y en los que se hacen uso de aprendizaje automático. Para representar una señal de audio en el dominio tiempo-frecuencia y producir una función de detección de inicio basada en la estructura espectral de la señal, se utilizan herramientas de transformación como la Transformada de Fourier de Tiempo-Corto (STFT), transformada de Wavelet, transformada de la constante Q, entre otras. Bello *et al.* [13] elabora una revisión de las técnicas mas utilizadas.

Los algoritmos de reconocimiento de patrones en su categoría de supervisados han sido usados en la tarea de detección de inicio de notas musicales, algunas de estas técnicas incluyen: Máquina de Soporte Vectorial (MSV), Redes Neuronales Artificiales (NNA por sus siglas en inglés) simples, NNA Recurrentes de memoria a corto plazo, NNA convolucional y perceptrón multicapa. A continuación, se hace un recuento de los trabajos que usan información espectral y los algoritmos de aprendizaje antes mencionados.

Limitándose a detectar los inicios de las notas musicales y/o generando un descriptor de alto nivel de la música diferente al ritmo encontramos a Rodet y Haillet [14] quienes ponen las bases para construir una función de detección basada en la energía, esta muestra buenos resultados en instrumentos percusivos. Bello y Sandler [15] proponen una función basada en la fase para aquellos instrumentos en donde la energía no es relevante para la detección del inicio, los instrumentos poco percusivos. Duxbury *et al* [16] combinan la información de la energía y la fase proponiendo una función de detección mas compleja.

Apoyando a la tarea de la detección de inicio con una técnica de aprendizaje máquina esta Constantini *et al*. [17] usando MSV. Lacoste y Eck [18] una NNA para clasificar marcos de un espectrograma en aquellos donde hay inicio y lo que no. Marchi et al. [19] usa una NNA recurrente de memoria a corto plazo (Long Short-Term Memory). Schlüter y Bock [20] usan una NNA convolucional para identificar bordes en un espectrograma considerándolo una imagen. Stasiak et al. [9] combina varias funciones de detección y una NNA del tipo perceptrón multicapas no lineal.

Los trabajos previos permiten construir un sistema MIR con un mayor alcance, sin limitarse a la tarea especifica de los inicios u orientarlos en generación de un descriptor de una nueva

característica de la música. La identificación de los inicios permite construir un descriptor de alto nivel basado en ellos. El ritmo es una de esas características de alto nivel. Una representación del ritmo que use los inicios o la distancia entre ellos puede explorarse. La representación debe de permitir analizar el ritmo. Es decir, el ritmo debe poder ser interpretado y comparado. Estos dos elementos permitirán generar una plataforma en un futuro para recuperar piezas musicales diferente a las disponibles actualmente, una basada en el ritmo.

La propuesta de un descriptor del ritmo

En la presente tesis se trabajan en los primeros elementos de un sistema MIR. Estos elementos realizan las tareas que se describen a continuación. Primero, acceder directamente a los archivos de audio, usando la estrategia de segmentación basada en marcos de tiempo. Segundo, los marcos se clasifican entre los que contienen un inicio de una nota musical y los que no, siendo esto es un descriptor de bajo nivel. Por último, este descriptor de los inicios es usado para generar un descriptor de alto nivel con los índices nPVI y rPVI. Nuestra propuesta es usar las representaciones del ritmo nPVI y rPVI en trabajos futuros para desarrollar herramientas con las cuales se pueda analizar, organizar y buscar piezas musicales basándose en el ritmo. Hacemos uso de grabaciones de instrumentos de percusión para ejemplificar como llegar al descriptor de alto nivel del ritmo.

El presente documento está organizado en cinco capítulos. El primer capítulo es la introducción al proyecto de investigación. El capítulo 2 contiene el marco teórico. En el capítulo 3 la metodología. El capítulo 4 los resultados de la investigación. En el capítulo 5, se exponen las conclusiones y recomendaciones, seguidas de algunas sugerencias de trabajos futuros.

Capitulo 2. Marco Teórico

"El ritmo es el alma de la vida. El universo entero gira en ritmo. Cada cosa y acto humano giran en ritmo." Baba Tunji

2.1. Las señales de audio digital

En la industria musical digital, el sonido emitido por los instrumentos musicales mientras los músicos están ejecutando piezas musicales, es transformado a una representación digital, a través del proceso conocido como grabación o digitalización. La digitalización del sonido consiste en tomar muestras de la señal de onda de un sonido a una cierta velocidad conocida como frecuencia de muestreo. La frecuencia de muestreo estándar es de 44,100 muestras por segundo, abreviado como 44.1 Khz (kilohertz). Este es el estándar de las grabaciones que se entregan a los consumidores finales. Los datos de la onda del sonido se pueden mostrar a través de una representación gráfica como en la Figura 1, esta representación es conocida como forma de onda. La señal corresponde a la grabación de una frase en el ritmo musical rock.

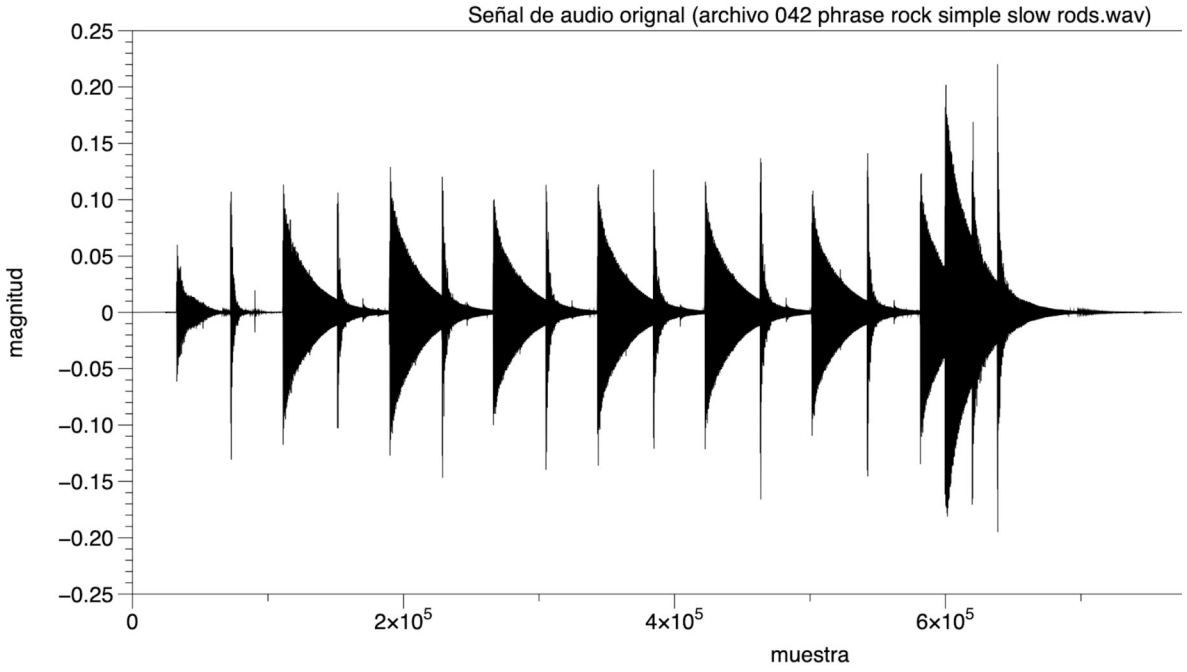

Figura 1. Señal de audio en forma de onda.

Bello et al. [13] menciona tres conceptos relacionados con la tarea de identificar inicios de notas musicales en una señal de audio. Estos son: ataque, transitorio e inicio. El ataque de una nota es el intervalo del tiempo en el cual la amplitud incrementa. El transitorio son los cortos intervalos de tiempo durante la cual la señal evoluciona rápidamente de alguna manera no trivial o relativamente impredecible. El inicio de una nota es el instante elegido como marca temporal a partir del cual se extiende el transitorio. En algunas ocasiones coincide con el inicio del transitorio, en algunas otras es previo al momento en el que el transitorio puede ser detectado. Un elemento adicional que caracteriza la presencia de un inicio es el incremento en el nivel de energía. Esto es una característica sobresaliente en los instrumentos percusivos. La Figura 2, muestra una representación de estos elementos en una onda y el posterior deterioro de la señal.

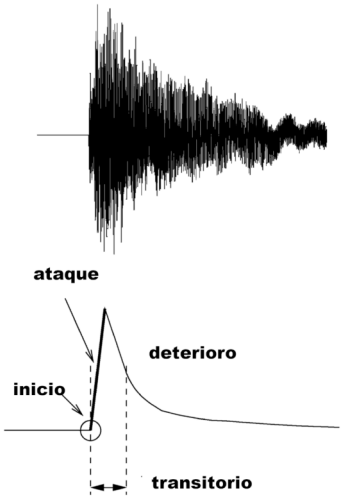

Figura 2. Nota musical con su inicio, ataque, transitorio y deterioro. Tomada y traducida de Bello *et al.* [13]

2.2. El espectrograma de un sonido digital

En un espectrograma el inicio de una nota musical es claramente visible como una mancha oscura a lo largo de todo el rango espectral, sobresaliendo en las altas frecuencias. Un espectrograma es creado a partir de la información espectral de una señal de audio. En la Figura 3, se muestra un audio en forma de señal de onda y su respectivo espectrograma. El audio es la grabación de los instrumentos de percusión que forman una batería. Los instrumentos comúnmente encontrados en una batería incluyen: bombo, caja, tarola y

contratiempo. La Figura es un acercamiento a los primeros cuatro golpes de bateria en el audio antes mencionado. Para producirla y el hacer el marcado de los inicios se usa el programa Sound Onset Labelizer de Leveau y Daudet [21]. El programa permite marcar los inicios de las notas musicales, en una interface donde es posible hacer acercamientos a diferentes niveles de la señal.

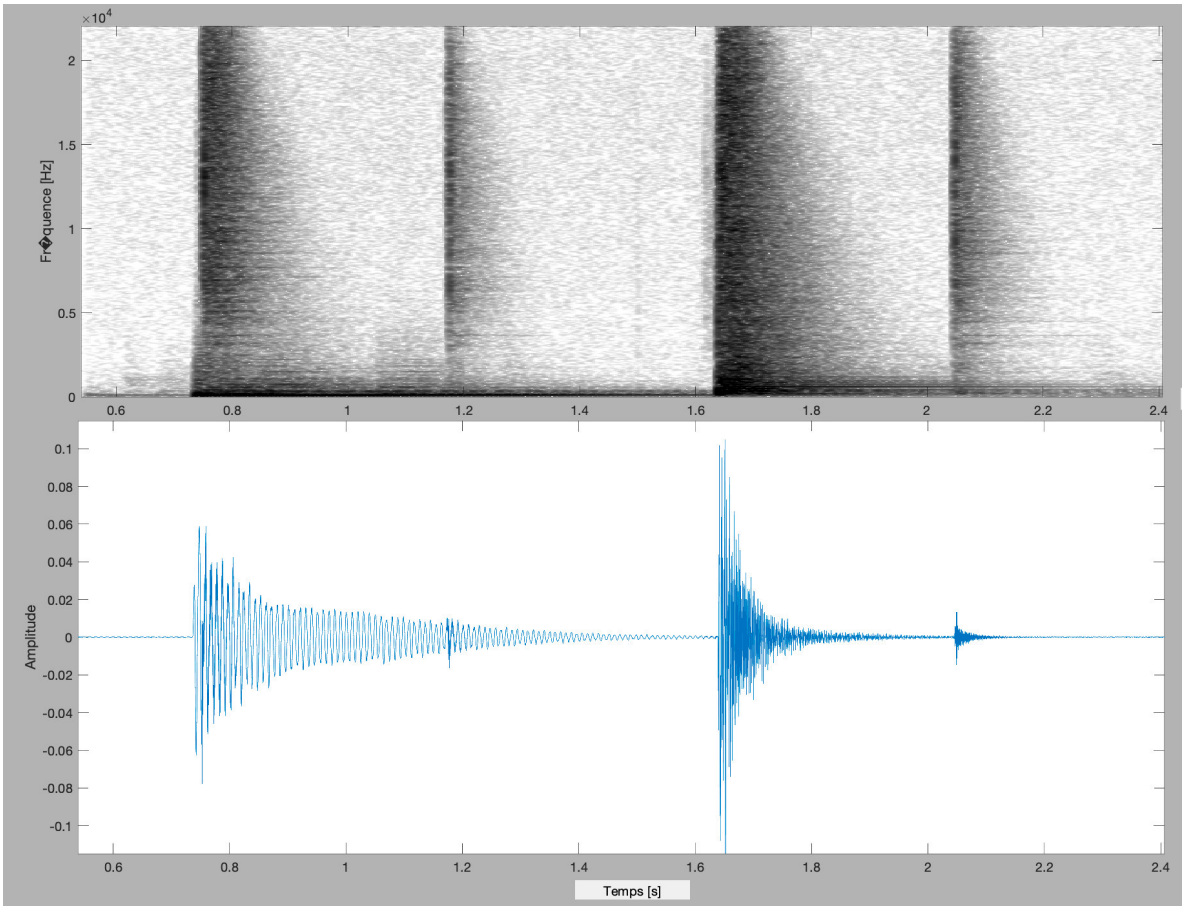

Figura 3. Forma de onda y espectrograma de una señal de audio digital.

En la Figura anterior, hay que resaltar como las manchas oscuras del espectrograma, parte superior, se alinean con los primeros momentos del desarrollo de las señales en la parte inferior. El segundo golpe o nota está traslapado con el primero, esto no impide ver en la parte superior la franja obscura, es mas clara en las frecuencias altas de la señal, evidencia del inicio.

2.3. La transformada de Fourier de Tiempo-Corto

La información espectral de un audio puede ser extraída a través del análisis de Fourier. Loy [22] cita a Fourier quién declara “cualquier vibración periódica, no importa cuán complicada parezca puede ser construida a partir de sinoidales cuyas frecuencias son múltiplos enteros de una frecuencia fundamental, escogiendo adecuadamente las amplitudes y fases“. Las ondas sinoidales diferentes a la fundamental se les denomina armónicos. La transformada rápida de Fourier (FFT por sus siglas en inglés) es un algoritmo eficiente usado para calcular la transformada discreta de Fourier (DFT). La DFT es usada en el análisis de Fourier cuando las ondas son representadas de manera discreta, es el caso del sonido almacenado digitalmente. Dicha transformada permite la representación de una señal en el dominio de frecuencia, cuando la señal originalmente estaba en el dominio del tiempo. Esta transformada normalmente se aplica sobre un segmento del tiempo de la señal. El segmento se extrae de la señal original con la ayuda de una ventana.

En la práctica, las señales de audio se analizan en segmentos de tiempo, con STFT. La Ecuación (1), muestra el análisis que se hace sobre las señales de audio basada en la STFT para extraer la información espectral de la señal X en Dixon [23].

$$\mathrm{X(n,k)} = \sum_{m=-\frac{N}{2}}^{\frac{N}{2}-1} x(hn+m)w(m)e^{-\frac{2j\pi mk}{N}} \qquad (1)$$

En donde $x(n)$ es la señal en el dominio del tiempo, $w(m)$ es un punto en una ventana de tamaño N, h es la magnitud del salto dado en la señal entre los diferentes segmentos, k son las frecuencias y j la constante imaginaria.

2.4. Función de detección basada en la energía

La información obtenida a partir del análisis espectral o de Fourier de una señal de audio se puede convertir en una o varias funciones de detección. Una función de detección se le define

como aquella función que intenta hacer coincidir sus máximos o mínimos con el momento en el tiempo en el cual está presente un inicio de una nota. En los instrumentos de percusión la función de detección llamada Spectral Flux (SF) ha mostrado buenos resultados en la detección de inicios. Esta función toma la energía presente en cada contenedor del análisis restándole la información del contenedor anterior en la misma frecuencia. La Ecuación (2), muestra la definición de la función de detección SF en Dixon [23]:

$$\mathrm{SF(n)} = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} H(|X(n,k)| - |X(n-1,k)|) \qquad (2)$$

En la ecuación anterior, $\mathrm{H(x)}$ se define en la Ecuación (3) y se le conoce como la función de rectificación de la mitad de onda. Esta función considera la parte positiva de la señal, resaltando los inicios. En la ecuación 2, $|X(n,k)|$ y $|X(n-1,k)|$ son la magnitud del número complejo en el contenedor actual y previo.

$$\mathrm{X(n)} = \frac{(x + |x|)}{2} \qquad (3)$$

En la Figura 4, tenemos la función de detección SF para el archivo de audio de los instrumentos de la batería. Vemos una serie de picos que se forman en aquellos segmentos de señal, o también llamados marcos, en donde hay un incremento en la energía. Es interesante remarcar en la Figura de que algunos picos tienen una mayor altura que otros. Esto obedece a la cantidad de energía liberada por el instrumento. Un instrumento como el bombo, libera más energía comparado con la tarola. Esto se ve reflejado en un pico de mayor altura para el primer instrumento, que para el segundo.

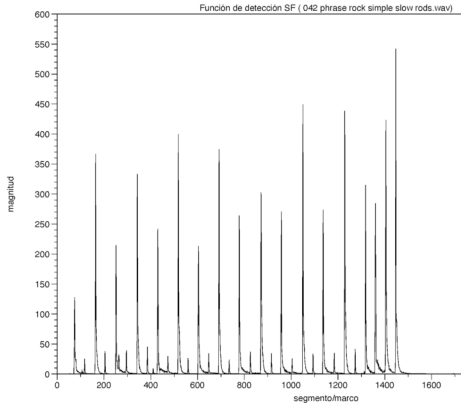

Figura 4. Función de detección SF.

El reto en la función es seleccionar aquellos picos en donde efectivamente haya un inicio de señal y descartar aquellos en donde no lo hay. Un patrón aparece alrededor del inicio. Un incremento súbito en el nivel de energía hasta llegar a un máximo local, seguido de una disminución con una duración de varios marcos. El patrón se repite con algún grado de variación en el tamaño y forma como lo indica Bello et al. [13]. Se discute a continuación, como una técnica de aprendizaje de máquina nos puede ayudar a reconocer este patrón y con él, los inicios de las notas.

2.5. Reconocimiento de patrones

Kuncheva [24] menciona el reconocimiento de patrones consiste en la asignación de etiquetas a objetos. A los objetos se les describe por un conjunto de características, también llamadas atributos. Existen dos tipos de problemas de reconocimiento de patrones: Los supervisados y no supervisados. En los problemas de la categoría de supervisados o aprendizaje supervisado se distinguen porque junto con los objetos podemos encontrar las etiquetas asignadas a ellos. Por el otro lado, en los problemas no supervisados la etiqueta no está disponible. En el primero de ellos un clasificador debe de aprender a distinguir las características que describen a un objeto de otro para identificar la etiqueta correspondiente a cada uno.

En la terminología de las técnicas de aprendizaje supervisado se le llama clase, a aquellos objetos que comparten características. Podemos encontrar c diferentes clases. Cada clase es

una etiqueta diferente a la que se le asigna a uno o varios objetos. Pero una etiqueta únicamente se le asigna a cada uno de ellos. Todas las clases forman parte del conjunto $\Omega = \{w_1, w_2, \dots, w_n, \}$. Los valores de las características de un objeto pueden ser consideradas como cualitativas o cuantitativas. Las características se presentan como un vector de n dimensiones $x = \{x_1,\ x_2, \dots, x_n\}$ donde $x_i \in \mathcal{R}^n$ en el caso de las cuantitativas. A $\mathcal{R}^n$ se le llama, el espacio de las características. Se le conoce como conjunto de datos al que pertenecen los datos de las características y las etiquetas de las clases, se le representa como $Z = \{z_1,\ z_2, \dots, z_n\}$, $z_j \in \mathcal{R}^n$.

Se le llama clasificador a la función que toma como argumentos las características y nos conduce a la etiqueta correspondiente, su definición formal aparece en la Ecuación (4):

$$D: \mathcal{R}^n \to \Omega \qquad (4)$$

Se le conoce como función discriminante a aquella función que permite identificar apropiadamente la etiqueta asignada a los objetos. Es posible tener varias funciones discriminantes, al conjunto que pertenecen todas ellas es $G = \{g_1(x), g_2(x), \dots, g_c(x)\}$. A los elementos de G, las funciones g se les define en la Ecuación (5):

$$g_i: \mathcal{R}^n \to \mathcal{R},\ \ i = 1, 2, \dots, c \qquad (5)$$

Llamamos regla de mayor membresía a la que nos permite encontrar la función discriminante que identifica correctamente la mayor cantidad de etiquetas para los objetos, se define en la Ecuación (6):

$$D(x) = w_{i^*} \in \Omega \leftrightarrow g_{i^*}(x) = \max_{i=1,2,\dots c} g_i(x) \qquad (6)$$

La región de decisión de una clase se le concibe como el conjunto de puntos para los cuales la función discriminante obtiene los mejores resultados. Todos los puntos pertenecientes a dicha región se les asigna la clase correspondiente. A los límites de estas regiones de decisión se les llama límites de clasificación. Si algún punto se encuentra dentro de los límites de clasificación, es probable que una clase diferente se le asigne. A estos casos se les conoce como traslape en las regiones de decisión. Un clasificador ideal sería uno en donde las

regiones de clasificación no se traslapan y por lo tanto siempre se les asigna a los objetos las etiquetas correspondientes.

La Figura 5, permite visualizar los conceptos anteriores, los puntos y los cuadros son las clases. Las áreas gris y blanca las regiones de clasificación. La franja que separa las regiones de clasificación representa los límites de clasificación. Como es visible hay un punto sobre el área gris, cuando sobre el área gris están los cuadrados, esto señala que las regiones están traslapadas. Esto dentro de un espacio de características $\mathcal{R}^2$ para un clasificador D.

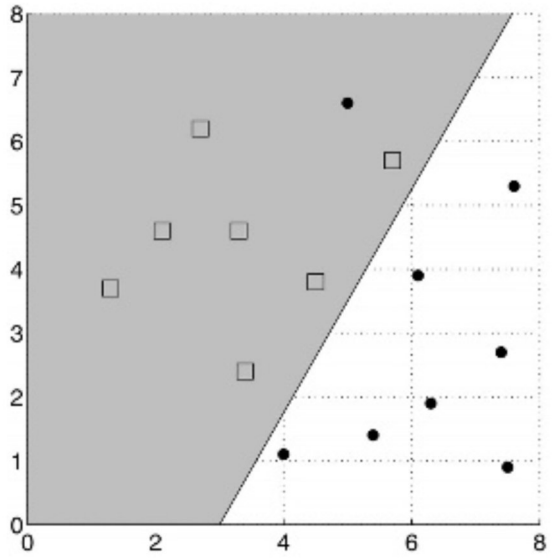

Figura 5. Representación de clases, regiones y límites de clasificación. Tomada de Kuncheva [23].

A la tarea de enseñarle a un clasificador a reconocer las características y las clases de los objetos se le conoce como proceso de entrenamiento. En la fase de aprendizaje el conjunto de datos Z, junto con una regla de aprendizaje se usa para que el clasificador sea capaz de asociar los atributos con la clase correspondiente. En una segunda fase, se verifica que el clasificador ha aprendido a asociar correctamente los atributos con las clases, llamada fase de pruebas. A continuación, explicaremos una de las técnicas usadas en una amplia variedad de problemas de reconocimiento de patrones y reconocida por su simplicidad, el k-vecinos más cercano.

2.6. Redes Bayesiana

Friedman, Geiger, Goldszmidt [25] definen una Red Bayesiana como un clasificador del tipo de aprendizaje supervisado que pretende representar las relaciones de Causa/Efecto en un dominio de problema. En este clasificador se hace uso de la regla de Bayes expresando la probabilidad condicional entre la Causa dado el Efecto. La Red Bayesiana se representa gráficamente a través de un grafo dirigido acíclico. La Figura 6, muestra una abstracción de la estructura de una Red Bayesiana, en donde los nodos representan las Causas y los Efectos. Las aristas representas las relaciones entre los efectos y las causas. En la Red no están permitidos ciclos, por lo que note que no existe una relación entre las causas, ni entre los efectos, solo de las Causas hacia los Efectos.

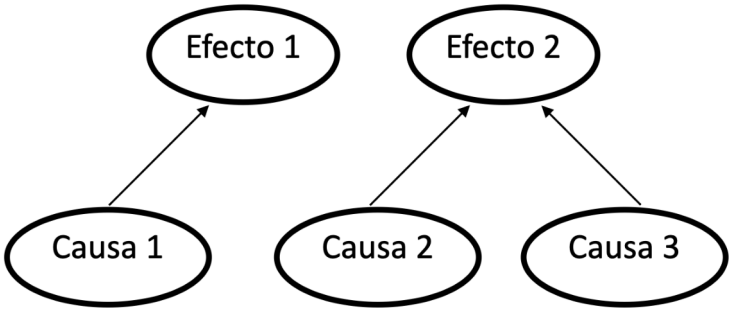

Figura 6. Grafo de una Red Bayesiana.

En el clasificador se suponen que las causas con independientes entre sí. En la Figura anterior, efecto 2, tiene dos causas, causa 2 y causa3, se entiende que causa 2 es independiente de causa 3 y viceversa.

"El objetivo es calcular la posterior distribución de probabilidad condicional de cada una de las posibles causas no observables a partir de la evidencia observada" [26]. La Red Bayesiana se construye a partir del Teorema de Bayes. La Ecuación 7 muestra el cálculo de probabilidades condicionales de las causas a partir de que el Efecto se conoce.

$$P(Causa \backslash Efecto) = \frac{P(Efecto|Causa)\,.\,P(Causa)}{P(Efecto)} \quad (7)$$

En una Red Bayesiana cada nodo es condicionalmente independiente de todos sus no descendientes dado el padre del nodo. Por lo que, la distribución de probabilidad conjunta de todas las variables aleatorias del grafo se factoriza en una serie de distribuciones de probabilidad condicional de variables aleatorias dados sus padres. Por lo anterior, es posible construir un modelo de probabilidad completo especificando solo la distribución de probabilidad condicional en cada nodo, Spiegelhalter citado por Horny [26].

2.7. Perceptrón multicapas.

Teniendo como inspiración el sistema nervioso del ser humano. Intentado imitar la forma como el cerebro realiza un conjunto de tareas con alta eficiencia como el reconocimiento de patrones. Las Redes Neuronales Artificiales han sido útiles en problemas de: categorización, aproximación, predicción, clasificación de patrones y otras mencionado en Jain, Mao y Mohiuddin [27]. Un modelo matemático representa la biología del sistema nervioso y sus componentes. El elemento básico de una ANN es una neurona. Un conjunto de neuronas procesando datos en paralelo forman capas. Un conjunto de capas forma una ANN. La ecuación 8, muestra el modelo de una neurona.

$$y = fa\left(\sum_{j=1}^{n} w_j x_j + s\right) \quad \textbf{(8)}$$

En la ecuación anterior, x_i son los valores entrantes, w_i los pesos sinápticos, s el sesgo y fa una función de activación. Los pesos sinápticos se multiplican con los valores de entrada. El sesgo representado de esta manera es un valor que normalmente no es una entrada sino un valor que se multiplica por 1, este 1 es parte de los pesos. Las funciones de activación mas comunes son: Umbral, lineal, sigmoidal y Gaussina.

Usando como criterio los patrones de conexión entre las capas y las neuronas, las NNA se agrupan en dos categorías: hacia adelante (feed-forward en inglés) y las recurrentes. Las primeras se caracterizan que los valores siempre van hacia adelante sin la presencia de ciclos. En las segundas, aparecen ciclos entre las capas o neuronas dirigiendo a algunos valores hacia atrás en la arquitectura. Las arquitecturas del grupo hacia adelante incluyen: perceptrón de una sola capa, perceptrón multicapa y redes de función base radial. Formando el grupo de

redes recurrentes encontramos: redes competitivas, SOM Kohonen's, redes de Hopfield y modelos ART.

Al igual que las personas aprenden a reconocer la voz de las otras personas que están a su alrededor o los sonidos de una pieza musical. Las NNA tienen que ser entrenadas para reconocer un patrón. Los tres principales paradigmas de aprendizaje son: supervisado, no supervisado e hibrido. En el supervisado a la red se le dan los datos de entrada y salida correspondientes. Por lo que la red aprende a asociar la entrada con su salida correspondiente. En el aprendizaje no supervisado, no existen salidas correctas a aprender, en su lugar, la red crea categorías con los patrones de entrada. En el aprendizaje hibrido se combinan los dos anteriores. Para aprender los patrones las NNA usan las denominadas reglas de aprendizaje, los cuatro tipos básicos de ellas son: error-corrección, Boltzmann, Hebbian y aprendizaje competitivo.

De las reglas de aprendizaje de error-corección los algoritmos mas representativos son: el aprendizaje del perceptrón y retro-propagación (back-propagation en inglés). El en primero de ellos, los datos entran a la neurona, se calcula el error o diferencia entre la salida esperada y producida para después ajustar los pesos y sesgos. El segundo, toma la idea básica del primero y la lleva mas allá. Este actualiza las salidas no solo de una neurona, lo hace con una capa de neuronas o hasta con una arquitectura completa.

2.8. Tabla de decisiones

Uno de los modelos de clasificación más sencillo en la teoría de algoritmos de aprendizaje supervisados son las tablas de decisiones. Al igual que todos los demás algoritmos supervisados las tablas de decisiones inician con un conjunto de datos Z. Una tabla de decisiones normalmente esta formada por dos componentes de acuerdo con Kohavi [28]:

1. Un esquema, un subconjunto de los atributos disponibles en el conjunto de datos.
2. Un cuerpo, está formado por varios conjuntos de las instancias etiquetadas. Las instancias en la tabla contienen los valores para los atributos y la etiqueta o clase asignada.

Los tipos de datos de los atributos en una tabla de decisiones pueden ser: categóricos, alfanuméricos o numéricos. Normalmente los atributos con datos continuos suelen ser

discretizados, evitando tener una tabla irrepresentable. Una práctica común es ordenar los atributos en orden ascendente o descendente. En algunas otras ocasiones se orden de acuerdo al peso que aportan para determinar la clase o por la clase misma.

Una vez construida la tabla de decisión se usa como clasificador T. Permitiendo encontrar la clase para un objeto no etiquetados e. El procedimiento para asignar la etiqueta consiste en buscar en la tabla T, la o las instancias del cuerpo de T donde coincidan los atributos con los atributos de objeto e. Usando como ejemplo el método de la clase mayoritaria. Si el resultado de la búsqueda no arroja ninguna coincidencia, se regresa como clase asignada a e la mayoritaria de la tabla. En el caso de que varias coincidencias sean encontradas, se le asigna a e la mayoritaria de ellas.

Una de las discusiones más relevantes en la construcción de tablas de decisiones es la selección de los atributos que son significativos. Varios métodos han sido propuestos en la literatura, todos ellos pretenden encontrar cuáles son los atributos que ayudan a elegir la mayor cantidad de ocasiones la clase que corresponde correctamente al objeto a clasificar.

La Tabla 1, muestra una visualización de una Tabla de decisiones. En ella se simplifica la estructura mapeándola en la tabla relacional, como lo hace Becker [29]. Las primeras columnas contienen a los atributos. Esta representación hace un resumen de todas las instancias para cada combinación de valores de los atributos, la columna peso, indica el número de registros contados en esa combinación y la última indica la proporción de registros de cada clase.

Tabla 1. Visualización de una tabla de decisiones en una tabla relacional.

Atributo 1	Atributo 2	Atributo 3	Atributo 4	Peso	Probabilidades
a	Si	50-60	5	10	0.3, 0.6, 0.2
a	Si	50-60	6	34	0.4, 0.7, 0.9
a	Si	60-70	7	123	0.2, 0.5, 0.8
a	No	60-70	7	5	0.3, 0.5, 0.1
a	NO	70-80	8	1	0.2, 0.6, 0.4
:	:	:	:	:	:
f	Si	50-60	5	10	0.3, 0.6, 0.2
f	Si	50-60	6	34	0.4, 0.7, 0.9
f	Si	60-70	7	123	0.2, 0.5, 0.8

2.9. Árbol de Hoeffding

Basados en la suposición de que un pequeño número de ejemplos pueden ser suficientes para escoger una división de valores de los atributos optima, los árboles de Hoedffing son un método de aprendizaje basado en árboles de decisiones propuesto por Domingos y Hulter [30]. El Árbol de Hoeffding, es un algoritmo muy rápido, se usa en la minería de flujos de datos, realizando la tarea de clasificación. Puede producir prácticamente idénticos resultados que los árboles de decisiones donde el aprendizaje se hace en lote. Pero a diferencia de ellos son capaces de aprender de manera gradual usando los nuevos datos del flujo.

El supuesto de los Árboles de Hoeffding es apoyado por el límite de Hoeffding, la Ecuación 9 muestra el valor estimado de la entropía del nodo dado un nivel de confianza.

$$\varepsilon = \sqrt{\frac{R^2 \ln(1/\delta)}{2N}} \quad (9)$$

Donde R es el rango de valores de una variable aleatoria, δ es la probabilidad deseada de que la estimación no esté dentro de ε el valor esperado y N es número de ejemplos recopilados en el nodo del árbol. En cada nodo el árbol está los datos estadísticos necesarios para dividir los atributos. En el caso de atributos discretos, se mantiene una tabla de 3 dimensiones para cada tripleta (x_i, v_i, k) un conteo $n_{i,j,k}$ con las instancias del entrenamiento con $x_i = x_j$, junto con un vector con los conteos de las clases.

El algoritmo toma un conjunto de ejemplos E, cada ejemplo contiene un conjunto de atributos y una clase que se les asigna, un parámetro de precisión δ, una función de evaluación G la cual brinda una medida para seleccionar el mejor atributo para asignar correctamente la clase. Cada nodo en el árbol maximiza G para uno de los atributos. Lo que se pretende es encontrar la menor cantidad de ejemplos N, con los cuales se cumpla con el límite de Hoeffding. La Figura 7, muestra el pseudocódigo del algoritmo de construcción del árbol de decisión.

```
ArbolHoeffding (E, δ )
  Parámetros: E es un conjunto de ejemplos y δ es el parámetro de confianza.
    1  Sea AH un árbol con una hoja (la raíz)
    2  Iniciar cuenta de la raíz n_ijk
    3  Para cada e (x, y) en E
    4      Llamar AgregarAH ((x, y), AH, δ)
AgregarAH ((x, y), AH, δ)
    1  Ordenar (x, y) en la hoja h usando AH
    2  Actualizar cuenta n_ijk en la hoja h
    3  Si los ejemplos vistos hasta h no son todos de la misma clase
    4    Entonces
    5      Calcular G para cada atributo
    6      Si G(mejor atributo)− G(Segund mejor) > ε
    7      Entonces
    8          Separa la hoja usando el mejor atributo
    9          Para cada rama
   10              Crear una nueva hoja
   11              Inicializar conteos
```

Figura 7. Pseudocódigo de un árbol de Hoeffding.

2.10. Algoritmo k-vecinos más cercano

Taunk et al. [31] define al algoritmo k-vecinos más cercanos (KNN por sus siglas en inglés) como uno no paramétrico de clasificación y regresión, el cual no hace ningún tipo de suposición sobre el conjunto de datos, reconocido por su simplicidad y efectividad, siendo del tipo de aprendizaje supervisado. Para entrenarlo, un conjunto de datos que contiene las clases y los atributos asociados a los objetos es usado. Posteriormente es capaz de predecir la clase para datos diferentes que no los acompaña la clase a la que pertenece.

Cuando KNN es usado en clasificación, usa diferentes características para determinar a cuál clase pertenece el objeto. El criterio usado es la cercanía de otros objetos en las regiones de clasificación. El valor de k es definitivo en el funcionamiento del algoritmo. Este es usado para determinar la pertenencia de un objeto a una clase. k indica el número de vecinos usados para determinar la clase. Un valor $k = 1$ indica que solo se usará al vecino más cercano para decidir la clase, con valor $k = 2$ se usarán los dos vecinos más cercanos y así sucesivamente. Con un mayor valor para k se da mayor claridad en la región de clasificación. Aunque usar valores muy grandes puede ser no deseable por razones de desempeño. Para determinar la clase a la cual pertenece un objeto se toma a la cual pertenecen la mayoría de sus vecinos cercanos. Dos pasos describen el algoritmo. Primero, analizar los k-vecinos más cercanos.

Segundo, usar las clases de los vecinos para determinar la clase del objeto a clasificar. A continuación, se explican estos pasos para comprender la idea general del algoritmo.

Sea $D = \{(x(1),\ y(1)), (x(2),\ y(2)),\ \ldots, (x(m),\ y(m)),\}$ un conjunto de datos con el cual se entrena el algoritmo KNN.

El paso 1: Es cargar el conjunto de entrenamiento.

El paso 2: Con cada nuevo objeto sin clasificar, hacer lo siguiente:

- Calcular la distancia Euclidiana de acuerdo con la Ecuación 10.

$$d = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \qquad (10)$$

- Encontrar los k vecinos mas cercanos.
- Asignar la clase a la que pertenecen el mayor número de vecinos cercanos.

El valor k es definitivo en el algoritmo este ayuda a construir los límites de clasificación y las pertenencias de los objetos a las clases. En la Figura 8 se ilustra el algoritmo anterior. Se desea clasificar el punto azul, para asignarle una clase, se usan los 4 vecinos más cercanos, tres verdes y 1 rojo. Dado que hay mayoría de verdes, se decide que el punto azul pertenece a la clase verde.

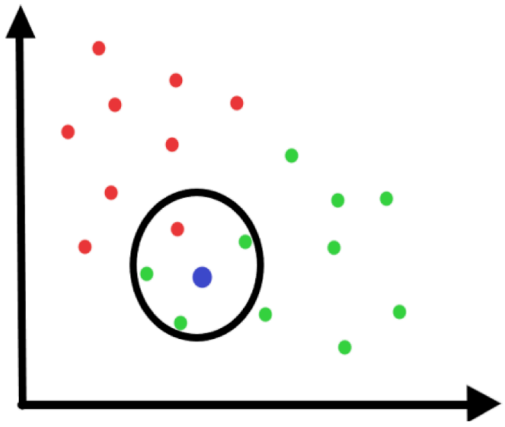

Figura 8. Clasificación de un objeto con el algoritmo KNN. Tomado de Taunk *et al.* [31].

2.11. Índice de Variabilidad en Pares

El índice de variabilidad en pares es utilizado para representar el ritmo en lenguaje, partituras de piezas musicales, interpretaciones musicales y grabaciones tomando la representación de forma de ondas del sonido. Permite hacer comparaciones entre diferentes combinaciones rítmicas e identificar similares. Grabe y Low [5] lo hacen con los lenguajes Inglés, Alemán, Holandés, Frances y Español. Patel y Daniele [7] lo hacen pero con el ritmo del lenguaje y de la música instrumental usando partituras. Daniele y Patel [6] comparan el ritmo del lenguaje y la música de varios idiomas y tradiciones musicales de diferentes culturas. McDonough, Danko y Zentz [32] contrastan los ritmos del habla y grabaciones de piezas musicales de artistas de Jazz Americano. Raju et al. [8] analizan los ritmos de partituras e interpretaciones musicales. Toussaint [4] compara los ritmos de tradiciones musicales de diferentes culturas. Daniele y Patel [33] analizan los cambios de ritmo en los patrones del habla y piezas musicales a lo largo de la vida de compositores.

El índice de variabilidad en pares mide el cambio en el tiempo en la aparición de parejas de eventos. El índice consiste en tomar la duración en el tiempo entre un inicio y otro para el caso de la música. En el lenguaje dependiendo el idioma se toma la duración en el tiempo de la aparición de 2 vocales o 2 consonantes. Las duraciones se toman en parejas, se encuentra la diferencia entre ellas, mientras se van sumando estas diferencias en las parejas adyacentes de duraciones. Son dos los índices disponibles en la literatura: El índice normalizado nPVI, y el índice crudo de variabilidad en pares (rPVI). El nPVI lo define Grabe y Low [5] como se muestra en la Ecuación 11:

$$nPVI = \left(\frac{100}{m-1}\right)\sum_{k=1}^{m-1}\left|\frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2}\right| \quad (11)$$

Donde d es la duración entre un evento y otro, d_k, la duración actual, d_{k+1}, la duración siguiente, m el número de duraciones. Una deficiencia encontrada en el nPVI es la dificultad para diferenciar ritmos binarios de ternarios como lo muestra Toussaint [4]. Un ritmo binario tiene dos pulsos, el primero acentuado o fuerte, y el segundo débil. Mientras que, el ritmo ternario tenemos tres pulsos, el primero fuerte y los otros dos débiles. El rPVI es definido en Grabe y Low [5] como en la Ecuación 12:

$$rPVI = \sum_{k=1}^{m-1} \frac{|d_k - d_{k-1}|}{m-1} \quad \textbf{(12)}$$

Los índices de variabilidad en pares nPV y rPVI han demostrado ser útiles en las tareas de representar, identificar similitudes y diferencias en ritmos del habla y la música. Por lo anterior, se esperan sean buenos descriptores de alto nivel y se pueda con ellos, organizar y recuperar piezas musicales usando como criterio el ritmo. El siguiente capítulo, que contiene el método, se muestra cómo partiendo de las señales de audio digitales llegamos a tener ambos índices.

Capitulo 3. Método

> “La educación musical es soberanía, porque mejor que cualquier otra cosa, el ritmo y la armonía encuentran camino hacia lo más profundo del alma y se apoderan de ella con la mayor fuerza posible.” Platón

En el presente capítulo se describen varias tareas, estas inician en las señales de audio y terminan en los descriptores de alto nivel del ritmo, los índices nPvi y rPvi. El presente trabajo hace uso de un conjunto de audios que contienen la ejecución de varios ritmos musicales en una batería. La Figura 9, presenta estas tareas. En la primera de ellas, inciso 1, se marcan los inicios de las notas musicales. En la segunda, inciso 2, se extrae la información espectral a través de la STFT. La tercera, inciso 3, la energía presente en cada marco se suma para conseguir una función de detección. La cuarta, inciso 4, los valores de la función de detección se agrupan en una ventana deslizante como preparación de los patrones que son introducidos en un clasificador. En la quinta, inciso 5, se entrena el clasificador para reconocer el patrón en las entradas. En la sexta, inciso 6, las salidas del clasificador identifican los marcos en donde están presentes los inicios de una nota musical. La séptima, inciso 7, se estiman las distancias entre los inicios. Por último, inciso 8, se calculan los descriptores nPvi y rPvi.

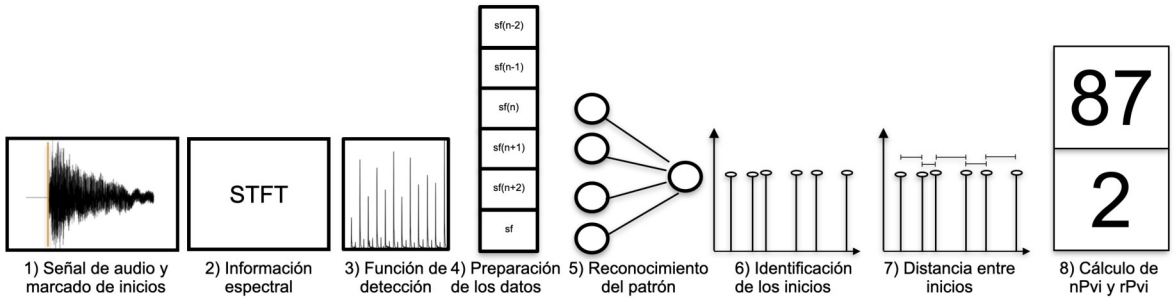

Figura 9. Algoritmo propuesto para extraer los descriptores nPvi y rPvi de la señal de audio.

3.1. El conjunto de datos

Un subconjunto de los audios que forman parte de la base de datos Enst-drums de Gillet y Richard [34] es tomado en el experimento. La base de datos Enst-drums contiene las grabaciones de bateristas mientras están ejecutando diversos géneros musicales. Las grabaciones incluyen varias frases de cada ritmo. Estas frases tienen una duración de unos

pocos segundos. En estos segundos están varios golpes con los instrumentos según el género. Contiene para cada baterista las grabaciones de cada instrumento, una mezcla seca y otra húmeda. En la industria de la grabación de sonido se entiende como mezcla seca aquella donde se combinan varios sonidos grabados en uno o varios canales de audio sin agregar algún efecto. Por el otro lado, una mezcla húmeda es aquella donde se combinan los sonidos y además se agregan efectos como reverberación, retardo y eco, por mencionar algunos. Las grabaciones se realizaron con una frecuencia de muestreo de 44.1 Khz y un tamaño de bit de 16. El subconjunto está formado por un archivo de cada ritmo para un solo baterista, tiene una cardinalidad de 6.

La Figura 10, muestra una imagen de uno de los bateristas de la base de datos Enst-drums. La batería contiene los instrumentos estándar, estos son: Bombo, Caja, tambores (toms) y los Platos.

Figura 10. Baterista frente a sus instrumentos en la base de datos de Enst-drums. Tomado de Gillet y Richard [34].

Los audios son parte de la mezcla seca. La Tabla 2, muestra los archivos seleccionados.

Tabla 2. Archivos de audios del experimento.

No.	Archivo de audio	Duración en segundos
1	036_phrase_disco_simple_slow_sticks.wav	16 s
2	042_phrase_rock_simple_slow_rods.wav	20 s

3	048_phrase_afro_simple_slow_mallets.wav	10 s
4	060_phrase_salsa_simple_slow_sticks.wab	20 s
5	066_phrase_shuffle-blues_simple_slow_brushes.wav	14 s
6	078_phrase_reggae_simple_slow_sticks	13 s

3.2. Marcado de los inicios

El conjunto de audios fue marcado; poniendo una marca en el momento del tiempo donde están presentes los inicios. Para realizar está tarea se usó el script de Matlab de Leveau y Daudet [21]. La Figura 11, muestra el marcado de los primeros inicios del archivo 042_phrase_rock_simple_slow_rods.wav. En la parte inferior las franjas verticales de color gris sobre la forma de onda de las señales, indican el lugar donde se colocaron las marcas de los inicios.

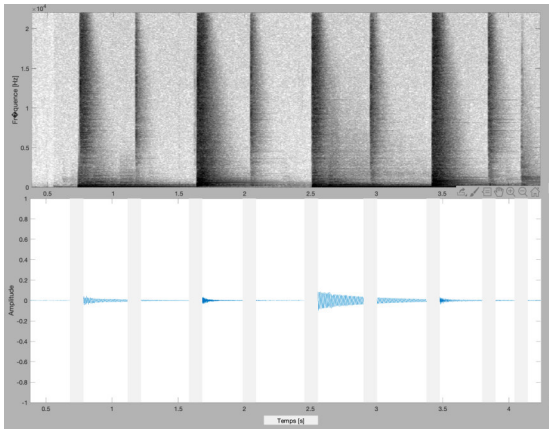

Figura 11. Marcado de los inicios del audio 042_phase_rock_simple_slow_roads.

3.3. Información espectral y función de detección

Las señales de audio fueron transformadas usando la STFT en Matlab. Los parámetros elegidos para ejecutar la transformada fueron: una ventana de hamming de tamaño 2048, un número de contenedores de 2000 para las frecuencias, un salto en la señal para producir una nueva ventana con un valor de 441 como se propone en Bello et al. [13]. El salto produce una taza de 100 marcos por segundo. Son 10 ms entre un marco y otro considerando el tiempo

de la señal de audio. Esto produce una matriz usada para generar la función de detección SF. Por lo que la función de detección tiene 100 puntos de datos por segundo.

3.4. Preparación de los datos

Los datos son preparados para ser ingresados a un clasificador como se hace en Stasiak et al. [9]. Una ventana deslizante sobre SF genera vectores de valores con los atributos. Dos tipos de preparaciones se hicieron. En ambas preparaciones se usaron tamaños de ventanas de 5, 7 y 9 valores de SF. La primera preparación, toma únicamente valores de la función de detección, en el caso de la ventana de tamaño 5 tiene la forma $x_1 = \{SF(n-2),\ SF(\mathrm{n}-1),\ SF(n),\ SF(n+1),\ SF(n+2)\}$. La segunda preparación, agrega un valor adicional $\overline{SF}$, la sumatoria de 10 valores anteriores y posteriores al marco actual que se está analizando. El vector para el mismo tamaño 5 es $x_2 = \{SF(n-2),\ SF(\mathrm{n}-1),\ SF(n),\ SF(n+1),\ SF(n+2),\ \overline{SF}\}$. En los vectores, n es el marco donde se considera un posible inicio. Teniendo así 3 conjuntos de la primera preparación y otros 3 de la segunda preparación. Estas preparaciones están formadas por $9,039$ patrones de la clase 0 y 201 de la clase 1.

Considerando el problema denominado desbalance en la cantidad de muestras de las clases, discutido en Ganganwar [35], se procedió a realizar un sobre muestreo. Esto crea 2 grupos de conjuntos de datos. En estos conjuntos se agregaron 1000 copias de los patrones con inicios y otros 1000 de los mismos patrones mas ruido. El ruido agregado no es mayor del 10% del nivel de energía en el marco. Esto genera 3 conjuntos de la primera preparación mas ruido y otros 3 de la segunda preparación también mas ruido. Esta preparación está formada por 9039 patrones de la clase 0 y 2201 de la clase 1. En total se tienen disponibles 12 conjuntos de datos. Los 12 son: dos grupos de 6, uno sin sobre muestro y el otro con sobre muestreo. Cada grupo de 6 tiene dos preparaciones: 3 con 5, 7 y 9 atributos, otros 3 formados: de 3, 7 y 9 atributos más un atributo adicional, la sumatoria $\overline{SF}$.

3.5. Entrenamiento y validación de clasificadores

Varios clasificadores fueron entrenados y validados con los conjuntos anteriores. Se hace uso de Weka de Frank et al. [36], una colección organizada de algoritmos de aprendizaje de máquina. Cinco diferentes algoritmos se entrenaron en busca de identificar cual muestra los mejores valores en las métricas ampliamente usadas y aceptadas en la comunidad de investigadores de reconocimiento de patrones. Dichas métricas usadas son: el porcentaje de instancias clasificadas correctamente, Precisión, Exhaustividad (recall en inglés) y Médida F (F-Measure en inglés). Los algoritmos son: Una red Bayesiana, un perceptrón múlticapa, un clasificador del k-vecinos más cercanos, una tabla de decisión y un árbol de Hoeffding. Estos clasificadores se entrenaron para asociar los atributos que representan la cantidad de energía de los marcos con las clases $\{0, 1\}$. El método de validación cruzada con 10 folds (en inglés cross-validation con 10 folds) es usado durante esta tarea.

Powers [37] discute las métricas usadas para evaluar el desempeño de un clasificador incluyendo: Precisión, Exhaustividad y Medida F. Antes de definirlos es importante entender que un clasificador puede clasificar correcta o incorrectamente las entradas con las clases asociadas. En nuestro contexto de la detección de marcos de tiempo de una señal de audio con inicio de notas musicales. Un Verdadero-Positivo (VP) es un marco que contiene un inicio y el clasificador lo clasifica como tal. Un Falso-Positivo (FP) es un marco sin un inicio pero que el clasificador lo identifica como con un inicio. Un Falso-Negativo (FN) es un marco sin inicio pero clasificado con uno. Un Verdadero-Negativo (VN) es un marco sin inicio y clasificado sin inicio. La Exhaustividad se entiende como la taza de Verdaderos Positivos y se calcula como lo muestra la Ecuación 13.

$$\text{Exhaustividad} = \frac{VP}{VP + FN} \quad \textbf{(13)}$$

A la Precisión se entiende como la proporción de Positivos que son correctamente clasificados como Positivos y la Ecuación 14 indica su cálculo.

$$Precisión = \frac{VP}{VP + FP} \quad \textbf{(14)}$$

La medida F o medida de precisión de una prueba se calcula siguiendo la Ecuación 15.

$$Medida\ F = 2\ x\ \frac{Precisión\ x\ Exhaustividad}{Precisión + Exhaustividad} \quad (15)$$

3.6. Cálculo de distancias y descriptores del ritmo.

Una vez identificados los marcos de tiempo con inicios, bajo el supuesto de que el inicio se encuentra en el centro del marco, se ubican los inicios en el dominio del tiempo. Se le asigna el valor de 1 a la distancia entre el primero y segundo. Este sirve como referencia para calcular la distancia entre las demás parejas. Si la distancia entre dos inicios es la mitad de la existente entre el primero y segundo, se le asigna un valor de 0.5. En el caso de que la distancia sea el doble entre los primeros dos, se le asigna 2. Estos valores de distancia son los usados para calcular los índices nPvi y rPvi. En el siguiente capítulo, se presentan los resultados de las tareas de marcado, entrenamiento de los algoritmos de aprendizaje de máquina, las distancias estimadas y los valores nPvi y rPvi de los audios.

Capitulo 4. Resultados

> “La música crea orden fuera del caos: El ritmo impone unanimidad sobre la divergencia, la melodía impone continuidad sobre la desarticulación y la armonía impone compatibilidad sobre la incongruencia.” Yehudi Menuhin

En el presente capítulo presentaremos y analizaremos los resultados de la realización de las tareas descritas en el capítulo anterior, el método. En la primera tarea se consiguen los archivos de audio marcados, por lo que tenemos la cantidad de inicios de las notas musicales. En la segunda y tercera tarea llegamos a la función de detección que es usada para crear los patrones que entran al clasificador en la tarea 4. En la tarea 5, para reconocer el patrón entrenamos varios clasificadores, veremos el desempeño de ellos. En la tarea 6, se tienen los inicios y en la 7 mostramos el cálculo de las distancias. Por último, en la 8 mostraremos los valores nPvi y rPvi de las frases de cada audio.

4.1. El marcado de los inicios de las notas musicales

La Tabla 3, muestra los seis archivos del experimento y la cantidad de inicios en cada uno de ellos. La suma de las duraciones de los audios es de 1 minuto y 13 segundos. La cantidad de inicios no es la misma en cada archivo. En total tenemos 201 inicios en el conjunto de datos.

Tabla 3. Cantidad de inicios marcados.

No.	Archivo de audio	Cantidad de inicios
1	036_phrase_disco_simple_slow_sticks.wav	41
2	042_phrase_rock_simple_slow_rods.wav	33
3	048_phrase_afro_simple_slow_mallets.wav	15
4	060_phrase_salsa_simple_slow_sticks.wav	45
5	066_phrase_shuffle-blues_simple_slow_brushes.wav	35
6	078_phrase_reggae_simple_slow_sticks.wav	31
	Total de inicios:	201

4.2. Preparación de los datos

La Tabla 4, presenta la cantidad de patrones creados a partir de la ejecución de las tareas 2, 3 y 4. Le recordamos al lector, al conjunto de datos con sobre muestreo se le agregaron 2000 patrones, de los cuales 1000 son copias de los actuales y los otros 1000 tienen ruido.

Tabla 4. Conformación de los conjuntos de datos.

	Sin sobre muestreo	Con sobre muestreo
Clase 0: Sin inicio	9,039	9,039
Clase 1: Con inicio	201	2,201
Total	9,240	11,240

4.3. Entrenamiento y validación de los algoritmos de aprendizaje máquina

La tarea 5, es la más compleja del algoritmo propuesto. En este proyecto de tesis se entrenaron varios clasificadores. Los clasificadores son: Una red bayesiana, un perceptrón multicapa, una tabla de decisión, un árbol de Hoeffding y un clasificador de k-vecinos más cercanos. Cada clasificador se entrena con los 12 conjuntos de datos disponibles. Recordemos, dos tipos de vectores de entrada se prepararon: el primero, tiene vectores de tamaño 5, 7 y 9. En el segundo tipo de vector a estos mismos tamaños de ventana se les agrega un valor adicional, por lo que los referiremos como 6, 8 y 10 atributos. Las clases a las que se asocian los atributos con $\{0, 1\}$, el 0 indica la no existencia de un inicio en el segmento de la señal de audio, 1 existe un inicio. Se hicieron algunos cambios a los parámetros para entrenar los algoritmos, ninguno de ellos produjo resultados significativos. Los resultados reportados son con los parámetros por omisión que ofrece Weka. Estos parámetros se encuentran en el Apéndice B.

Con cada algoritmo entrenado se calculan los valores de porcentaje de instancias correctamente clasificadas, Precisión, Exhaustividad y Medida F. Estas medidas se calculan para las clases y una media ponderada. Permiten identificar el clasificador con un mejor desempeño. El método de prueba usado es la validación cruzada con 10 folds. Teniendo en mente que el problema de desbalance de clases afecta el desempeño del clasificador, el análisis de los resultados debe de considerarlos. El programa Weka cuenta con la implementación de los algoritmos clasificadores y calcula las medidas de validación de desempeño.

La Tabla 5, presenta los resultados del entrenamiento de Redes Bayesianas con los 6 conjuntos de datos sin sobre muestreo. Los valores en la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 5. Desempeño del clasificador Red Bayesiana con preparaciones sin sobre muestreo.

Atributos	% Correctos	*Precisión*	*Exhaustividad*	Medida F
5	87.2078	0.973	0.872	0.914
6	85.0649	0.974	0.851	0.902
7	94.4697	0.978	0.945	0.958
8	93.7446	0.978	0.937	0.954
9	95.2165	0.979	0.952	0.963
10	94.2965	0.979	0.943	0.957

La Tabla 6, muestra el desempeño del clasificador Redes Bayesianas. Es evidente la diferencia de desempeño de los clasificadores en cada clase. La poca cantidad de valores de la clase 1 entrena de manera deficiente al clasificador para su identificación.

Tabla 6. Desempeño en las clases del clasificador Red Bayesiana sin sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.992	0.109	0.876	0.685	0.931	0.188
6	0.993	0.099	0.853	0.725	0.918	0.174
7	0.994	0.245	0.949	0.745	0.971	0.368
8	0.995	0.224	0.941	0.770	0.967	0.348
9	0.995	0.279	0.956	0.765	0.975	0.409
10	0.995	0.248	0.946	0.805	0.970	0.379

La Tabla 7, muestra los resultados del entrenamiento y su validación para el algoritmo de Red Bayesiana con los conjuntos de datos donde hay sobre muestreo.

Tabla 7. Desempeño del clasificador Red Bayesiana con preparaciones con sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	85.0267	0.891	0.850	0.861
6	84.0125	0.890	0.840	0.852
7	93.8167	0.943	0.938	0.940
8	93.6477	0.943	0.936	0.938
9	94.4306	0.948	0.944	0.945
10	94.0658	0.946	0.941	0.942

La Tabla 8, presenta el desempeño del clasificador Red Bayesiana para las clases con los conjuntos de datos donde hay sobre muestreo.

Tabla 8. Desempeño en las clases del clasificador Red Bayesiana con sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.967	0.577	0.842	0.883	0.900	0.698
6	0.971	0.557	0.826	0.899	0.893	0.688
7	0.978	0.799	0.944	0.913	0.961	0.853
8	0.980	0.790	0.941	0.920	0.960	0.850
9	0.980	0.819	0.951	0.919	0.965	0.866
10	0.981	0.801	0.944	0.926	0.962	0.859

El clasificador de la red bayesiana muestra leve aumento de desempeño conforme incrementa el número de atributos. Poniendo atención a como se clasifican las clases, vemos que el sobre muestreo tiene una influencia positiva para clasificar la clase 1. El entrenamiento con un mejor desempeño, sin tener una diferencia importante con el que muestra el segundo mejor es el uso de 9 atributos. Logra clasificar correctamente 10573 instancia de un total de 11240. Veamos ahora el desempeño de los otros tipos de clasificadores diferentes.

La Tabla 9, contiene los resultados del entrenamiento y validación del perceptrón multicapas con los conjuntos de datos donde no hay sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 9. Desempeño del clasificador perceptrón multicapas con preparaciones sin sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	98.7554	0.987	0.988	0.987
6	98.8853	0.988	0.989	0.988
7	98.7879	0.987	0.988	0.987
8	98.7879	0.987	0.988	0.987
9	98.8312	0.988	0.988	0.988
10	98.8095	0.987	0.988	0.987

La Tabla 10, presenta los valores resultantes de Precisión, Exhaustividad y Medida F para clases 0 y 1 del preceptrón multicapas con los conjuntos de datos sin sobre muestreo.

Tabla 10. Desempeño en las clases del clasificador Perceptrón multicapas sin sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.992	0.743	0.995	0.650	0.994	0.693
6	0.992	0.794	0.996	0.655	0.994	0.718

7	0.992	0.759	0.995	0.645	0.994	0.697
8	0.992	0.768	0.996	0.630	0.994	0.994
9	0.992	0.777	0.996	0.645	0.994	0.705
10	0.992	0.788	0.996	0.615	0.994	0.691

La Tabla 11, contiene los resultados del entrenamiento y validación del perceptrón multicapas con los conjuntos de datos donde hay sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 11. Desempeño del clasificador perceptrón multicapas con preparaciones con sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	97.4911	0.991	0.978	0.984
6	97.6957	0.992	0.980	0.986
7	97.6512	0.990	0.980	0.985
8	97.4199	0.989	0.979	0.984
9	97.5445	0.976	0.975	0.976
10	97.5356	0.976	0.975	0.976

La Tabla 12 contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del preceptrón multicapas con los conjuntos de datos con sobre muestreo.

Tabla 12. Desempeño en las clases del clasificador Perceptrón multicapas con sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.991	0.914	0.978	0.962	0.984	0.938
6	0.992	0.920	0.980	0.966	0.986	0.943
7	0.990	0.922	0.980	0.961	0.985	0.941
8	0.989	0.916	0.979	0.956	0.984	0.936
9	0.990	0.919	0.979	0.960	0.985	0.939
10	0.990	0.917	0.979	0.961	0.985	

El clasificador perceptrón multicapa no muestra un mejor desempeño con un aumento del número de atributos. También para este clasificador el sobre muestreo tiene un impacto positivo en la clasificación de la clase 1. El clasificador entrega los mejores resultados con 6 atributos. Logra clasificar correctamente 10981 instancias de un total de 11240.

La Tabla 13, contiene los resultados del entrenamiento y validación del clasificador tabla de decisión con los conjuntos de datos sin sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 13. Desempeño del clasificador Tabla de decisión con preparaciones sin sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	97.9004	0.974	0.979	0.975
6	97.9004	0.974	0.979	0.975
7	98.0736	0.978	0.981	0.979
8	98.0736	0.978	0.981	0.979
9	98.0087	0.977	0.980	0.978
10	97.987	0.977	0.980	0.978

La Tabla 14, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador Tabla de decisión con los conjuntos de datos sin sobre muestreo.

Tabla 14. Desempeño en las clases del clasificador Tabla de decisión sin sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.984	0.531	0.995	0.260	0.989	0.349
6	0.984	0.531	0.995	0.255	0.989	0.345
7	0.987	0.577	0.993	0.410	0.990	0.480
8	0.987	0.581	0.994	0.395	0.990	0.470
9	0.986	0.559	0.993	0.380	0.990	0.452
10	0.986	0.553	0.993	0.365	0.990	0.440

La Tabla 15, contiene los resultados del entrenamiento y validación del clasificador tabla de decisión con los conjuntos de datos con sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 15. Desempeño del clasificador Tabla de decisión con preparaciones con sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	96.6459	0.966	0.966	0.966
6	97.5801	0.976	0.976	0.976
7	97.9715	0.980	0.980	0.980
8	98.1584	0.982	0.982	0.982
9	97.9893	0.980	0.980	0.980
10	97.8292	0.978	0.978	0.978

La Tabla 16, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador Tabla de decisión con los conjuntos de datos con sobre muestreo.

Tabla 16. Desempeño en las clases del clasificador Tabla de decisión con sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.978	0.917	0.980	0.911	0.979	0.914
6	0.985	0.939	0.985	0.937	0.985	0.938
7	0.985	0.959	0.990	0.937	0.987	0.948
8	0.988	0.954	0.989	0.952	0.989	0.953
9	0.985	0.958	0.990	0.939	0.988	0.948
10	0.984	0.954	0.989	0.934	0.987	0.944

El clasificador tabla de decisión no muestra un mejor desempeño con un aumento del número de atributos en el primer digito decimal. También para este clasificador como en los anteriores, el sobre muestreo tiene un impacto positivo en la clasificación de la clase 1. El clasificador da los mejores resultados con 9 atributos. Logra clasificar correctamente 11014 instancias de un total de 11240.

La Tabla 17, contiene los resultados del entrenamiento y validación del clasificador Árbol de Hoeffding con los conjuntos de datos sin sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 17. Desempeño del clasificador Árbol de Hoeffding con preparaciones sin sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	98.29	0.982	0.983	0.983
6	98.2143	0.983	0.982	0.982
7	98.0519	0.981	0.981	0.981
8	98.0628	0.982	0.981	0.981
9	97.9762	0.981	0.980	0.980
10	97.987	0.981	0.980	0.981

La Tabla 18, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador Árbol de Hoeffding con los conjuntos de datos sin sobre muestreo.

Tabla 18. Desempeño en las clases del clasificador Árbol de Hoeffding sin sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.984	0.531	0.995	0.260	0.989	0.349
6	0.984	0.531	0.995	0.255	0.989	0.345
7	0.987	0.577	0.993	0.410	0.990	0.480
8	0.987	0.581	0.994	0.395	0.990	0.470
9	0.986	0.559	0.993	0.380	0.990	0.452
10	0.986	0.553	0.993	0.365	0.990	0.440

La Tabla 19, contiene los resultados del entrenamiento y validación del clasificador árbol de Hoeffding con los conjuntos de datos con sobre muestreo.

Tabla 19. Desempeño del clasificador árbol de Hoeffding con preparaciones con sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	96.379	0.965	0.964	0.964
6	97.3488	0.974	0.973	0.974
7	97.6335	0.978	0.976	0.977
8	97.3665	0.974	0.974	0.974
9	97.4822	0.976	0.975	0.975
10	97.5089	0.976	0.975	0.975

La Tabla 20, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador Árbol de Hoeffding con los conjuntos de datos con sobre muestreo.

Tabla 20. Desempeño en las clases del clasificador Árbol de Hoeffding con sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.978	0.917	0.980	0.911	0.979	0.914
6	0.985	0.939	0.985	0.937	0.985	0.938
7	0.985	0.959	0.990	0.937	0.987	0.948
8	0.988	0.954	0.989	0.952	0.989	0.953
9	0.985	0.958	0.990	0.939	0.988	0.948
10	0.984	0.954	0.989	0.934	0.987	0.944

El clasificador tabla árbol de Hoeffding no muestra un mejor desempeño con un aumento del número de atributos. Además, este clasificador como en los anteriores, el sobre muestreo tiene un impacto positivo en la clasificación de la clase 1. El clasificador da los mejores resultados con 7 atributos. Logra clasificar correctamente 11012 instancias de un total de 11240.

La Tabla 21, contiene los resultados del entrenamiento y validación del clasificador k-vecinos más cercanos con los conjuntos de datos sin sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 21. Desempeño del clasificador k-vecinos más cercanos con preparaciones sin sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	98.7987	0.988	0.988	0.988
6	98.8312	0.988	0.988	0.988
7	98.7446	0.987	0.987	0.987
8	98.7446	0.987	0.987	0.987
9	98.7446	0.987	0.987	0.987
10	98.7554	0.988	0.988	0.988

La Tabla 22, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador k-vecinos mas cercanos con los conjuntos de datos sin sobre muestreo.

Tabla 22. Desempeño en las clases del clasificador k-vecinos mas cercanos sin sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.994	0.724	0.994	0.720	0.994	0.722
6	0.994	0.740	0.994	0.710	0.994	0.724
7	0.993	0.712	0.994	0.705	0.994	0.709
8	0.994	0.710	0.994	0.710	0.994	0.710
9	0.993	0.714	0.994	0.700	0.994	0.707
10	0.994	0.707	0.993	0.725	0.994	0.716

La Tabla 23, contiene los resultados del entrenamiento y validación del clasificador k-vecinos más cercanos con los conjuntos de datos con sobre muestreo. Los valores de la tabla son las medias ponderadas de las métricas: Precisión, Exhaustividad y Medida F.

Tabla 23. Desempeño del clasificador k-vecinos más cercanos con preparaciones con sobre muestreo.

Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
5	99.1103	0.991	0.991	0.991
6	99.1726	0.992	0.992	0.992
7	99.1993	0.992	0.992	0.992
8	99.226	0.993	0.992	0.992
9	99.226	0.993	0.992	0.992
10	99.2527	0.993	0.993	0.993

La Tabla 24, contiene los valores de Precisión, Exhaustividad y Medida F para clases 0 y 1 del clasificador k-vecinos mas cercanos con los conjuntos de datos con sobre muestreo.

Tabla 24. Desempeño en las clases del clasificador k-vecinos más cercanos con sobre muestreo.

	Precisión		*Exhaustividad*		Medida F	
Atributos	Clase 0	Clase 1	Clase 0	Clase 1	Clase 0	Clase 1
5	0.999	0.960	0.990	0.996	0.994	0.978
6	0.999	0.962	0.990	0.997	0.995	0.979
7	0.999	0.963	0.991	0.998	0.995	0.980
8	1.000	0.963	0.991	0.999	0.995	0.981
9	1.000	0.962	0.990	1.000	0.995	0.981
10	1.000	0.964	0.991	1.000	0.995	0.981

El clasificador k-vecinos más cercano muestra leves mejoras del desempeño con un aumento del número de atributos. De la misma manera que los anteriores, en este clasificador, el sobre muestreo tiene un impacto positivo en la clasificación de la clase 1. El clasificador da los mejores resultados con 10 atributos. Logra clasificar correctamente 11156 instancias de un total de 11240.

La Tabla 25 muestra un resumen de los mejores resultados de cada tipo de algoritmo. En general los cinco clasificadores logran clasificar correctamente más del 94% de los patrones. Todos con valores de Precisión, Exhaustividad y Medida F mayores al 0.94. Todos ellos son conseguidos con sobre muestreo. De los resultados disponibles, vemos los máximos para el algoritmo del k-vecinos mas cercanos.

Tabla 25. Mejores resultados de los clasificadores.

Clasificador	Atributos	% Correctos	Precisión	*Exhaustividad*	Medida F
Red Bayesiana	9	94.4306	0.948	0.944	0.945
Árbol de Hoeffding	7	97.6335	0.978	0.976	0.977
Perceptrón múlticapa	6	97.6957	0.992	0.980	0.986
Tabla de decisión	8	98.1584	0.982	0.982	0.982
k-vecinos mas cercanos	**10**	**99.2527**	**0.993**	**0.993**	**0.993**

La Tabla 26 muestra la matriz de confusión para el mejor desempeño del experimento. El algoritmo k-vecinos más cercanos con 10 atributos. Este clasificador consigue clasificar correctamente 11156 instancias de un total de 11240.

Tabla 26. Matriz de confusión del clasificador k-vecinos más cercanos con 10 atributos.

	Resultado del clasificador	
Clase	0	1
0	8956	83

1	1	2200

Los algoritmos de reconocimiento de patrones son una herramienta útil en la clasificación de marco de tiempo en las señales de audio usando la función de detección basada en la energía SF. Veamos ahora como usar los marcos con inicio para crear un descriptor de alto nivel del ritmo.

4.4. El cálculo de las distancias entre los inicios

Al identificar los inicios de las notas musicales nos es posible calcular las distancias entre ellos. Las distancias que se muestran en esta sección usan los inicios marcados a mano. La Figura 12, muestra las distancias en milisegundos existentes entre los inicios en la frase musical en el estilo afro. Hay que poner atención en la irregularidad de las distancias. De acuerdo con la definición de los índices, distancias irregulares generan valores altos. Mientras distancias regulares valores bajos.

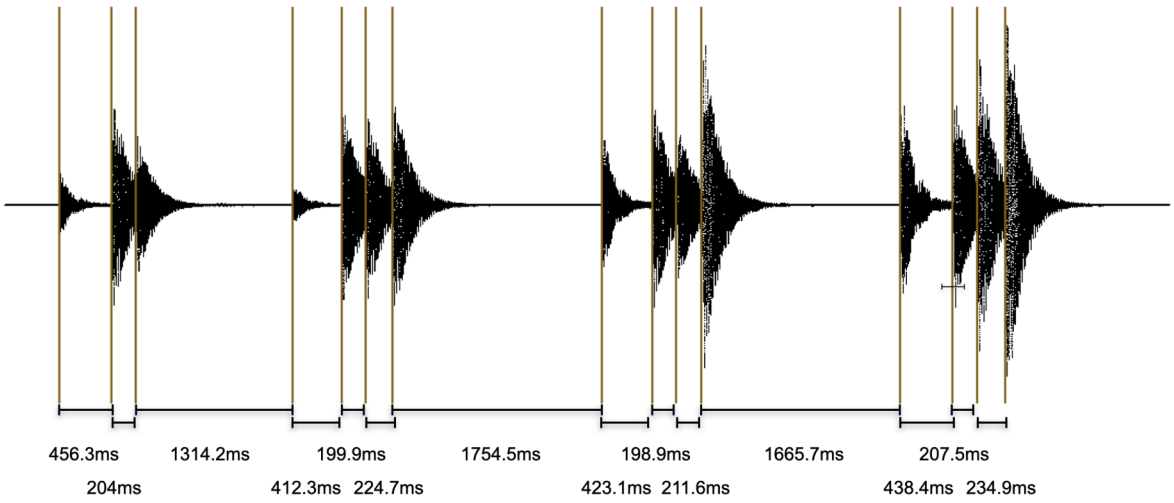

Figura 12. Distancias entre los inicios en milisegundos del archivo 048_phrase_afro_simple_slow_mallets.wav.

La Figura 13, muestra la misma frase en el estilo afro. A diferencia de la figura anterior ahora tenemos las distancias relativas a los primeros dos inicios. Esta primera distancia es usada como medida de referencia para las demás. Teniendo un conjunto de valores que permiten identificar las diferencias entre ellas. Un valor de 0.5 indica que la distancia entre los dos

inicios es la mitad de la existencia entre los primeros dos. Por otro lado, una distancia de 3 expresa, una distancia el triple que la de los primeros dos.

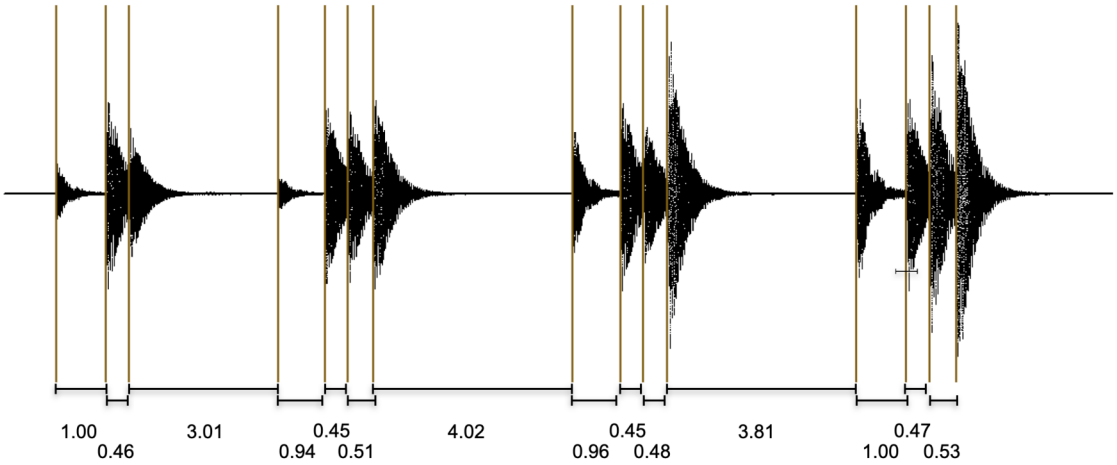

Figura 13. Distancias relativas de los inicios del archivo 048_phrase_afro_simple_slow_mallets.wav.

Las distancias calculadas para los seis archivos de los audios del experimento se encuentran en el Apéndice. A continuación, los valores de los descriptores de alto nivel del ritmo.

4.5. Los valores de los descriptores de alto nivel nPvi y rPvi

Los indices de variablidad en pares nPvi y rPvi toman las distancias relativas entre los inicios. Miden la medida de cambio entre las parejas adyacentes de distancias. La primera de ellas normaliza estas distancias. Mientras que la segunda las muestra de una manera cruda. La Tabla 27, muestra los valores de los índices de los 6 archivos del experimento. Resalta como el archivo con la frase en el estilo musical rock contiene distancias entre los inicios mas regulares. Por el otro lado, la frase del estilo afro sus distancias son mas irregulares.

Tabla 27. Descriptores nPvi y rPvi para los audios del experimento.

No.	Archivo de audio	nPvi	rPvi
1	036_phrase_disco_simple_slow_sticks.wav	35.22	0.4960
2	042_phrase_rock_simple_slow_rods.wav	9.63	0.0822
3	048_phrase_afro_simple_slow_mallets.wav	85.76	1.4973
4	060_phrase_salsa_simple_slow_sticks.wab	34.98	0.2729
5	066_phrase_shuffle-blues_simple_slow_brushes.wav	50.64	0.4044
6	078_phrase_reggae_simple_slow_sticks	57.98	0.4105

Fuera del alcance de este estudio está el análisis de estos valores. Pero algunas preguntas que se pueden plantear al ver los valores son: ¿la cantidad de inicios es una variable significativa en el valor de los índices?, ¿es posible comparar estos índices si la cantidad de inicios es diferente? y ¿varias frases de un mismo estilo, género o tradición arrojarán índices semejantes? El siguiente capítulo muestra las conclusiones de este trabajo y los trabajos futuros en el tema de los descriptores del ritmo propuestos en esta tesis.

Capitulo 5. Conclusiones y trabajos futuros

> "… de donde yo vengo, nosotros decimos: El ritmo es el alma de la vida, porque el universo gira en torno al ritmo y cuando salimos de ritmo, es cuando estamos en problemas." Babatunde Olatunji

En el actual proyecto de tesis se presentó la propuesta de generar un descriptor del ritmo accediendo directamente a los archivos de audio musical. Para demostrar la posibilidad técnica se usó un conjunto de audios con grabaciones de ejecuciones de una batería. Las ejecuciones contienen varios ritmos. Los ritmos tienen una cantidad diferente de golpes. La estrategia general para conseguirlo fue: tomar la información de la forma de onda de los audios, marcar los inicios, segmentar la señal de audio en ventanas de tiempo, crear una función de detección con la información espectral de la ventana, entrenar un clasificador, identificar los segmentos con inicios, a partir de ellos calcular las distancias entre los mismos, generar una medida de distancias entre los inicios relativa a los primeros dos para resaltar la diferencia entre las mismas y estas distancias relativas se usan para generar los descriptores de alto nivel del ritmo, los índices nPvi y rPvi.

La tarea de marcar los inicios en los archivos de audio consume mucho tiempo. Una herramienta útil para realizarla es aquella que permita visualizar el audio en su forma de señales de onda y un espectrograma. El identificar el momento en el tiempo de los inicios involucra el hacer acercamientos y alejamientos en diferentes niveles de la forma de onda y el espectrograma. Los programas que automáticamente identifiquen los inicios ahorran tiempo en esta tarea.

La función de detección basada en la energía en los marcos de tiempo, flujo espectral, en este trabajo mostró resultados favorables para los instrumentos de percusión que forman una batería. Con otros instrumentos percusivos y no percusivos habría que probar otras funciones de detección, en la literatura encontramos opciones como las basadas en la fase o las que combinan tanto la energía y la fase. La amplia diversidad de instrumentos hace está tarea compleja.

El reconocimiento de inicios a través de algún algoritmo de aprendizaje máquina es una tarea viable. Se demostró en el presente trabajo que varios de ellos pueden realizarla. Trabajo para otros ingenieros o investigadores interesados en el tema está disponible. Investigar el desempeño de los clasificadores con otras funciones de detección o generar una propuesta que sea capaz de cubrir el problema general para instrumentos percusivos y no percusivos sería deseable. Como lo intenta Stasiak [9] al combinar varias funciones de detección en una NNA.

La precisión del cálculo en las distancias entre los inicios es diferente a hacerlo por una persona que por una máquina. La interrogante de si esta precisión afecta los valores de los descriptores de alto nivel de forma significativa no se cubrió en el presente trabajo. Los valores de los descriptores nPvi y rPvi creados por programas informáticos serán diferentes de los calculados por personas, es un tema para analizar.

En la literatura podemos ver que el descriptor nPvi es capaz de diferenciar valores para diferentes ritmos musicales, inclusive en varias tradiciones musicales y culturas. Habría que trabajar en extender esta idea. Revisar si un estilo, género o tradición musical contiene diferentes ritmos y si estos ritmos generan valores diferentes o similares a ritmos en otro estilo, género o tradiciones. Tener claro cuáles son las condiciones en las que el descriptor permite hacer comparaciones debe ser trabajada. La cantidad de inicios en las frases a comparar parece ser algo significativo, en la literatura solo se comparan frases con cantidad de inicios iguales. En cuanto si es posible tener números diferentes en la cantidad de inicios entre las frases a comparar y que permita identificar una relación entre ellos, es otro tema para considerar.

Un sistema de recuperación de piezas musicales basado en el ritmo es posible. En este trabajo se demostró que generar un descriptor de alto nivel del ritmo es viable técnicamente. Sin embargo, varias interrogantes están pendientes de resolver. Por ejemplo, ¿son únicamente los descriptores nPvi o rPvi necesarios para comparar los ritmos en un universo mayor de ritmos, géneros, estilos o tradiciones musicales? o ¿es necesario acompañar estos descriptores con otros elementos de datos que den contexto a los valores?, como pudieran

ser los nombres de los géneros, estilos o tradiciones musicales para realizar búsquedas en subconjuntos de piezas musicales.

En un bosquejo general de los elementos de un MIR basado en el ritmo debería tener al menos seis componentes. Un primer componente para agregar piezas musicales; con algunos mecanismos para conseguir: la frase o frases más representativas del ritmo de la pieza, la cantidad de inicios que forman dicha frase, la identificación de estos inicios en el tiempo y algunos metadatos como el género, estilo o tradición musical. Un segundo componente para calcular los índices nPvi y rPvi, junto a varias herramientas estadísticas para analizar los valores del descriptor. Un tercer componente, una base de datos que almacene sistemáticamente los descriptores y los metadatos. Un cuarto componente para agrupar los valores de los descriptores e identificar piezas con ritmos similares. Un quinto para buscar un valor específico o un rango de valores para una búsqueda de un ritmo y las piezas correspondientes. Un sexto, un conjunto de interfaces para diversos tipos de usuarios. Usuarios expertos que permitan agregar piezas con su información técnica y usuarios que consuman piezas musicales introduciendo el descriptor del ritmo de alguna manera, uno posible es el tarareo del ritmo.

Referencias

[1] P. Gronow y I. Saunio, International history of the recording industry, London: A&C Black, 1999.

[2] M. A. Casey, V. Remco, M. Goto, M. Leman, C. Rhodes y M. Slaney, "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE,* vol. 96, nº 4, pp. 668--696, 2008.

[3] W. A. Sethares y D. Bañuelos, Rhythm and transforms, vol. 1, New York: Springer, 2007.

[4] G. T. Toussaint, "The pairwise variability index as a tool in musical rhythm analysis," de *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Science of Music*, Thessalonika, Greece, 2012.

[5] E. Grabe y E. L. Low, "Durational variability in speech and the rhythm class hypothesis," de *Papers in Laboratory Phonology*, Berlin, 2002.

[6] J. R. Daniele y A. D. Patel, "The interplay of linguistic and historical influences on musical rhythm in different cultures," de *Proceedings of the 8th International Conference on Music Perception and Cognition*, Evanston, Illinois, 2004.

[7] A. D. Patel y J. R. Daniele, "An empirical comparison of rhythm in language and music," *Cognition,* vol. 87, pp. B35--B45, 2003.

[8] M. Raju, E. L. Asu y J. Ross, "Comparison of rhythm in musical scores and performances as measured with the pairwise variability index," *Musicae scientiae,* vol. 14, nº 1, pp. 51--71, 2010.

[9] B. Stasiak, J. Mońko y A. Niewiadomski, "Note onet detection in musical signals via neural-network-based multi-ODF fusion," *International Journal of Applied Mathematics and Computer Science,* vol. 26, nº 1, pp. 203-213, 2016.

[10] P. M. Kumar, J. Sebastian y H. A. Murthy, "Musical onset detection on carnatic percussion instruments," de *2015 Twenty First National Conference on Communications (NCC)*, Mumbai, India, 2015.

[11] D. Wulandari, A. Tjahyanto y Y. S. Suprapto, "Gamelan music onset detection based on spectral features," *TELKOMNIKA (Telecommunication Computing Electronics and Control),* vol. 11, nº 1, pp. 07--118, 2013.

[12] M. Tian, A. Srinivasamurthy, M. Sandler y X. Serra, "A study of instrument-wise onset detection in Beijing opera percussion ensembles," de *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Florence, Italy, 2014.

[13] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies y M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on speech and audio processing,* vol. 13, nº 5, pp. 1035--1047, 2005.

[14] X. Rodet y F. Jaillet, "Detection and modeling of fast attack transients," de *Proc. Int. Computer Music Conf.*, 2001.

[15] J. P. Bello y M. Sandler, "Phase-based note onset detection for music signals," de *003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (IEEE Cat. No.03TH8684)*, New Paltz, NY, USA, 2003.1285811.

[16] C. Duxbury, J. P. Bello, M. Davies, M. Sandler y others, "Complex domain onset detection for musical signals," de *Proc. 6th Conf. Digital Audio Effect (DAFx-03)*, London, UK, 2003.

[17] G. Costantini, R. Perfetti y M. Todisco, "Event based transcription system for polyphonic piano music," *Signal processing,* vol. 89, nº 9, pp. 1798--1811, 2009.

[18] A. Lacoste y D. Eck, "A supervised classification algorithm for note onset detection," *EURASIP J. Adv. Signal Process.,* vol. 2007, nº 1, pp. 1--13, 2006.

[19] E. Marchi, G. Ferroni, F. Eyben, L. Gabrielli, S. Squartini y B. Schuller, "Multi-resolution linear prediction based features for audio onset detection with bidirectional LSTM neural networks," de *2014 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Florence, Italy, 2014.

[20] J. Schlüter y S. Böck, "Improved musical onset detection with convolutional neural networks," de *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, 2014.

[21] P. Leveau y L. Daudet, "Methodology and tools for the evaluation of automatic onset detection algorithms in music," de *Proc. ISMIR*, Barcelona, Spain, 2004.

[22] G. Loy, Musimathics, Volume 2: The Mathematical Foundations of Music, vol. 2, MA, USA: The MIT Press, 2007.

[23] S. Dixon, "Onset detection revisited," de *in Proc. of the 9th Int. Conf. on Digital Audio Effects (DAFx)*, Montreal, Quebec, Canada, 2006.

[24] L. I. Kuncheva, Combining pattern classifiers: methods and algorithms, Hoboken, NJ, USA: Wiley, 2014.

[25] N. Friedman, D. Geiger y M. Goldszmidt, "Bayesian Network Classifiers," *Mach. Learn.,* vol. 29, nº 2, pp. 131-163, 1997.

[26] M. Horný, "Bayesian networks," *Boston University School of Public Health,* vol. 17, 2014.

[27] A. K. Jain, J. Mao y M. K. Mohiuddin, "Artificial neural networks: A tutorial," *Computer,* vol. 29, nº 3, pp. 31--44, 1996.

[28] R. Kohavi, "The Power of Decision Tables," de *8th European Conference on Machine Learning*, NY, 1995.

[29] B. G. Becker, "Visualizing Decision Table Classifiers," de *Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258)*, CA, USA, 1998.

[30] P. M. Domingos y G. Hulten, "Mining high-speed data streams," de *Proc. 6th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Boston, MA, USA, 2000.

[31] K. Taunk, S. De, S. Verma y A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," de *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, India, 2019.

[32] J. McDonough, H. Danko y J. Zentz, "Rhythmic structure of music and language: An empirical investigation of the speech cadence of American jazz masters Louis Armstrong and Jelly Roll Morton," *University of Rochester Working Papers in the Language Sciences,* vol. 3, nº 1, pp. 45--56, 2007.

[33] J. R. Daniele y A. D. Patel, "Stability and change in rhythmic patterning across a composer's lifetime: a study of four famous composers using the nPVI equation," *Music Perception,* vol. 33, pp. 255--265, 2015.

[34] O. Gillet y G. Richard, "Enst-drums: an extensive audio-visual database for drum signals processing," de *International Society for Music Information Retrieval Conference (ISMIR)*, Victoria, Canada, 2006.

[35] V. Ganganwar, "An overview of classification algorithms for imbalanced datasets," *Int. J. Emerging Technol. Adv. Eng,* vol. 2, nº 4, pp. 42--47, 2012.

[36] E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, I. H. Witten y L. Trigg, "Weka-a machine learning workbench for data mining," de *Data mining and knowledge discovery handbook*, MA, USA, 2009.

[37] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *ArXiv,* vol. abs/2010.16061, 2020.

[38] S. J. Downie, "Music information retrieval," *Annual review of information science and technology,* vol. 37, nº 1, pp. 295--340, 2003.

[39] L. Kuncheva, Fuzzy classifier design, vol. 49, Springer-Verlag, 2000.

Apéndice A. Distancias entre los inicios.

El archivo 036_phrase_disco_simple_slow_sticks contiene una frase en el estilo disco. En la frase se marcaron 41 inicios. El conjunto D_1 contiene por elementos las distancias relativas entre ellos. $D_1 = \{1.0000,\ 0.8103,\ 1.8492,\ 1.8365,\ 2.0511,\ 0.8514,\ 0.9042,\ 2.0333,\ 1.8249,\ 2.0504, 0.8817,\ 0.9873,\ 2.0000,\ 1.8419,\ 1.9569,\ 0.9569,\ 0.8589,\ 2.1235,\ 1.8460,\ 1.9370,\ 0.9269,\ 0.9407,\ 2.0681,\ 1.8832,\ 2.0430,\ 0.9375,\ 0.9946,\ 2.0582,\ 1.7638,\ 1.9713,\ 0.9722,\ 0.9182,\ 2.0770,\ 1.8288,\ 1.9022,\ 0.9255,\ 0.9264,\ 1.8676,\ 1.9726, 0.9667\}$, con una cardinalidad $|D_1| = 40$.

El archivo 042_phrase_rock_simple_slow_rods.wav es la grabación de una frase en el estilo musical rock. La cantidad de inicios marcada es de 33. El conjunto D_2 muestra las distancias entre ellos. $D_2 = \{1.0000,\ 1.0673,\ 0.9327,\ 1.0746,\ 1.0269,\ 1.0894,\ 0.9698,\ 0.5615, 0.4537,\ 0.9982,\ 1.0347,\ 0.9577,\ 1.0136,\ 1.0136,\ 0.9926,\ 1.0148,\ 0.9924,\ 1.0772,\ 1.0446,\ 1.0469,\ 0.9430,\ 1.0680,\ 1.0601,\ 0.9965,\ 0.9869,\ 1.0994,\ 1.0446,\ 1.0129,\ 1.0066,\ 0.9890,\ 1.0146,\ 0.9823\}$, con una cardinalidad $|D_2| = 32$.

En el archivo 048_phrase_afro_simple_slow_mallets.wav está una frase en el estilo afro. Este archivo tiene 15 inicios. El conjunto D_3 tiene los inicios relativos de la frase. $D_3 = \{1.0000, 0.4675,\ 3.0124,\ 0.9451,\ 0.4583,\ 0.5149,\ 4.0214,\ 0.9699,\ 0.4558,\ 0.4849,\ 3.8179,\ 1.0049,\ 0.4755,\ 0.5383\}$, con una cardinalidad $|D_3| = 14$.

El archivo de audio 060_phrase_salsa_simple_slow_sticks.wav tiene una frase en el estilo musical salsa. La frase está formada por 45 golpes. Las distancias entre los inicios de los golpes se muestran en el conjunto D_4. $D_4 = \{1.0000,\ 1.0196,\ 1.0402,\ 0.9921,\ 0.9870,\ 2.0698,\ 0.4989,\ 0.5143,\ 1.0467,\ 0.4995,\ 0.5085,\ 1.0451,\ 0.4876,\ 0.5273,\ 1.0480,\ 0.4995,\ 0.5352,\ 1.0445,\ 0.4904,\ 0.5338,\ 1.0136,\ 1.0490,\ 0.9822,\ 0.4846,\ 0.5246,\ 0.9945,\ 0.4408,\ 0.5661,\ 0.4680,\ 0.4931,\ 0.4889,\ 0.4847,\ 0.9873,\ 0.4590,\ 0.5434,\ 0.4859,\ 0.4687,\ 0.4368,\ 0.4986,\ 0.9759,\ 0.4359,\ 0.4976,\ 0.482,\ 0.5174,\ 0.9508\}$, con una cardinalidad $|D_4| = 45$.

El audio 066_phrase_shuffle-blues_simple_slow_brushes.wav es una frase en el estilo musical blues. La frase contiene 35 inicios. Las distancias relativas entre los golpes se muestran en el conjunto D_5. $D_5 = \{1.0000,\ 0.5701,\ 1.0704,\ 0.5410,\ 1.1246,\ 0.5579,\ 1.1077,\ 0.5810,\ 1.1764,\ 0.5548,\ 1.1096,\ 0.5635,\ 1.0991,\ 0.5495,\ 0.9720,\ 0.5693,\ 1.0440,\ 0.5647,\ 1.0152,\ 0.5554,\ 1.0864,\ 0.5660,\ 1.0424,\ 0.4811,\ 0.4885,\ 0.5435,\ 0.4631,\ 0.4920,\ 0.4779,\ 0.5012,\ 0.9790,\ 0.5640,\ 0.9384,\ 0.8593\}$, con una cardinalidad $|D_5| = 34$.

El archivo de audio 078_phrase_reggae_simple_slow_sticks.wav es una grabación de una frase en el estilo musical reggae. La frase rítmica está formada por 31 golpes, contiene la misma cantidad de inicios. El conjunto D_6 contiene las distancias relativas entre los golpes. $D_6 = \{1.0000,\ 0.6047,\ 0.0039,\ 0.9574,\ 0.6420,\ 1.0174,\ 0.6163,\ 0.9826,\ 0.5698,\ 1.5002,\ 1.0055,\ 0.5759,\ 0.9972,\ 0.6811,\ 0.9413,\ 0.6700,\ 1.0346,\ 0.5423,\ 1.0249,\ 0.5607,\ 1.0642,\ 0.6678,\ 0.9668,\ 0.5781,\ 0.9766,\ 0.6176,\ 0.8767,\ 0.6522,\ 0.9788,\ 1.5062\}$, con una cardinalidad $|D_6| = 30$.

Apéndice B. Parámetros de los algoritmos de aprendizaje en Weka.

Los parámetros por omisión configurados en Weka para el algoritmo Red Bayesiana son: Un tamaño de lote 100, un estimador simple de las tablas de probabilidades con un alpha de 0.5, numero de decimales 2, el algoritmo de aprendizaje “Hill climbing” y sin el uso de árbol de AD para acelerar los conteos.

Los parámetros por omisión configurados en Weka para el algoritmo Perceptrón multicapa son: Un tamaño de lote de 100, un capa oculta con un número de neuronas = (número de atributos + clases) /2, funciones de activación sigmoid en la capa oculta y linealpara la de salida, una taza de aprendizaje = 0.3, un momemtum de 0.2, con un filtro de nominal a binario, los atributos normalizados, las clases normalizadas, posiciones decimales 2, con un valor de 20 valores en el error que se incrementa antes de detener el aprendizaje. El algoritmo de aprendizaje usado backpropagation.

Los parámetros por omisión configurados en Weka para el algoritmo Tabla de decisión son: Un tamaño de lote de 100, un número de folds = 1, una medida de evaluación de precisión de clase discreta y una clase numérica, posiciones decimales 2, un espacio de búsqueda con “greedy hill climbing augmented” con facilidades de “backtracking”, dirección hacia adelante, y terminación de búsqueda de 5.

Los parámetros por omisión configurados en Weka para el algoritmo Árbol de Hoeffding son: Un tamaño de lote de 100, un periodo de gracia de 200, un umbral usado para romper “ties” = 0.05, una estrategia de predicción “Naive Bayes adaptive”, una fracción mínima de peso ganada de 0.01, un umbral de predicción de “Naive Bayes” 0, un criterio de separación “info gain Split” y un error permitido en el criterio de decisión de 1.0 E-7

Los parámetros por omisión configurados en Weka para el algoritmo k-vecinos más cercanos son: k=1, tamaño del lote = 100, sin las distancias ponderadas, algoritmo de búsqueda lineal, numero de posiciones decimales 2 y un tamaño de la ventana 0. Una semilla de 0 para inicializar el generador de números aleatorios con los que se configuran los valores iniciales de los pesos. Un porcentaje de 0 para el conjunto de validación.