



IDENTIFICACIÓN ESTRUCTURAL AUTOMÁTICA DE FRAGMENTOS MUSICALES

Carlos Andrés Torres Rodríguez

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2015

IDENTIFICACIÓN ESTRUCTURAL AUTOMÁTICA DE FRAGMENTOS MUSICALES

Carlos Andrés Torres Rodríguez

Trabajo de investigación presentado como requisito parcial para optar al título de:
Magíster en Ingeniería
Ingeniería de Sistemas y Computación

Director:
Ingeniero Luís Fernando Niño Vásquez
Ph. D. Computer Science

Asesor:
Maestro Horacio Alberto Lapidus
Pedagogo Musical
M.Ed. Tecnologías de la Información aplicadas a la Educación

Línea de investigación:
Sistemas Inteligentes
Grupo de Investigación: LISI

Universidad Nacional de Colombia
Facultad de Ingeniería
Departamento de Ingeniería de Sistemas e Industrial
Bogotá, Colombia
2015

A mi madre, mi padre y hermanos

AGRADECIMIENTOS

Este trabajo de investigación realizado en la Universidad Nacional de Colombia es un esfuerzo en el cual aportaron directa o indirectamente distintas personas opinando, evaluando, corrigiendo, escuchando, etc.

En primer lugar, expreso mis agradecimientos al grupo de investigación BIOLISI, por adoptarme, por las reuniones de los miércoles y sábados, su gran colaboración y la oportunidad brindada de trabajar en conjunto.

Agradezco también a Luís Fernando Niño Vásquez, director de la investigación, por su orientación y el tiempo prestado para la realización.

A Carlos Manuel Estévez-Breton, estudiante de Ph.D., por su colaboración en los trabajos comunes realizados para este proyecto.

A Horacio Lapidus, profesor de Educación Musical y M.Ed., Tecnologías de la Información Aplicadas a la Educación, por su tiempo en el acompañamiento de este proyecto desde su experiencia musical.

A Catherine M. y Felipe V. por sus aportes y colaboración con el diseño y estilo del documento.

A Ivi por su compañía, colaboración y paciencia.

A los músicos integrantes del grupo de validación de resultados del algoritmo por su disposición y tiempo.

A Lona y Yitas por acompañarme en la escritura del documento.

Agradecimientos finales a todos aquellos que de una u otra forma colaboraron para que se llevara a cabo este trabajo.

RESUMEN

La identificación estructural automática es una de las líneas de investigación que forma parte de la Tecnología Musical, en la cual se aprovechan las técnicas ofrecidas por la Inteligencia Artificial (en especial el Aprendizaje Computacional) para simular el análisis de un músico experimentado en la identificación y descripción del diseño y la forma de una pieza musical.

Los métodos actuales reportados en la literatura, ofrecen soluciones variadas en cuanto a la segmentación de piezas musicales y a la relación entre las divisiones propuestas, pero describen aún de forma muy simple y general las características sobresalientes, por tal razón se propuso una nueva forma de abordar el problema.

Se desarrolló una metodología para la identificación estructural automática de piezas musicales que consta de: un proceso de fragmentación, una caracterización de los fragmentos contenidos, y una estrategia de búsqueda en profundidad de la mejor fragmentación posible.

En el proceso de fragmentación se propusieron nuevas heurísticas para reducir el espacio de búsqueda y así recorrer las posibles fragmentaciones en un tiempo razonable. Dentro de las heurísticas se encuentra un novedoso algoritmo de particionamiento en frases fundamentado en conceptos de percepción de pausas rítmicas.

Para la representación computacional de los fragmentos se propusieron nuevas características basadas en conceptos teóricos de forma y estructura.

La identificación de los elementos estructurales y los rasgos característicos se realizó por medio del análisis de relaciones de similitud desde una aproximación de vistas múltiples por medio de técnicas de agrupamiento no supervisado.

El algoritmo propuesto entrega como respuesta los rasgos característicos de la pieza en cuanto al agrupamiento de fragmentos similares y la discriminación de fragmentos diferenciales en cada una de las características exploradas, presentando de esta manera una descripción gráfica y numérica del diseño formal y permitiendo inferir con argumentos la estructura subyacente en la pieza musical.

Palabras claves: Tecnología Musical, Inteligencia Artificial, Aprendizaje Computacional, Algoritmo de Búsqueda, Agrupamiento, Aprendizaje no Supervisado, Forma Musical.

ABSTRACT

Automatic Extraction of Musical Structure is one of the research areas in Music Technology, in which Artificial Intelligence offered techniques (especially the Machine Learning) used to simulate the analysis of an experienced musician in the identification and description of a musical piece design and form.

Current methods reported in the literature offer a variety of solutions in terms of musical piece segmentation and the relationship between the proposed partitions, and still describe in a simple and general way the outstanding features, for this reason, a new approach is proposed.

A methodology for automatic extraction of musical structure has been developed, which consist of: a fragmentation process, a content fragments characterization, and a deep search strategy for the best possible fragmentation.

In the fragmentation process new heuristics are proposed in order to reduce the search space and thus explore the possible fragmentations within a reasonable time. A novel phrase partitioning algorithm is among the proposed heuristics, which is based on rhythmic breaks perception concepts.

For the computational representation of the fragments, new features based on theoretical concepts of form and structure were proposed.

Identification of structural elements and distinctive features was performed through similarity analysis from a multi-view approaching, using unsupervised clustering techniques.

The proposed algorithm delivers as response the piece distinctive traits regarding the clustering of similar fragments and discrimination of differential fragments in each of the scanned features, thus showing a graphical and numerical description for the formal design and allowing arguments to infer the underlying structure of a musical piece.

Key words: Music Technology, Artificial Intelligence, Machine Learning, Search Algorithm, Clustering, Unsupervised Learning, Musical Form.

LISTA DE FIGURAS

<i>Figura 1.</i> Idea musical completa dividida en motivos y variaciones.	28
<i>Figura 2.</i> Ejemplo ilustrativo de análisis de patrones en texto.	33
<i>Figura 3.</i> Metodología general por fases.	35
<i>Figura 4.</i> Esquema de la fase de análisis.	35
<i>Figura 5.</i> Esquema de la fase de diseño.	36
<i>Figura 6.</i> Esquema del “proceso de identificación estructural”	37
<i>Figura 7.</i> Ejemplo de jerarquía musical.	38
<i>Figura 8.</i> Esquema de la fase de implementación.	40
<i>Figura 9.</i> Esquema de la fase de pruebas.	40
<i>Figura 10.</i> Histograma de eventos en melodías ERK.	41
<i>Figura 11.</i> Ejemplo de melodía ERK etiquetada.	42
<i>Figura 12.</i> Esquema de la fase de evaluación.	42
<i>Figura 13.</i> Ejemplo de todas las posibles fragmentaciones para una pieza corta.	45
<i>Figura 14.</i> Ejemplo gráfico de reducción de espacio de búsqueda.	46
<i>Figura 15.</i> Ejemplo para corte aceptado y rechazado por limitación de cortes.	46
<i>Figura 16.</i> Histograma y frecuencia de duraciones para una melodía de ejemplo.	47
<i>Figura 17.</i> Ejemplo de omisión de fragmentos repetidos.	48
<i>Figura 18.</i> Representación numérica para la pieza de ejemplo.	49
<i>Figura 19.</i> Representación computacional de un fragmento ejemplo.	49
<i>Figura 20.</i> Ejemplos de líneas rítmicas.	51
<i>Figura 21.</i> Ejemplos de distribuciones de alturas.	51
<i>Figura 22.</i> Ejemplos de líneas melódicas.	51
<i>Figura 23.</i> Características perspectivas para el fragmento ejemplo.	52
<i>Figura 24.</i> Esquema del algoritmo general.	60
<i>Figura 25.</i> Esquema del algoritmo de identificación estructural.	60
<i>Figura 26.</i> Esquema del algoritmo de identificación estructural en profundidad.	61
<i>Figura 27.</i> Esquema del algoritmo de selección de mejor fragmentación.	63
<i>Figura 28.</i> Canción <i>Was Hoert Man Denn Neuess Vom Kriegen</i>	65
<i>Figura 29.</i> Fragmentación mayor del caso ejemplo.	67
<i>Figura 30.</i> Visor jerárquico para el tiempo total.	68
<i>Figura 31.</i> Visor jerárquico para la proporción sonido/silencio.	68
<i>Figura 32.</i> Visor MSD y jerárquico para la densidad rítmica.	69
<i>Figura 33.</i> Visor MDS y jerárquico para el Índice potencial de tensión.	69
<i>Figura 34.</i> Visor MDS y jerárquico para la línea rítmica.	70
<i>Figura 35.</i> Señalamiento sobre la partitura del primer grupo identificado.	70
<i>Figura 36.</i> Señalamiento sobre la partitura del segundo grupo identificado.	71
<i>Figura 37.</i> Señalamiento sobre la partitura del tercer grupo identificado.	71
<i>Figura 38.</i> Visor MDS y jerárquico para la línea melódica.	72
<i>Figura 39.</i> Visor MDS y jerárquico para la distribución de alturas.	72
<i>Figura 40.</i> Visor MDS y jerárquico para el Kernel lineal.	73

<i>Figura 41.</i> Visor MDS y jerárquico para la parametrización adaptativa.	74
<i>Figura 42.</i> Pieza musical simplificada.	75
<i>Figura 43.</i> Fragmentación en profundidad del caso ejemplo.	75
<i>Figura 44.</i> Visor MDS y jerárquico para el tiempo total en profundidad.	76
<i>Figura 45.</i> Visor jerárquico para la Proporción sonido/silencio en profundidad.	76
<i>Figura 46.</i> Visor MSD y jerárquico para la densidad rítmica en profundidad.	77
<i>Figura 47.</i> Visor MDS y jerárquico para el índice potencial de tensión en profundidad.	77
<i>Figura 48.</i> Visor MDS y jerárquico para la línea rítmica en profundidad.	78
<i>Figura 49.</i> Visor MDS y jerárquico para la línea melódica en profundidad.	78
<i>Figura 50.</i> Visor MDS y jerárquico para la distribución de alturas en profundidad.	79
<i>Figura 51.</i> Visor MDS y jerárquico para el Kernel lineal en profundidad.	79
<i>Figura 52.</i> Visor jerárquico para la parametrización adaptativa en profundidad.	80
<i>Figura 53.</i> Análisis de sensibilidad para el parámetro <i>Umbral de corte (fdi)</i>	82
<i>Figura 54.</i> Histograma para la media armónica.	83
<i>Figura 55.</i> Ejemplo de enunciado de evaluación.	84
<i>Figura 56.</i> Resultados de la evaluación manual.	85

LISTA DE TABLAS

<i>Tabla 1.</i> Comparación de resultados para modelos de identificación de <i>frases</i>	31
<i>Tabla 2.</i> Características iniciales del fragmento ejemplo.	50
<i>Tabla 3.</i> Caracterización completa del fragmento ejemplo.	53
<i>Tabla 4.</i> Parámetros de configuración de la prueba.	66

GLOSARIO

Glosario A - para la terminología musical

Armonía: Percepción del sonido en forma vertical o simultánea en forma de acordes.

Estructura (arquitectura): es la disposición y orden de las *partes* o ideas musicales dentro de un todo. También puede entenderse como un sistema de conceptos coherentes enlazados, como ejemplos podemos citar las estructuras típicas ABA, AB, AA'A.

- Primaria: constituida por una sola frase independiente de la introducción o coda.
- Binaria: formada por dos secciones A y B.
- Ternaria: formada por tres secciones A, B y C.
- Rondó: ABACA.

Grado: posición de cada nota dentro de una escala musical. Los grados se designan mediante números romanos correlativos I, II, III, IV, V, VI y VII.

Pulso: es una unidad básica que se emplea para medir el tiempo, dividiéndolo en partes iguales o intervalos regulares.

Ritmo: relaciones de duración entre los sonidos de una pieza. Puede ser regular o irregular, según sean aplicables o no patrones métricos constantes tales como pulsos, compases, etc.

Textura musical: es la forma en que los materiales melódicos, rítmicos y armónicos se combinan en una composición, determinando así la cualidad sonora global de una pieza. Designa la forma de relacionarse las diversas voces que intervienen en una pieza musical (entendiendo como voces diversas líneas melódicas simultáneas, sean vocales o instrumentales).

Tonalidad: se refiere a música de origen europeo desde el siglo XVI al XIX, en que las creaciones usan escalas diatónicas y la selección y uso de las de notas musicales obedece a un ordenamiento jerárquico basado en una nota principal (y su acorde / escala) hacia la cual tienden a resolverse las demás: la tónica.

Glosario B - Para la terminología de computación

Explosión combinatoria: para los algoritmos que emplean búsqueda por fuerza bruta, en la mayoría de casos reales, el número de soluciones candidatas es prohibitivamente alto, dado que el tamaño del espacio de búsqueda se encuentra determinado por la combinación de las variables en análisis. La explosión combinatoria afecta directamente el tiempo de operación de un algoritmo llevándolo en algunos casos a valores inaceptables de horas, días o años.

Fuerza bruta: también conocida como búsqueda combinatoria o exhaustiva, es una técnica computacional que consiste en recorrer todas las posibles combinaciones para la solución de un problema.

Heurística: estrategia, método o criterio utilizado para hacer más sencilla la solución a un problema o para reducir su tiempo de búsqueda, algunas veces haciendo uso de conocimiento previo o experiencia en el campo de acción.

K-Means: método de agrupamiento ampliamente utilizado en minería de datos.

Modelos ocultos de Markov: modelo estadístico utilizado en el reconocimiento de patrones. Su objetivo es inferir unos parámetros desconocidos a partir de unas observaciones.

LISTADO DE TÉRMINOS

AI: Artificial Intelligence, en español conocida como IA o Inteligencia Artificial, es el estudio del diseño de agentes inteligentes.

EFSC: Essen Folk Song Collection, Colección de canciones folclóricas de Alemania y países cercanos, ensamblada en el *Gesamthochschule* de la Universidad Essen.

ERK: Colección de canciones folclóricas Alemanas bajo el título “*La selección de las más exquisitas canciones populares*” (Deutscher Liederhort, Auswahl der vorzüglicheren Deutschen Volkslieder) originalmente recolectadas por *Ludwing Erk*.

GPR: Grouping Preference Rules, Agrupación por Reglas de Preferencia, es un algoritmo para identificación de frases musicales.

ICA: Independent Component Analysis, Análisis de Componentes Independientes, es un método computacional que sirve para separar una señal multivariante en sus componentes aditivos suponiendo que la señal de origen tiene independencia estadística.

IDyOM: Information Dynamics Of Music, Información Musical Dinámica, es un algoritmo para identificación de frases musicales.

LBDM: Local Boundary Detection Model, Modelo de Detección de Límites Locales, es un algoritmo para identificación de frases musicales.

MDS: Multidimensional Scaling, escalado multidimensional, es una técnica estadística utilizada para la visualización y exploración de datos en un espacio de similitud de dos dimensiones, cuando la naturaleza de los elementos en análisis es de una dimensionalidad generalmente mayor.

MIDI: Musical Instrument Digital Interface, protocolo de comunicación serial estándar entre computadores, sintetizadores, secuenciadores, controladores y otros dispositivos musicales electrónicos. La información que representa no se considera audio digital sino una descripción del evento sonoro. También se conoce con este nombre el formato de archivos que almacena la información producida por un dispositivo MIDI.

MITs: Music Intelligent Tutoring System, Sistemas Tutores Inteligentes para Música.

NMF: Non-Negative Matrix Factorization, Factorización de Matrices no negativas, consiste en la descomposición de una matriz cuyos elementos son no negativos en el producto de dos matrices también no negativas, aplicada en computación en los campos de visión por computador, sistemas de recomendación y agrupamiento de elementos.

PCA: Principal Component Analysis, Análisis de Componentes Principales, es una técnica estadística utilizada para reducir la dimensionalidad de un conjunto de datos teniendo preferencia por las dimensiones más importantes.

PMI: Pointwise Mutual Information, Información Mutua entre Puntos, es una medida de asociación utilizada en teoría de la información. También es un algoritmo para identificación de frases musicales.

SOM: Self-Organizing Maps, Mapas Auto-Organizados, es un tipo de red neuronal entrenada de forma no supervisada para producir una representación discreta y de baja dimensionalidad del espacio de entrada (mapa) de unas muestras de entrenamiento.

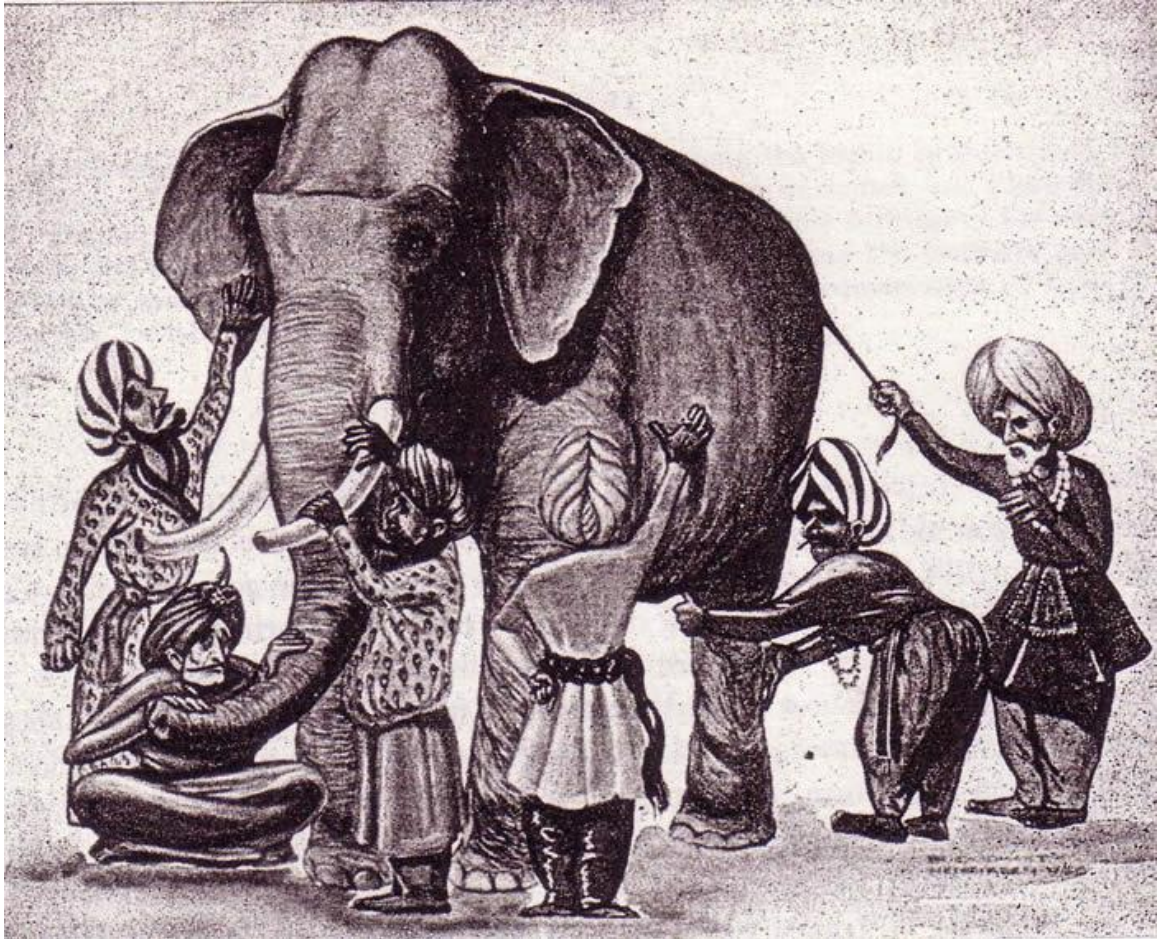
SVD: Singular Value Decomposition, Descomposición en Valores Singulares, es un método para la factorización de matrices, generalmente aplicada en computación al cálculo de la pseudo inversa y a la predicción de datos.

TP: Transition Probabilities, Probabilidades de Transición.

CONTENIDO

RESUMEN	9
LISTA DE FIGURAS	11
LISTA DE TABLAS	13
GLOSARIO	15
GLOSARIO A - PARA LA TERMINOLOGÍA MUSICAL.....	15
GLOSARIO B - PARA LA TERMINOLOGÍA DE COMPUTACIÓN.....	16
INTRODUCCIÓN	23
1. INTERESES DE INVESTIGACIÓN	25
1.1. PROBLEMA.....	25
1.2. APLICACIONES POTENCIALES.....	25
1.3. PREGUNTA DE INVESTIGACIÓN	26
1.4. OBJETIVOS	26
1.4.1. <i>Objetivo General</i>	26
1.4.2. <i>Objetivos Específicos</i>	26
2. MARCO TEÓRICO	27
2.1. COMPONENTE MUSICAL.....	27
2.1.1. <i>Elementos</i>	27
2.1.2. <i>Diseño formal</i>	28
2.1.3. <i>Patrones [8]</i>	29
2.2. COMPONENTE COMPUTACIONAL	30
2.2.1. <i>Tecnología en música</i>	30
2.2.2. <i>Inteligencia artificial</i>	30
2.2.3. <i>Otros conceptos importantes</i>	32
2.2.4. <i>Referentes de otros dominios</i>	33
2.3. ESTADO DEL ARTE Y LIMITACIONES	34
3. METODOLOGÍA	35
3.1. ANÁLISIS	35
3.2. DISEÑO	36
3.2.1. <i>Proceso de identificación estructural</i>	37
3.3. IMPLEMENTACIÓN	39
3.4. PRUEBAS	40
3.5. EVALUACIÓN	42
4. FRAGMENTACIÓN DE LA PIEZA MUSICAL	45
4.1. POSIBLES FRAGMENTACIONES.....	45
4.2. REDUCCIÓN DEL ESPACIO DE BÚSQUEDA.....	45
4.2.1. <i>Limitación por cortes</i>	46
4.2.2. <i>Identificación de Frases y Semifrases</i>	47

4.2.3.	<i>Omisión de fragmentos repetidos</i>	48
5.	CARACTERIZACIÓN DE FRAGMENTOS	49
5.1.	CARACTERÍSTICAS GENERADAS	50
5.1.1.	<i>Características simples</i>	50
5.1.2.	<i>Características compuestas</i>	50
5.1.3.	<i>Perspectiva</i>	52
5.2.	EJEMPLO	53
6.	AGRUPAMIENTO Y SELECCIÓN	55
6.1.	AGRUPAMIENTO	55
6.1.1.	<i>Comparación entre fragmentos</i>	55
6.1.2.	<i>Esquema de agrupamiento</i>	56
6.1.3.	<i>Caracterización del agrupamiento</i>	56
6.2.	SELECCIÓN	57
6.2.1.	<i>Mejor agrupamiento</i>	57
6.2.2.	<i>Maximización</i>	57
7.	IMPLEMENTACIÓN	59
7.1.	MODELO DE REPRESENTACIÓN	59
7.2.	ALGORITMOS	59
7.2.1.	<i>Identificación estructural</i>	60
7.2.2.	<i>Identificación estructural en profundidad</i>	61
7.3.	FORMATOS	63
8.	RESULTADOS	65
8.1.	PARÁMETROS	65
8.2.	FRASES Y SEMIFRASES	66
8.3.	RELACIONES ENTRE FRAGMENTOS	67
8.4.	RELACIONES DE PROFUNDIDAD	74
9.	PRUEBAS Y EVALUACIÓN	81
9.1.	PRUEBAS AUTOMÁTICAS	81
9.2.	EVALUACIÓN POR EXPERTOS	84
9.2.1.	<i>Resultados cuantitativos</i>	85
9.2.2.	<i>Resultados cualitativos</i>	85
	CONCLUSIONES	87
	TRABAJOS FUTUROS	89
	BIBLIOGRAFÍA	91



Fuente: Imagen de autor desconocido, descargada de www.thecollider.com el 10/11/2014

"... Luego que cada ciego palpó un *fragmento* del elefante, el rey se dirigió a uno de ellos y le preguntó: '¿Dime hombre ciego, has visto al elefante? Cuéntame, ¿qué cosa es un elefante?' ..."

Los seis ciegos y el elefante
Parábola india

INTRODUCCIÓN

Entre los años 2000 y 2010 se realizó a título del autor de este trabajo, el desarrollo y puesta en marcha de la plataforma "Armonía al Teclado", un proyecto dirigido por el profesor Horacio Lapidus y que se desarrolló con varias fases que cubrieron el análisis e implementación de una herramienta prototipo para la automatización del proceso de prácticas estudiantiles, el desarrollo de un portal de contenidos y prácticas en línea, la experimentación y análisis de datos con estudiantes del Conservatorio de Música (Facultad de Artes) de la Universidad Nacional de Colombia, y finalmente la puesta en producción y apertura al público en general de la plataforma [1].

Durante las etapas descritas, se recorrió un importante camino en la comprensión desde el punto de vista pedagógico, musical y de ingeniería sobre el uso de herramientas automatizadas para el apoyo y asistencia del proceso de aprendizaje en las diferentes áreas musicales.

Gracias a este acercamiento, utilizando la ingeniería como apoyo al fortalecimiento del proceso de aprendizaje dentro del área musical y debido a la carencia de productos y documentos similares en el mercado (donde se encuentran numerosas aplicaciones para producción, análisis, composición y ejecución autónoma de instrumentos, pero con una escasa oferta de apoyo pedagógico en algunas áreas), resultó de gran interés profundizar en el desarrollo y fortalecimiento de las líneas de aplicación del conocimiento ingenieril en dicho espacio. Campos como: Sistemas Inteligentes, Representación del Conocimiento, Multimedia, Educación Virtual, Interacción Humano Computador y Reconocimiento de Patrones, son, además de interesantes, pertinentes en términos de su materialización como herramientas de apoyo a los ámbitos músico-pedagógicos.

De acuerdo con lo expuesto se inició una búsqueda, recopilando información sobre cómo se lleva a cabo el día a día de las prácticas estudiantiles y cuáles son las herramientas de docencia, especializadas en el ámbito musical. Como resultado de este periodo de revisión del estado del arte, se plantearon ideas sobre posibles nuevas aplicaciones de acompañamiento y tutoría interactiva, así como de mejoras a ideas y aplicaciones anteriores donde cobró una importancia relevante, la línea de acción relativa al análisis automático de la estructura musical.

En consecuencia, se propuso una metodología y se desarrolló un algoritmo para la identificación estructural automática de piezas musicales, soportado en la caracterización y relación de los diferentes elementos contenidos; este resultado demostró la posibilidad de identificar de manera automática y precisa la estructura subyacente de una pieza musical utilizando análisis de patrones de similitud.

El análisis de las relaciones de similitud entre fragmentos se abordó desde una aproximación novedosa que hace uso de la exploración de rasgos característicos por vistas múltiples, esta idea se propuso tomando como referencia la enseñanza de la antigua parábola india "Los

ciegos y el elefante” cuyo texto expresa la incapacidad para reconocer una realidad en su totalidad al no tener en cuenta todos los elementos que la componen y la importancia de valorar diferentes perspectivas.

El presente documento expone la metodología propuesta con cada una de sus fases, enseguida, presenta los detalles del proceso de *identificación estructural*, describiendo a profundidad las tareas de fragmentación, caracterización, agrupamiento y selección. Una vez hecho esto, se evidencian los detalles destacados en la implementación del algoritmo computacional, se presenta un caso de ejemplo de aplicación, y se finaliza con la descripción de las pruebas y su respectiva evaluación.

1. INTERESES DE INVESTIGACIÓN

1.1. PROBLEMA

Gran parte de las tareas realizadas por los músicos tienen como requisito un grado de comprensión de la forma o estructura de una pieza musical; generar música (interpretar, acompañar, componer, improvisar), analizar (entender, caracterizar, organizar) y enseñar, son algunas de las más comunes.

Durante el análisis, el músico percibe algunas de las variables propias del pasaje escuchado o leído, y utilizando criterios previamente aprendidos, encuentra diferentes rasgos característicos que él organiza para agrupar o discriminar fragmentos, determinar repeticiones, motivos, secuencias de acordes y patrones rítmicos, que al ser identificados con cierta experiencia evidencian la estructura subyacente de la pieza musical.

Una de las características de este proceso es la diversidad de resultados que puede presentar un análisis, ya que para diferentes tipos de músicos los criterios y medidas pueden cambiar y afectar la identificación de los rasgos presentes, haciendo poco probable una evaluación única o exacta del ejercicio.

En este orden de ideas, se entiende que el análisis requiere la dedicación de una persona experta en el área que aplique su conocimiento. En algunos casos esta labor se vuelve costosa, tediosa y/o repetitiva, por lo cual sería de gran utilidad la identificación automática por medio de un sistema inteligente.

1.2. APLICACIONES POTENCIALES

Un mecanismo, que de manera automática describa los rasgos de diseño formal de una pieza musical puede utilizarse para desarrollar herramientas que apoyen diferentes labores musicales, como lo describen las investigaciones documentadas en la literatura [2, 3, 4].

A continuación se mencionan algunas de las aplicaciones más importantes desde el punto de vista de esta investigación:

- a. Generación automática: cuando un músico interpreta una pieza musical que desconoce a priori va aprendiendo su estructura a medida que la recorre, para anticipar en el tiempo ciertos momentos clave e introducir elementos de expresión que destaquen la emotividad de la pieza, generalmente relacionada con el material característico con que ha sido compuesta. En este sentido, la interpretación realizada por un sistema computacional podría acercarse a la realizada por un humano si en esta, se considera la información adicional sobre la estructura de la obra, que permite

integrar dinámicas relacionadas a sus rasgos característicos. De forma similar, se puede aplicar la generación por computador para acompañar la interpretación de un solista adicionando material “improvisado” o para apoyar el proceso de composición, recomendando nuevas ideas generadas a partir del material característico de la pieza.

- b. Análisis automático: la identificación estructural es igualmente importante para el análisis automático de grandes volúmenes de información, con el fin de caracterizar las piezas musicales con sus rasgos propios y así poder emplear herramientas computacionales de organización y recuperación de información musical.
- c. Apoyo a la pedagogía: dada la importancia del análisis estructural en las labores del músico, la realización repetida y progresiva de ejercicios de identificación se encuentra incluida como un elemento de su formación básica, asistido en su desarrollo, evaluación y corrección por un profesor experto en el tema. La delegación hacia un sistema automático de las "tediosas y repetitivas" labores del docente en este proceso, liberaría gran parte de su tiempo para dedicarse a un seguimiento más general y completo de la evolución de sus estudiantes. Por otro lado, se presentaría una mejora en el escenario de práctica, ya que los estudiantes demuestran más confianza y menos tensión en sus prácticas al no depender de una reacción humana frente a sus errores que además estarían siendo registrados y evaluados [1]. También puede ser útil en el desarrollo de entornos de prácticas asistidas de interpretación, improvisación o acompañamiento, que generalmente necesita la interacción con otro(s) estudiante(s).

1.3. PREGUNTA DE INVESTIGACIÓN

¿Es posible identificar de manera automática y precisa la estructura de una pieza musical dada?

1.4. OBJETIVOS

1.4.1. Objetivo General

Identificar automáticamente la estructura de una pieza musical dada.

1.4.2. Objetivos Específicos

- Proponer una representación computacional de las características para los elementos musicales del problema.
- Analizar y validar características propias de los elementos musicales y características abstraídas de las relaciones entre ellos a partir de la experimentación.
- Diseñar una metodología para la selección y uso de características musicales que aporten a la identificación estructural de una pieza musical.
- Implementar un algoritmo de identificación estructural automática para piezas musicales.

2. MARCO TEÓRICO

La investigación que aquí se expone se vale de referentes tanto del área musical como del área computacional, por tal motivo y en pro de una mejor exposición del marco teórico, este capítulo se divide en dos componentes en los que se profundizan los temas de interés respectivos a cada uno de ellos.

2.1. COMPONENTE MUSICAL

Para soportar los conceptos musicales expuestos se hará referencia a textos traducidos al español del libro *Form in Tonal Music* [5], texto de referencia en el plan de estudios para análisis formal de la música de algunos programas académicos (Universidad de los Andes y otros).

2.1.1. Elementos

Para una comprensión más amplia del diseño formal es necesario definir primero los elementos que hacen parte del discurso musical; algunos de estos elementos están descritos en términos de otros conceptos musicales básicos que se pueden encontrar en la literatura musical introductoria [5, 6, 7].

- **Pieza musical:** composición con un sentido completo. La percepción del sentido de completitud de una pieza musical es relativa a su posible ubicación dentro de una creación de mayor extensión o a su comparación con otras composiciones. Para el presente estudio se entenderá la pieza musical como el elemento mayor propuesto para el análisis estructural.
- **Fragmento:** aparte o fracción de una *pieza musical*, adquiere sentido cuando contiene como mínimo dos notas musicales (secuencia) construyéndose así una diferencia de alturas o intervalo.
- **Periodo:** *fragmento* compuesto por dos o más *frases*.
- **Frase:** *fragmento* que tiene un sentido completo en sí mismo, divisible en *semifrases*.
- **Semifrase:** *fragmento* construido a partir de la unión de dos o tres *motivos*.
- **Tema:** *fragmento* formado por uno o más *motivos*, tiene un antecedente y un consecuente, los cuales le dan el carácter de idea musical concreta. Estas partes pueden estar unidas por un *fragmento* eslabón.

- Motivo: *fragmento* melódico corto que presenta un diseño característico y ocurre una y otra vez en una *pieza musical* como elemento unificador, generalmente con algunas variantes en sus reapariciones.

2.1.2. Diseño formal

El análisis de la morfología musical tiene como componente importante la organización de aspectos tales como la melodía, la frase, el periodo y las partes, elementos que conforman (de lo simple a lo complejo) una composición musical tonal -observada desde este punto de vista- como un universo de los elementos en cuestión [8].

A las unidades melódicas más breves se las conoce como motivos; para que el discurso musical tenga un sentido más contundente se hace uso de la variación motívica como herramienta de construcción de frases coherentes. Estas variaciones o transformaciones pueden consistir en repetición, adaptación armónica (por enlace, por transporte, por cambio de nivel), progresión, espejo, cambios de la curva melódica, expansión y compresión rítmica y combinaciones de las anteriores.

Dichas variaciones pueden ser aplicadas a un motivo inicial e integradas con otros motivos contrastantes, para generar una frase completa, tal y como se muestra en la Figura 1.

Figura 1. Idea musical completa dividida en motivos y variaciones.

The image shows a musical score for Cello in 6/8 time, divided into three systems. The first system (measures 1-3) shows motive m1 (a) and m2 (b) under phrase A. The second system (measures 4-6) shows motive m3 (c) and m4 (d) under phrase B. The third system (measures 7-9) repeats phrase A with motives m1 (a) and m2 (b). The score is written in bass clef with a key signature of one sharp (F#) and a time signature of 6/8.

En la imagen se presentan los motivos m1 y m2 que se desarrollan a partir de una misma célula y forman la frase A. La frase B, presenta los nuevos motivos contrastantes m3 y m4, también desarrollados a partir de una misma célula transportada hacia abajo. Finalmente se repite la frase A con una pequeña variación para finalizar.

Los motivos se varían, distribuyen y relacionan para formar partes mayores, conocidas como frases, las cuales aportan un sentido más concreto del lenguaje al fragmento. Las frases a su vez, se agrupan secuencialmente para definir periodos que finalmente integran elementos mayores llamados partes.

Tonalidad, tema y textura: "Elementos que al combinarse son útiles para segmentar la música en frases y secciones y así proporcionar las bases primarias para interpretar el diseño formal de una pieza" [5]

Profundizando en este último enunciado y de manera retrospectiva, los cambios armónicos, de melodía y de ritmo, son los elementos con mayor importancia, que al analizarse en forma conjunta permiten segmentar las frases en motivos e ideas unitarias, con base en sus relaciones.

2.1.3. Patrones [8]

Se define como un elemento recurrente basado en la repetición, la periodicidad y la simetría, aunque puede tener variaciones en algunas de sus presentaciones. En el caso musical, existen patrones en los diferentes niveles de división de una pieza (Motivo, Semifrase, Frase, etc.) los cuales con una interpretación adecuada pueden revelar su estructura subyacente.

Segmentación/Unión de Unidades Formales: En los procesos de subdivisión operan factores de segmentación y en los de agrupamiento factores de unión. En la música tonal, los principales factores de Segmentación/Unión son:

- a. Proximidad: opera la ley de la proximidad de la Gestalt (los elementos más próximos tienden a ser agrupados)
- b. Similitud: opera la ley de similitud de la Gestalt (los elementos similares tienden a ser agrupados). La similitud puede operar en parámetros o nociones muy variados (motivo, ritmo, timbre, métrica, etc.)
- c. Procesos comunes de alto nivel: la presencia de un proceso continuo, a lo largo de varias partes, tiende a unificar a estas en una parte de jerarquía superior. Por ejemplo, las fluctuaciones métricas (acelerando, ritardando)¹ o dinámicas (crescendo, decrescendo)² proveen una forma de enlace entre varias unidades, si estas participan de un mismo proceso.
- d. Conclusividad: se refiere a la sensación de reposo o conclusión que presenten las unidades formales.

¹ Acelerando (o accel.) es una indicación para acelerar gradualmente el tiempo de una pieza hasta que se indique otra cosa. Ritardando (o rit.) es una indicación para disminuir gradualmente el tiempo de la música.

² Crescendo (en italiano "creciendo") indica que se debe aumentar gradualmente la intensidad del sonido, es decir, un matiz dinámico de transición. Decrescendo o disminuyendo (del italiano "decreciendo") indica que se debe reducir gradualmente la intensidad del sonido, es decir, un matiz dinámico de transición.

2.2. COMPONENTE COMPUTACIONAL

2.2.1. Tecnología en música

Comprende la aplicación de la tecnología como apoyo a gran parte de los procesos realizados en el ámbito musical (grabación, composición, almacenamiento, interpretación, análisis, etc.).

El extenso y variado listado de líneas de investigación en el área de la tecnología musical incluye áreas tales como: tecnologías multimediales, ejecución instrumental y composición automática, Sistemas Tutores Inteligentes para Música, en adelante (MITS), entrenamiento auditivo asistido, enseñanza virtual (E-Learning) y evaluación automática [3], entre otros, soportados generalmente en las ciencias básicas aplicadas como la matemática y la estadística.

2.2.2. Inteligencia artificial

Desde que se acuñó el término Inteligencia Artificial, en adelante (IA) en 1956, se han planteado varias definiciones. Desde un punto de vista práctico y pertinente con esta investigación, se asume que la IA se remite al diseño de agentes inteligentes o comportamientos inteligentes en artefactos, y el aprendizaje como evidencia de esos comportamientos [9].

No obstante, cabe mencionar que uno de los intereses más fuertes en el uso de la computación, como apoyo a las prácticas musicales, ha sido guiado en gran medida por la necesidad de suplir o complementar dichas labores a través de la emulación de lógicas previamente abstraídas de un experto humano [1]. Campos como la percepción rítmica, la composición, la ejecución [3] y la evaluación automática se hacen más robustos desde un enfoque tecnológico de inteligencia artificial aplicada.

2.2.2.1. *Representación del conocimiento*

Una de las áreas de la inteligencia artificial es la representación lógica del conocimiento, la cual abarca para el énfasis musical, desde la representación de estructuras básicas, tales como los acordes, compases, armaduras, claves, tiempos y ritmos [10], hasta conceptos más abstractos y complejos como la relación entre elementos y la armonía musical [11].

2.2.2.2. *Identificación estructural automática*

Esta área se focaliza en la descripción conceptual y jerárquica de los elementos representativos de la música [12] siendo así el soporte para el funcionamiento de otras áreas del día a día musical asistido por computador.

La identificación estructural se realiza aplicando herramientas de detección de relaciones y patrones [2] (redes neuronales, algoritmos genéticos, árboles de decisión [3]) a partir de una entrada de usuario, que puede ser, un acorde, una partitura, una célula rítmica, un objeto sonoro (señal de audio) o algún otro elemento estructural.

2.2.2.3. Identificación de frases musicales

Dentro de la identificación estructural, la identificación de frases tiene una importancia predominante dado que estas corresponden a uno de los niveles de organización más comunes para la percepción auditiva al exponer una idea completa.

En la literatura se encuentran descritos algunos modelos computacionales que realizan identificación de *frases* [4] y se listan a continuación:

- Grouper.
- LBDM: Local Boundary Detection Model.
- GPR2a: Grouping Preference Rules (Rest).
- GPR2b: Grouping Preference Rules (Attack-point).
- GPR3a: Grouping Preference Rules (Change in register).
- GPR3d: Grouping Preference Rules (Length change).
- PMI: Pointwise Mutual Information.
- TP: Transition Probabilities.
- IDyOM: Information Dynamics Of Music.
- Hybrid: Combina Grouper, LBDM, GPR2a e IDyOM.

Igualmente, la literatura ofrece un análisis comparativo de los resultados alcanzados por estos modelos, los cuales se encuentran expuestos en la Tabla 1.

Tabla 1. Comparación de resultados para modelos de identificación de frases.

Model	Precision	Recall	F1
Hybrid	0.87	0.56	0.66
Grouper	0.71	0.62	0.66
LBDM	0.70	0.60	0.63
IDyOM	0.76	0.50	0.58
GPR2a	0.99	0.45	0.58
GPR2b	0.47	0.42	0.39
GPR3a	0.29	0.46	0.35
GPR3d	0.66	0.22	0.31
PMI	0.16	0.32	0.21
TP	0.17	0.19	0.17
Always	0.13	1.00	0.22
Never	0.00	0.00	0.00

El comparativo hace referencia a las medidas de evaluación comunes en el área de Recuperación de Información (*precision, recall y f1*). El proceso y los resultados se encuentran descritos en [4].

En términos generales, basta con aclarar que de acuerdo con la **Tabla 1**, el modelo tiene una mejor respuesta cuando su medida *f1* (*media armónica*) es mayor.

2.2.3. Otros conceptos importantes

2.2.3.1. *Minería de datos*

Es un campo interdisciplinario de las ciencias de la computación que tiene como objetivo la generación de información estructurada por medio de la extracción de información de un conjunto de datos, haciendo uso de técnicas de análisis de patrones [13].

- **Minería de secuencias:** Es un caso particular de la minería de datos, en el que se trata de encontrar patrones en conjuntos estructurados [14]. En el campo del análisis musical automático se han utilizado estos métodos sobre eventos sonoros elementales [15], formaciones secuenciales de notas [2], o en corpus completos de canciones para la identificación de patrones distintivos [16]. Se pueden obtener reglas de asociación a partir del descubrimiento de patrones basados en frecuencias.

2.2.3.2. *Aprendizaje computacional*

Conocido como aprendizaje automático o aprendizaje de máquina, esta disciplina da cuenta de la construcción y estudio de algoritmos que pueden aprender de los datos [17]. Se considera un subcampo interdisciplinario entre las ciencias de la computación y la estadística.

- a. **Aprendizaje supervisado:** es un conjunto de técnicas para deducir una función o comportamiento a partir de una serie de datos de entrenamiento o elementos previamente etiquetados o categorizados, los cuales son una representación del resultado deseado [18].
- b. **Aprendizaje no supervisado:** es un modelo de aprendizaje automático en el cual el resultado es generado a partir de las observaciones de una serie de variables aleatorias sin que exista un conocimiento a priori [19]. Técnicas ampliamente documentadas y utilizadas como: k-means, mixture models, agrupamiento jerárquico, modelos ocultos de Markov, PCA, SVD, análisis de componentes independientes (ICA), factorización de matrices no negativas (NMF) y mapas auto-organizados (SOM), utilizan algoritmos de aprendizaje no supervisado [20].

2.2.3.3. *Agrupamiento*

También conocido como Análisis de Grupos, es una técnica del aprendizaje computacional que tiene como objetivo relacionar una serie de elementos de acuerdo con un criterio que suele ser de distancia o similitud [17].

- a. **Agrupamiento no supervisado:** El objetivo del aprendizaje no supervisado es tratar de encontrar la estructura de grupos oculta en los datos sin un entrenamiento previo [17].
- b. **Agrupamiento jerárquico:** Es un método de agrupamiento no supervisado ampliamente utilizado en la minería de datos que tiene como objetivo inferir la jerarquía entre los objetos utilizando como referente las distancias entre ellos [21].

- c. Método de Kernel: Es una técnica del aprendizaje computacional empleada en el análisis de patrones, con el objetivo de buscar y estudiar los tipos de relaciones existentes entre los elementos de un conjunto de datos. Para esta investigación, se utilizaron los conceptos de matriz de distancias y Kernel de cadenas, como insumos para la comparación de elementos con características heterogéneas y los agrupamientos jerárquicos [22, 23, 26, 28].

2.2.4. Referentes de otros dominios

El análisis de patrones tiene un amplio espectro de referencias en dominios diferentes que abarcan desde el reconocimiento de patrones de escritura, hasta la identificación de interacción entre proteínas [23, 24] pero, para el caso de esta investigación, un ejemplo que ilustra mejor el tipo de aplicación que se busca dar a las herramientas computacionales, es el siguiente:

Figura 2. Ejemplo ilustrativo de análisis de patrones en texto.

<p>So began the story of chunking, one of the most useful and least understood phenomena in human cognition...</p> <p>sobeganthestoryofchunkingoneofthemostusefulandleastunderstoodphenomenainhumancognition...</p> <p>sobe*gan*the*sto*ry*of*chunk*ing*one*ofthe*most*use*fulan*dle*astund*er*sto*od phen*omen*ain*huma*nco*gnition...</p>

Ejemplo de análisis de patrones sobre texto para encontrar los espacios (divisores de palabras) por correlación. Fuente: [25]

El ejemplo ilustrativo de la Figura 2, describe cómo, al eliminar los espacios que separan las palabras del texto enunciado en la primera parte de esta, se produce la secuencia de letras presentada en segunda instancia. El problema es entonces reconstruir la ubicación de los espacios a partir de la secuencia de caracteres.

De acuerdo con la literatura consultada, se entiende que por medio de un algoritmo fundamentado en principios de baja entropía en la relación de secuencias de caracteres de tamaño variable, se puede llegar a inferir la partición tal como se ejemplifica en la línea final de la Figura 2 donde se evidencia un resultado aceptable [25].

Siguiendo con el análisis, si se hace una comparación entre la secuencia de caracteres sin espacios presentada en la segunda línea, y una pieza musical, se puede encontrar una relación directa, donde existe un lenguaje subyacente y una necesidad de abstraer de forma automática

los puntos de corte para separar los elementos que lo conforman, utilizando como material de referencia únicamente la cadena de símbolos de entrada y el principio de baja entropía que se puede traducir en este caso como alta correlación entre elementos identificados.

2.3. Estado del Arte y Limitaciones

Para sintetizar de manera breve un estado del arte referente específicamente a la identificación estructural automática, podemos mencionar que sus aproximaciones se han implementado dividiendo las piezas musicales en frases de alto nivel jerárquico por medio de la aplicación de algoritmos que hacen uso del espectro acústico o por medio de los eventos musicales, contemplando las variables ritmo y melodía.

Dentro de los métodos documentados en el área, podemos identificar algunas limitaciones y deficiencias:

- a. Únicamente se utilizan las variables de ritmo y melodía, despreciando otros atributos naturales que presentan los fragmentos musicales como son el acento, la intensidad, etc.
- b. Se utilizan las variables ritmo y melodía, de una forma completamente lineal, cuando de ellas se pueden extraer otras características importantes y representativas de los fragmentos como son la densidad rítmica, proporciones sonoras, distribuciones de tonos, potenciales de tensión, etc.
- c. Las soluciones empleadas utilizan criterios de decisión soportados en una valoración unificada de la distancia, llevando las variables ritmo y melodía, a una única forma rítmico-melódica o bien definiendo una valoración única como resultado de alguna operación entre las comparaciones parciales, esto impide que los fragmentos puedan ser valorados como similares en unos aspectos y diferentes en otros.

De esta forma podemos concluir que los algoritmos se han elaborado específicamente para piezas musicales fundamentados claramente en las relaciones de ritmo y melodía, dejando de lado aquellas que de forma novedosa desarrollan su estructura y relaciones en otras características.

3. METODOLOGÍA

Este capítulo describe la metodología propuesta para el desarrollo de la investigación, proponiendo desde un punto de vista general un modelo de ejecución por fases, descrito gráficamente en la Figura 3.

Figura 3. Metodología general por fases.

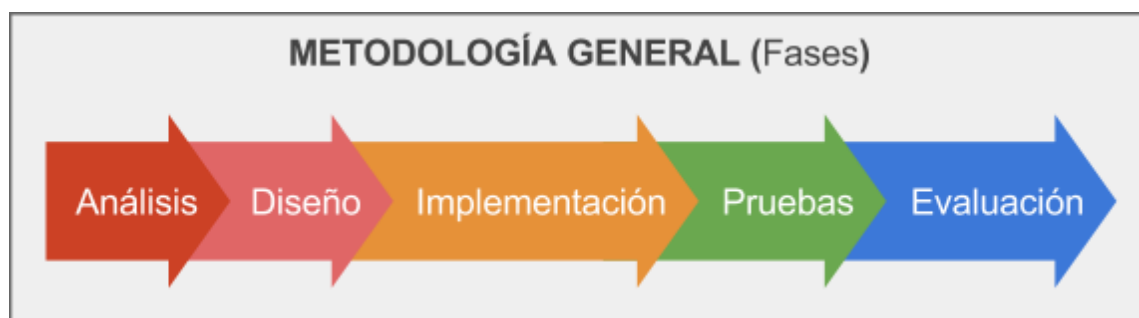
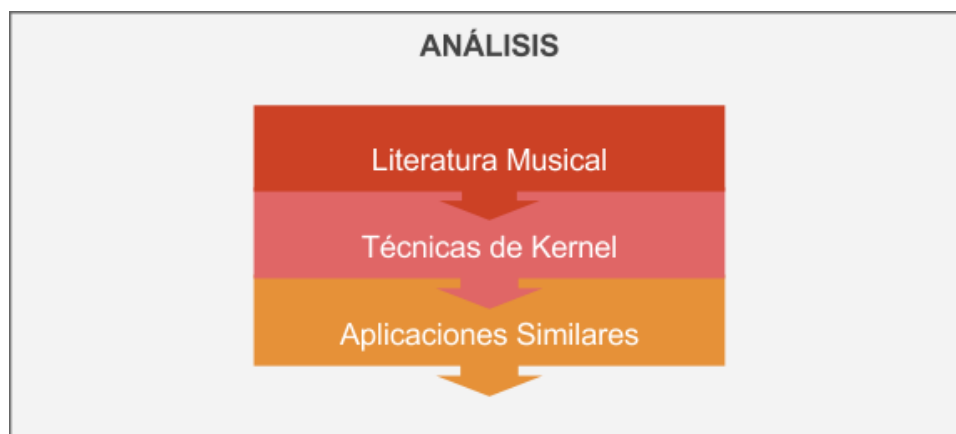


Diagrama en fases de la metodología general aplicada.

3.1. ANÁLISIS

La fase de análisis se encuentra definida como una secuencia de actividades que incluyó la revisión del material bibliográfico y el estado del arte, tal y como se expone en la Figura 4.

Figura 4. Esquema de la fase de análisis.



Como se observa en el flujo de actividades de análisis, una vez revisada la literatura musical, en la segunda instancia y debido al importante papel que en los últimos años han tenido los métodos geométricos y estadísticos en la solución de problemas de clasificación y búsqueda de patrones; se realizó una revisión del uso de las técnicas de Kernel para determinar relaciones entre los componentes del discurso musical, estas técnicas se encuentran

fundamentadas en el uso de medidas de similitud, utilizando las diversas características de los elementos del problema. La decisión de seleccionar esta técnica en lugar de otras utilizadas con anterioridad como las redes neuronales y las cadenas de Markov, se soporta en la facultad especial de operar sobre características de naturaleza heterogénea, estas características propias de los fragmentos musicales se profundizarán en el capítulo 5.

Finalmente, esta fase tuvo en cuenta las áreas de la biología [26], el procesamiento de texto [22] y la visión por computador [27] donde se cuenta con referentes exitosos del uso de las técnicas propuestas por el aprendizaje computacional para la solución de problemas similares.

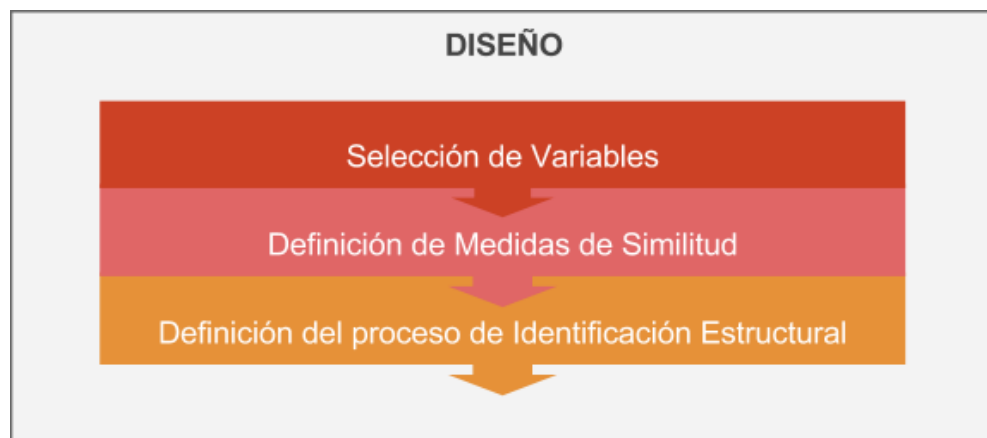
Igualmente, en la bibliografía relacionada con aplicaciones tecnológicas en el área de la música se encuentra un amplio espectro de técnicas para el análisis de diferentes patrones musicales que se tuvieron en cuenta.

Vale la pena aclarar que en el área de la *tecnología musical* se han encontrado desarrollos referentes al análisis estructural desde el punto de vista físico [15] abordando temas de procesamiento de señales de audio. Para el presente caso, la metodología propuesta se enfocó en la representación simbólica musical a partir de *partituras* digitales, basadas en eventos numéricos, abordando el problema en un sentido netamente estructural.

3.2. DISEÑO

La fase de diseño tiene como objetivo, definir las variables, las medidas de similitud y los procesos relacionados con la identificación estructural como se ilustra en el esquema de la Figura 5.

Figura 5. Esquema de la fase de diseño.



Los elementos musicales en su descripción básica, tienen un conjunto de propiedades que se pueden utilizar de primera mano para aplicar al proceso de identificación estructural.

Adicionalmente, haciendo uso de las posibilidades ofrecidas por los métodos de Kernel [28] se planteó una extracción de características complementarias, generadas a partir de las relaciones subyacentes entre elementos primarios, para así realizar un análisis de similitud en un dominio superior; esto aumentó la cantidad de variables y con ello la dimensión del problema, lo cual permitió realizar un análisis con nuevas perspectivas de observación, mejorando así la comparación clásica punto a punto [3] y teniendo en cuenta semejanzas entre variables de un nivel de complejidad mayor.

Con nuevas características complementarias para los elementos musicales es importante resaltar que los métodos de Kernel presentan una amplia variedad de posibilidades para el análisis de diferentes tipos de datos y exponen ventajas en la comparación de información heterogénea. Por tal motivo se propuso realizar una identificación y selección de las medidas de similitud más adecuadas a las características extraídas de los nuevos elementos.

3.2.1. Proceso de identificación estructural

Teniendo claramente definidas las características naturales de los elementos musicales y las nuevas características generadas, se definió un proceso para la identificación estructural tal y como se ejemplifica en el esquema de la Figura 6.

Figura 6. Esquema del “proceso de identificación estructural”.



Para una mejor comprensión del proceso en mención, es importante aclarar primero la idea de *sección* y *fragmento*.

"Sección: Puede entenderse y reconocerse como un pasaje musical auto-contenido definido por cierto grado de asociación de su material temático, tonalidad, ritmo y/o textura" [5]

La *sección* es un *fragmento*, también conocido como unidad formal y ha sido denominada de diversas maneras por muchos autores: *períodos*, *frases*, *semifrases*, *incisos*, *miembros de frase*, etc. Las secciones de una obra tienen diversos niveles jerárquicos y estos solo pueden evaluarse de acuerdo a su relación con toda la obra. Habitualmente la jerarquía de las partes está en relación directa con sus dimensiones relativas y el material contenido. En la Figura 7 se presenta un ejemplo.

Figura 7. Ejemplo de jerarquía musical.

The image shows a musical score for Cello in G major and 6/8 time. The score is divided into two main sections, A and B. Section A is further divided into sub-sections 'a' and 'b'. Section B is divided into sub-sections 'c' and 'd'. Motives are labeled m1, m2, m3, and m4. Red brackets indicate the hierarchical grouping of notes into motives, sub-phrases, and phrases.

En el ejemplo expuesto, la máxima jerarquía corresponde a las secciones AB, la primera subdivisión a las semifrases ordenadas como aba-cdc' y la segunda subdivisión a los motivos (m1, m2 m3, m4 en su respectivo orden).

Retomando la acotación descrita para la pieza musical, y fundamentado en la anterior definición de *sección*, se puede extender la aseveración para todos los tipos de fragmentos:

“La percepción del significado de un fragmento es subjetiva a su papel dentro de una estructura superior o a su comparación con otros fragmentos.” [5]

Esta subjetividad del significado de un fragmento, mencionada tanto en los libros de teoría musical [5, 6] como en los referentes de investigación en identificación estructural [8, 12], permite visualizar todos los posibles cortes realizados a la pieza musical como fragmentos, donde algunos de ellos posiblemente tomen el rol de *motivo*, *semifrase*, *frase*, *tema*, *eslabón* o *melodía* según la relación con sus fragmentos vecinos, internos o de alguno que lo contenga.

Hecha esta aclaración conceptual, a continuación se describen cada una de las fases del proceso de identificación estructural.

3.2.1.1. Fragmentación

Una *fragmentación* se define como la división en partes o fragmentos específicos de una pieza musical. Para una comprensión más clara del término, se puede hacer la analogía con una cinta que ha sido cortada en unos puntos determinados.

El proceso inicia entonces cuando se toma la pieza musical completa y se corta de diferentes formas, para así tener un abanico de fragmentaciones candidatas; cada una de estas fragmentaciones es una forma única y diferente de dividir la pieza completa en partes menores.

3.2.1.2. *Caracterización*

Para cada uno de los posibles fragmentos que se generaron en el paso anterior, se realizó el proceso de extracción de características. De esta manera cada uno de los fragmentos obtuvo una identidad fundamentada en sus propiedades naturales.

3.2.1.3. *Agrupamiento*

Es importante aclarar que este paso se aplicó de forma independiente a cada una de las fragmentaciones candidatas, teniendo en cuenta únicamente los fragmentos que lo conformaban.

El proceso de agrupamiento se refiere entonces a la identificación de grupos diferenciables dentro de la población de fragmentos analizados; dicha identificación está fundamentada en criterios de similitud que emergen de la comparación de sus características propias.

3.2.1.4. *Selección*

El proceso finalizó con la *selección* de la fragmentación que presentó el mejor agrupamiento posible. El mejor agrupamiento es aquel que tiene una mayor verosimilitud que los demás.

Como resultado de este proceso se obtuvo, por un lado, una forma única de fragmentar la pieza musical que se caracteriza por tener las mejores condiciones de relación entre sus fragmentos identificados, y por otro lado, se encontró también una descripción de dichas relaciones que se remitió a las características del agrupamiento existente entre los fragmentos.

3.3. IMPLEMENTACIÓN

Para el desarrollo del algoritmo se propuso una metodología de implementación de software evolutiva basada en prototipos, donde se inició con un análisis del algoritmo a implementar, se procedió con un diseño de la solución propuesta y se desarrolló un prototipo³ que finalmente se expuso a pruebas comparativas con una o varias piezas musicales.

El siguiente ciclo inició con un nuevo análisis que incluyó una evaluación del resultado obtenido, esto con la colaboración de un músico experto. Esta evaluación permite comprender las diferencias de la solución automática con la solución esperada, para así ajustar el diseño y los parámetros en una nueva iteración.

El esquema de la fase de implementación mostrado en la Figura 8 presenta de manera gráfica y resumida la metodología propuesta para la implementación.

³ Ver código fuente en el CD anexo

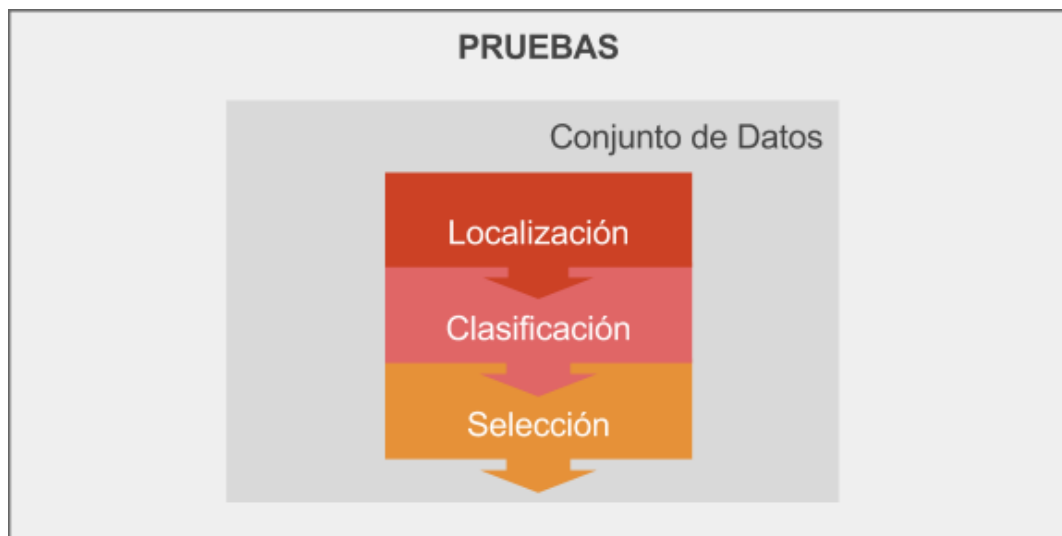
Figura 8. Esquema de la fase de implementación.



3.4. PRUEBAS

La materia prima para el desarrollo de las pruebas es el conjunto de datos, el cual debe ser localizado, clasificado y seleccionado como se muestra en el esquema de la fase de pruebas de la Figura 9.

Figura 9. Esquema de la fase de pruebas.



Este esquema ilustra el proceso propuesto para la definición del conjunto de datos al cual se le aplicaron las pruebas.

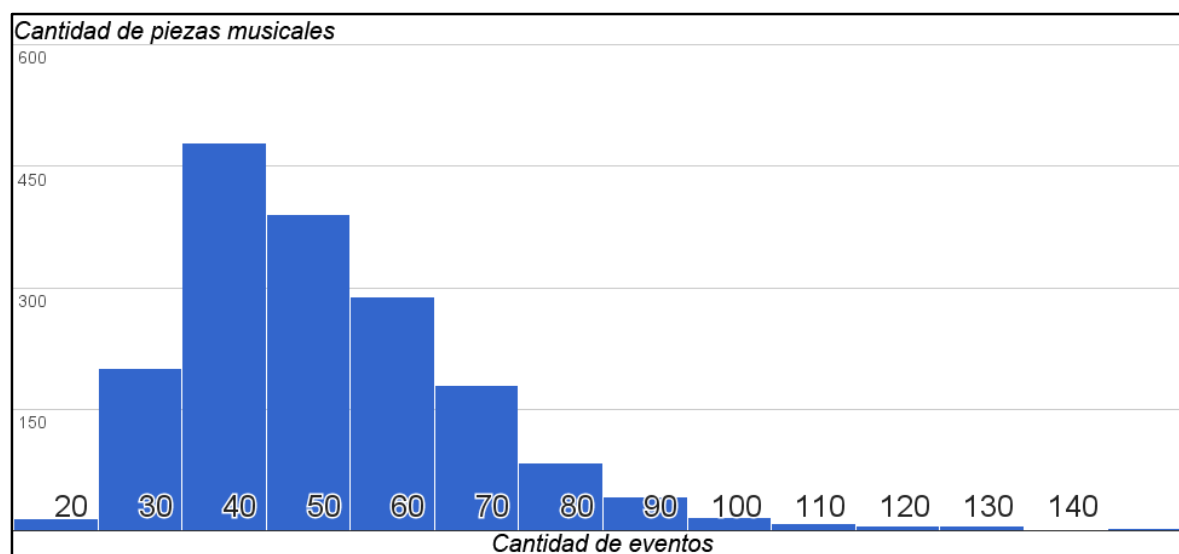
Inicialmente se realizó una búsqueda de archivos MIDI, donde se encontraron bibliotecas de gran variedad, compuestas por melodías de jazz, tango, música clásica, corales y música popular, pero tras la revisión de la bibliografía relacionada con la aplicación de la *tecnología musical* en la identificación de frases, se localizó una base de datos estandarizada con la que se han realizado las pruebas de numerosos trabajos en el campo [4].

El conjunto de datos EFSC - Essen Folk Song Collection (Colección Essen de canciones folclóricas) es una biblioteca de melodías de Alemania y países cercanos que fue ensamblada en el *Gesamthochschule* de la Universidad Essen [29]. Las melodías se encuentran almacenadas en formato Kern (archivos .krn).

La colección EFSC contiene un grupo de melodías folclóricas alemanas llamado ERK [30] donde se encuentran un total de 1.705 piezas musicales, con 78.995 eventos musicales o notas en total.

Como se muestra en la Figura 10, la cantidad de notas musicales por melodía, oscila entre 30 y 150, y la mayor cantidad de piezas con alrededor de 45 sonidos contenidos.

Figura 10. Histograma de eventos en melodías ERK.



Histograma de eventos para las melodías del conjunto de datos de prueba

Este conjunto de datos presenta dos ventajas sobre las bibliotecas MIDI localizadas inicialmente; para comenzar, ERK fue etiquetado en un primer nivel de fragmentación de forma manual por músicos expertos, como lo ilustra el ejemplo de melodía ERK que se presenta en la Figura 11, por otra parte, ERK tiene resultados documentados de aplicación de otros modelos de fragmentación contra los cuales se puede comparar un nuevo modelo.

Figura 11. Ejemplo de melodía ERK etiquetada.

BIN I NET A PUERSCHLE

Los arcos punteados representan la fragmentación etiquetada. Descargado de <http://kern.ccarh.org/>

3.5. EVALUACIÓN

Para esta fase, se siguió el proceso ilustrado en el esquema de la Figura 12.

Figura 12. Esquema de la fase de evaluación.



Es importante destacar que el conjunto de datos ERK [30] fue utilizado en la validación automática de las fragmentaciones para los niveles de jerarquía superior (identificación de *frases* y *semifrases*) aprovechando la particularidad de incluir en su contenido, las marcas de donde inician y terminan las *frases/semifrases* de cada una de sus piezas. Adicionalmente, se

tuvo la ventaja de poder comparar los resultados obtenidos con los trabajos anteriores en el área.

Por otra parte y en dos sentidos, primero para la validación del resultado producido por el algoritmo en cuanto a las relaciones identificadas entre los fragmentos (familiaridad), y segundo, para la validación de los resultados en niveles de jerarquía menor y dado que no se localizó ningún otro conjunto de datos con etiquetados manuales de estos aspectos, se propuso la realización de una validación manual con la colaboración de músicos con conocimiento en diseño formal.

Posteriormente se realizó una lectura y análisis de los resultados obtenidos tanto en la validación automática como en la aplicación de las pruebas manuales para así obtener una medición objetiva de la eficacia del algoritmo desarrollado y por ende, del proceso de identificación estructural propuesto.

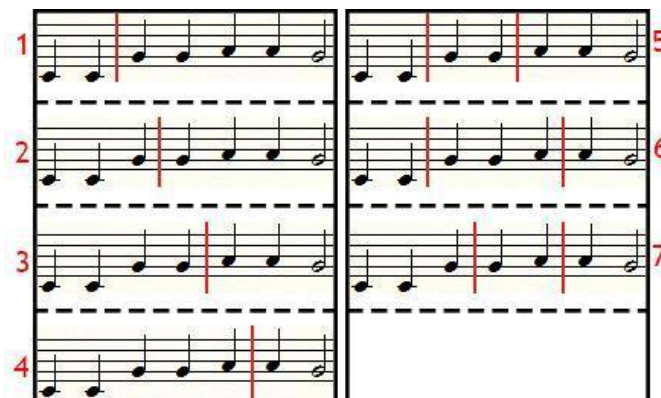
4. FRAGMENTACIÓN DE LA PIEZA MUSICAL

4.1. POSIBLES FRAGMENTACIONES

Según la metodología propuesta, el primer proceso aplicado a la *pieza musical* consiste en fragmentarla en todas las combinaciones posibles, donde cada una de estas, tiene una configuración de cortes única.

La iteración sobre las posibles fragmentaciones candidatas, categoriza al actual proceso como la solución a un problema de búsqueda, que se puede realizar inicialmente de forma exhaustiva como se muestra en el ejemplo de la Figura 13.

Figura 13. Ejemplo de todas las posibles fragmentaciones para una pieza corta.



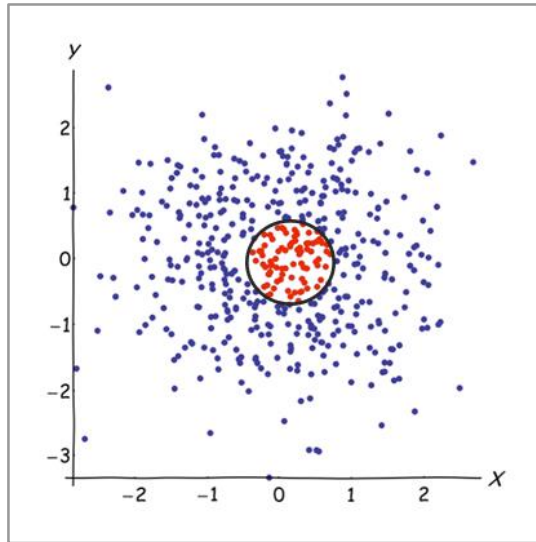
En cada una de las posibles fragmentaciones, la línea vertical roja muestra la ubicación en la cual se propone(n) el(los) punto(s) de corte.

En esta pieza corta la cantidad de fragmentaciones posibles se reduce a 7, pero esta cantidad crece a medida que la longitud total de la pieza aumenta; con piezas de mayor tamaño, se pueden presentar problemas de explosión combinatorial elevando el tiempo de cálculo computacional. Este problema es característico de las aplicaciones que emplean la búsqueda exhaustiva (fuerza bruta), aunque se garantice la exploración del total de configuraciones de cortes posibles.

4.2. REDUCCIÓN DEL ESPACIO DE BÚSQUEDA

Con relación a la búsqueda exhaustiva anterior, se genera un problema de eficiencia que se puede resolver aplicando heurísticas que reduzcan de forma significativa el espacio explorado. Para una mayor comprensión de este concepto, se presenta en la Figura 14 un ejemplo gráfico de cómo una población es reducida aplicando alguna condición o discriminación.

Figura 14. Ejemplo gráfico de reducción de espacio de búsqueda.



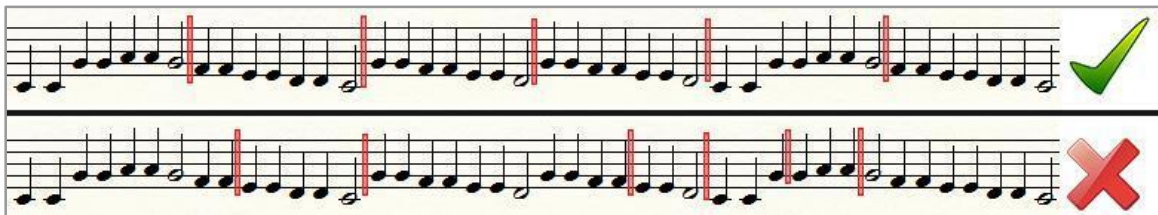
En la imagen se ejemplifican como puntos la totalidad de los elementos en un espacio de búsqueda supuesto, el cual es limitado por algún método para reducir su tamaño. Los puntos en rojo representan la población reducida resultante.

Para la reducción del espacio de exploración se plantearon los tres procesos descritos a continuación:

4.2.1. Limitación por cortes

Esta condición permite filtrar las posibles fragmentaciones rechazando aquellas que incluyan fragmentos de un tamaño inferior a un mínimo o superior a un máximo definidos.

Figura 15. Ejemplo para corte aceptado y rechazado por limitación de cortes.



Para el ejemplo de la Figura 15, se configuró una limitación de cortes entre 4 y 8 elementos. La fragmentación superior fue aceptada porque todos sus cortes forman fragmentos de 7 notas, cantidad incluida en el intervalo, pero la fragmentación inferior fue rechazada porque incluye fragmentos de 3, 9 y 11 notas, cantidades fuera del intervalo permitido.

Esta heurística no tiene un fundamento musical robusto dado que los tamaños de los elementos musicales representativos de una pieza musical no tienen una longitud definida a priori, por tal razón, su configuración inadecuada puede hacer que la exploración pierda calidad.

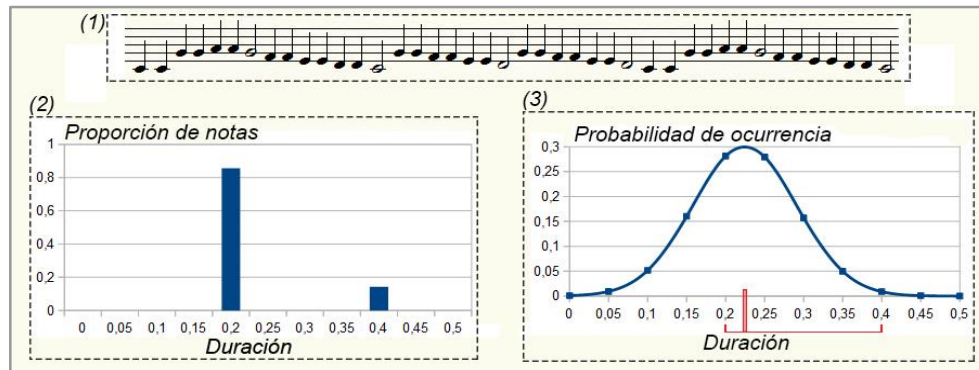
4.2.2. Identificación de Frases y Semifrases

Una heurística que tiene mayor soporte conceptual a partir de la percepción musical y por lo tanto mejores posibilidades de reducción del espacio de búsqueda con menor pérdida de calidad es la identificación de frases.

Debido a la alta relación que existe entre la percepción de ideas con sentido completo y la sensación rítmica de pausa entre ellas (esto proveniente del estudio de la música como lenguaje) podemos definir como frases a los fragmentos separados por dichos espacios de pausa (espacios de tiempo sin registro de ataque).

Una característica a tener en cuenta para obtener mejores resultados en la búsqueda de frases, se refiere a la relatividad existente entre las pausas y la duración más frecuente de la pieza musical, para esto se planteó el procedimiento ejemplificado en la Figura 16.

Figura 16. Histograma y frecuencia de duraciones para una melodía de ejemplo.



Se expone una melodía de ejemplo (1) acompañada de un histograma de sus duraciones (2) y una distribución normal que se infiere a partir de este (3).

El objetivo inicial fue encontrar el ritmo frecuente de la pieza, para ello se construyó un histograma de las duraciones entre ataques y se ajustó a una distribución normal. Esto presentó como resultado una media y una varianza inferida con la cual se construyó un filtro que permitió reconocer cuáles de los tiempos entre ataques de la pieza eran tan grandes como para concluir que son pausas.

Para tener mayor dominio del filtro discriminador de pausas se incluyó un parámetro que permitió controlar el umbral del corte.

Esta forma de abordar el problema permitió encontrar frases de manera relativa al contenido propio y sin conocer a priori la configuración rítmica de la pieza.

No obstante los resultados, se incluyen las siguientes salvedades que pueden afectar la calidad de la actual heurística:

- a. Algunas piezas musicales no tienen una clara discriminación rítmica de sus frases.
- b. Dependiendo de su contenido rítmico, algunas piezas musicales pueden tener una configuración donde se presenten dificultades en la percepción general de un ritmo frecuente.
- c. Por otro lado, el uso inadecuado del parámetro de umbral del filtro puede llevar también a la pérdida de calidad de la discriminación de las frases.

4.2.3. Omisión de fragmentos repetidos

En general, las piezas musicales hacen reutilización de fragmentos característicos. Con base en este precedente y aclarando que el modelo propuesto para la identificación de patrones no depende de la ubicación de los fragmentos, se pudo aplicar una heurística de omisión de repeticiones, la cual reduce el tamaño total de la pieza en una medida proporcional a la cantidad de repeticiones identificadas y por lo tanto, se reduce el espacio de búsqueda, sin afectar la calidad de la solución.

La heurística consiste en buscar e identificar dentro de los *Fragmentos* encontrados en un paso previo, aquellos que se presenten repetidos (o bien, con un alto grado de similitud); de esta forma, al generar los posibles cortes en profundidad, se aplicará una fragmentación simétrica a las repeticiones identificadas. En la Figura 17 se presenta un ejemplo para dar mayor claridad al proceso.

Figura 17. Ejemplo de omisión de fragmentos repetidos.

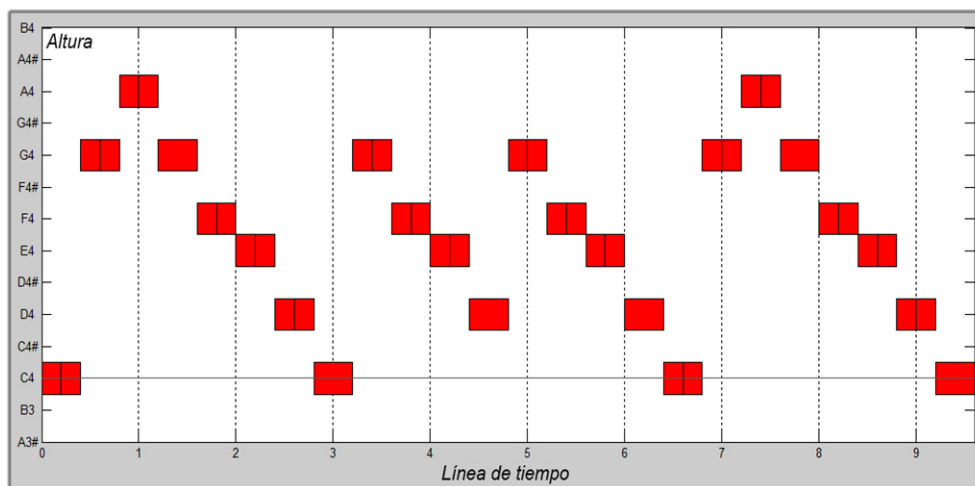
The diagram shows two staves of musical notation in bass clef, 6/8 time, with a key signature of one sharp (F#). The top staff contains four phrases: 'a' (quarter notes G, A, B), 'b' (quarter notes C, D, E), 'a' (quarter notes G, A, B), and 'b' (quarter notes C, D, E). A red arrow points from the top staff to the bottom staff. The bottom staff shows the same melody but with only two phrases: 'a' (quarter notes G, A, B) and 'b' (quarter notes C, D, E). The first 'a' phrase is highlighted with a green box, and the first 'b' phrase is highlighted with a blue box. The second 'a' phrase is also highlighted with a green box, and the second 'b' phrase is highlighted with a blue box. This illustrates the process of omitting repeated fragments to reduce the search space.

Esta figura expone en la partitura superior una pieza musical de ejemplo en la que se encuentran identificadas cuatro semifrases (abab); abajo se reduce la melodía analizada a dos segmentos (ab) presentando enseguida sus respectivas repeticiones. El tamaño de la pieza en análisis se reduce a la mitad (de 24 a 12) minimizando el espacio total de búsqueda de 256 (4^4) a 16 (4^2) posibilidades.

5. CARACTERIZACIÓN DE FRAGMENTOS

En este capítulo se profundizan los procesos relacionados con la representación computacional y la caracterización de los fragmentos musicales.

Figura 18. Representación numérica para la pieza de ejemplo.



En la representación gráfica para la pieza de ejemplo de la Figura 18, se observa el tiempo transcurrido en el eje X y el tono (altura) en el eje Y, con estas dos dimensiones se describen los eventos sonoros o notas musicales incluyendo para cada una de ellas su punto de ataque o tiempo de inicio y su punto de finalización con el cual se infiere la duración de dicho evento.

Para una mejor comprensión de la representación numérica, en la Figura 19 y la Tabla 2 se exponen la forma gráfica y numérica de las características iniciales para un fragmento ejemplo de 7 eventos; se les llaman *características iniciales* porque incluyen los atributos propios de los eventos contenidos en el fragmento, pero no tienen un significado que identifique como tal al fragmento musical.

Figura 19. Representación computacional de un fragmento ejemplo.



Tabla 2. Características iniciales del fragmento ejemplo.

Evento	Inicio	Duración	Altura
#1	0.00	0.20	C4
#2	0.20	0.20	C4
#3	0.40	0.20	G4
#4	0.60	0.20	G4
#5	0.80	0.20	A4
#6	1.00	0.20	A4
#7	1.20	0.20	G4

5.1. CARACTERÍSTICAS GENERADAS

A partir de las características iniciales y con el objetivo de extraer atributos que vislumbraran perspectivas funcionales del fragmento, se generaron nuevas características que agregaron valor a la representación y que incluyeron una diversidad de datos de tipo heterogéneo:

5.1.1. Características simples

Algunas de estas características generadas son de naturaleza numérica y por eso se les llama simples:

a. Rítmica

- Duración total: tiempo desde el inicio del primer evento hasta el fin del último.
- Densidad rítmica: cantidad de notas por unidad de tiempo o por un múltiplo de ésta.
- Proporción sonido/silencio: indicador de sonido efectivo del segmento contra espacios de ausencia de sonido.

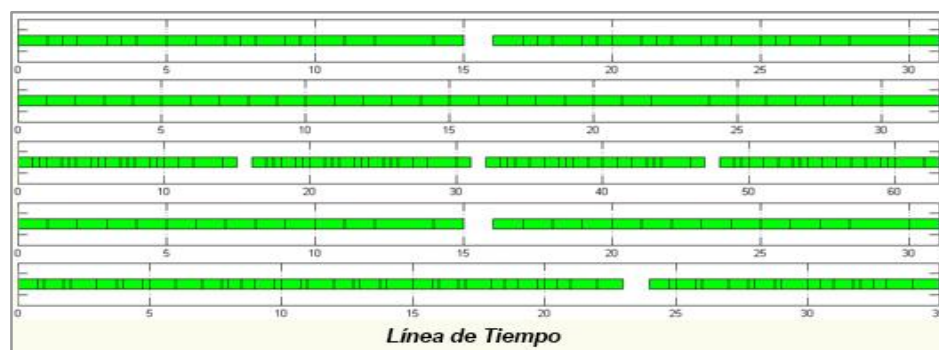
b. Armónica

- Índice potencial de tensión (disonancia): indicador de presencia de semitonos en el fragmento.

5.1.2. Características compuestas

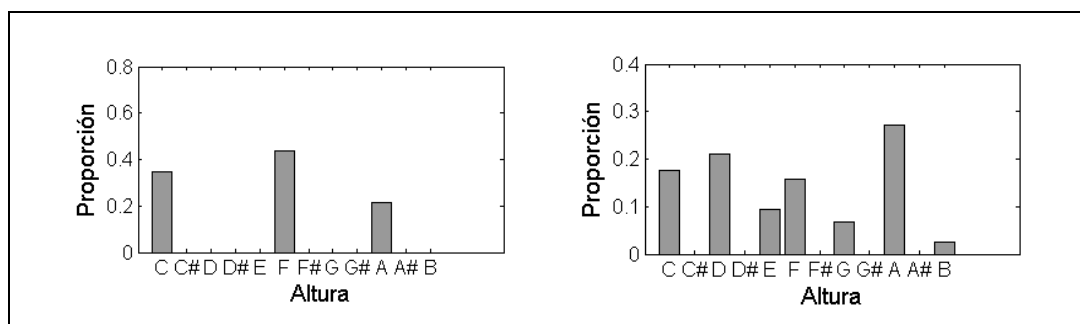
- a. Rítmica: Línea Rítmica - Cadena o secuencia de duraciones de los sonidos. En la Figura 20 se presentan ejemplos gráficos de líneas rítmicas.

Figura 20. Ejemplos de líneas rítmicas.



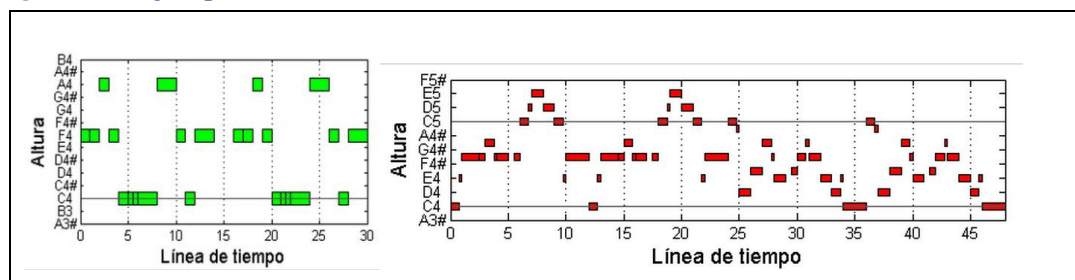
- b. Armónica: distribución de Alturas - Frecuencia para cada una de las 12 notas musicales (histograma). En la Figura 21 se observan ejemplos gráficos de distribuciones de alturas.

Figura 21. Ejemplos de distribuciones de alturas.



- c. Melódica: Línea Melódica - Cadena o secuencia de alturas sucesivas de los sonidos (sin considerar duraciones). Para mayor comprensión la ilustración de la Figura 22, presenta ejemplos gráficos de líneas melódicas.

Figura 22. Ejemplos de líneas melódicas.



5.1.3. Perspectiva

Adicionalmente, para las variables de cadena o secuencia fue posible realizar algunas transformaciones para permear la representación a problemas conocidos de traslación, expansión y compresión del ritmo y la melodía; estos problemas son producto de procesos aplicados a los fragmentos en el desarrollo creativo como parte de la variación motívica, como lo confirma el siguiente texto (el término *perspectiva* fue adaptado de la palabra inglesa *Viewport*).

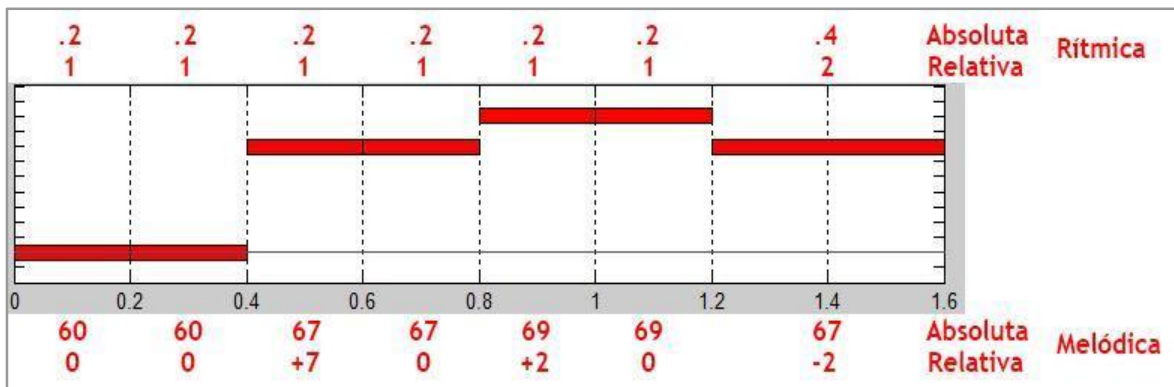
"Dada una idea melódica, un compositor puede usar varios procesos temáticos como: repetición, secuencia, fragmentación, extensión o la introducción de una idea contrastante." [5]

Las transformaciones propuestas fueron las siguientes:

- Línea rítmica relativa: Secuencia de duraciones referidas a la unidad definida (la más frecuente, la inicial, etc).
- Línea melódica relativa: Secuencia de cambios de altura con referencia a un tono dado (la más frecuente, la inicial, etc).

La Figura 23 muestra la forma numérica de la representación perspectiva rítmica y melódica para el fragmento de ejemplo.

Figura 23. Características perspectivas para el fragmento ejemplo.



Para la experiencia que se describe actualmente, se utilizaron las transformaciones relativas con referencia a la duración inicial para la línea rítmica y con referencia a la altura inicial para la línea melódica, como dos nuevas características generadas.

5.2. EJEMPLO

En la Tabla 3 y para mayor comprensión de la representación final, se expone el fragmento ejemplo con sus características generadas en detalle.

Tabla 3. Caracterización completa del fragmento ejemplo.

Característica	Valor
Duración Total	<i>1.6</i>
Densidad Rítmica	<i>$7/1.6 = 4.375$</i>
Proporción Sonido/Silencio	<i>$suma(duraciones)/1.6 = 1$</i>
Línea Rítmica	<i>{.2, .2, .2, .2, .2, .2, .4}</i>
Línea Rítmica Relativa	<i>{1, 1, 1, 1, 1, 1, 2}</i>
Distribución de Alturas	<i>{2, 0, 0, 0, 0, 0, 0, 3, 0, 2, 0, 0}</i>
Índice Potencial de Tensión	<i>0</i>
Línea Melódica	<i>{60, 60, 67, 67, 69, 69, 67}</i>
Línea Melódica Relativa	<i>{0, 0, 7, 0, 2, 0, -2}</i>

Características finales para el ejemplo mostrado en la Figura 19.

6. AGRUPAMIENTO Y SELECCIÓN

Tomando como referencia los dos capítulos anteriores, a continuación se expone, cómo, por medio de la exploración sobre el espacio de búsqueda generado en la fragmentación de la pieza musical y utilizando la caracterización de los fragmentos para cada una de las fragmentaciones posibles, se puede encontrar alguna que represente la mejor forma de particionar la obra evidenciando rasgos estructuralmente importantes.

"La similitud y el contraste son fundamentales en la organización formal" [5]

Teniendo en cuenta el texto citado, se puede definir que una organización formal coherente se encuentra utilizando elementos de comparación (similitud/contraste) entre las características propias de los fragmentos de una misma partición.

Esto en secuencia, introduce la posibilidad de calcular un índice de "organización formal" para cada fragmentación, permitiendo así comparar las diferentes fragmentaciones para seleccionar aquella cuyo índice sea el mejor, remitiéndose todo el problema a una solución soportada en la teoría de optimización.

6.1. AGRUPAMIENTO

"En muchas composiciones la forma o el diseño general es novedoso y escapa a la clasificación estándar, aun así tienen un diseño formal que se puede discernir fragmentando en segmentos de longitud variable y agrupándolos en diseños formales reconocibles. (...) Cualquier composición tendrá un diseño formal (...) La estructura formal proviene de la manera específica en que se generó el contenido de la composición." [5]

Las premisas teóricas anteriores son el soporte musical para afirmar que con la técnica de agrupamiento no supervisado se pueden identificar aún estructuras formales de piezas musicales novedosas.

6.1.1. Comparación entre fragmentos

La base fundamental para la aplicación de las técnicas de agrupamiento es la comparación entre fragmentos.

Dentro de las técnicas ofrecidas por los métodos de Kernel se encuentra la matriz de distancias, espacio al cual se pueden inducir distancias entre elementos heterogéneos que en otras condiciones no podrían compararse. Para calcular las distancias entre pares de elementos se utilizan funciones de comparación, en seguida se presentan aquellas que se utilizaron en las características de los fragmentos.

- a. Resta aritmética: para las características de *tiempo total*, *densidad rítmica*, *proporción sonido/silencio* e *índice potencial de tensión* se puede utilizar el valor absoluto de la resta aritmética ya que son variables numéricas simples.
- b. Distancia de edición: para las secuencias numéricas de *línea rítmica* y *melódica* (también para las relativas) se definió que se podía utilizar algún algoritmo de diferencia entre secuencias, el cual permitía dar un valor real que representaba la distancia entre el par de cadenas; se planteó utilizar distancia de Levenshtein⁴.
- c. Resta de histogramas: para la distribución de alturas se propuso utilizar un algoritmo que tomara el par de histogramas y calculara la diferencia en cada una de las 12 posiciones de conteo, luego debía tomar el promedio de estas 12 diferencias y así obtendría un valor único que representara la distancia entre los fragmentos desde esta característica.

6.1.2. Esquema de agrupamiento

Terminada la caracterización de los fragmentos y su comparación en masa para llevarlos a una matriz de Kernel, se les aplicó una técnica de agrupamiento por cada una de las características identificadas, esto arrojó como resultado un esquema de agrupamiento.

6.1.3. Caracterización del agrupamiento

De la misma forma como se caracterizó un fragmento musical, se puede caracterizar cualquier otro elemento, por supuesto, utilizando valores propios de su función; de esta manera, para un esquema de agrupamiento se pueden calcular las siguientes características:

- a. Similitud intra-grupo: inverso de la distancia intra-grupo (distancia promedio entre elementos del mismo grupo), favorece los grupos con elementos más cercanos. En seguida se presenta la ecuación para este índice:

$$S_i = \left(1 - \frac{1}{m} \sum_{j=2}^{n_i} \sum_{r=1}^{j-1} |X_r - X_j| \right) \quad (1)$$

Ecuación 1. El índice se calcula para cada uno de los i grupos identificados, donde j y r son las variables de iteración dentro de los n elementos de X (conjunto de fragmentos comparados). Finalmente, el número de comparaciones posibles entre fragmentos es representado por la variable m .

- b. Índice de agrupamiento: proporción de la pieza que pertenece a algún grupo, favorece agrupamientos con mayor número de elementos agrupados (menor cantidad de fragmentos huérfanos).

⁴ Se implementó una versión propia del algoritmo de Levenshtein en la cual se da un valor de edición relativo a la diferencia de altura o distancia rítmica según el caso. La implementación del algoritmo se encuentra en el CD anexo.

$$Lg = \frac{\sum^C L_x}{\sum^T L_x} \quad (2)$$

Ecuación 2. El índice se calcula para la totalidad de la pieza musical, donde C representa el conjunto de fragmentos agrupados (Cluster) y T el conjunto total de fragmentos. L representa la longitud del elemento evaluado (x).

- c. Índice de generalización: indicador de agrupación por fragmentos mayores, favorece grupos formados por fragmentos de mayor longitud.

$$G_i = \frac{1}{N_i} \sum_{i=1}^{n_i} L_i \quad (3)$$

Ecuación 3. Donde N representa la cantidad de fragmentos en el grupo i . L la longitud de cada elemento.

6.2. SELECCIÓN

6.2.1. Mejor agrupamiento

"Una base tradicional para clasificar los tipos de diseño musical es por el número y la relación de las secciones que se presentan en una composición." [5]

El enunciado anterior es el soporte teórico para definir que los índices expuestos para la caracterización del agrupamiento debían tomar sus mayores valores en conjunto, con los grupos que mejor se comportaron como representantes de un diseño formal coherente, al evaluar características de cantidad de elementos agrupados, mejores relaciones y un nivel de generalización mayor; por lo tanto, la suma ponderada de estos índices reveló a los mejores exponentes, como se sintetiza en la siguiente ecuación.

$$CTI = \frac{k_1}{n_c} \sum_{i=1}^{n_c} S_i + k_2 Lg + \frac{2k_3}{n_c \sum^T L_x} \sum_{i=1}^{n_c} G_i \quad (4)$$

Ecuación 4. El índice totalizador, es una ponderación con las constantes k de los índices parciales (Ecuaciones 1, 2 y 3) y la cantidad de grupos representada por la variable n_c .

6.2.2. Maximización

Finalmente, para encontrar la fragmentación con mejor índice de *organización formal*, se aplicó un proceso simple de maximización del índice totalizado de agrupamiento que consistió en realizar un recorrido sobre todas las fragmentaciones candidatas, seleccionando el mejor representante de la población explorada.

7. IMPLEMENTACIÓN

Se desarrolló una herramienta de software que sigue los lineamientos de la metodología propuesta y que corresponde al proceso de identificación estructural para piezas musicales. Dicha pieza de software fue desarrollada bajo el entorno de programación *Eclipse* y utilizando la versión 6 del lenguaje de programación *Java*, aprovechando las facilidades de implementación que ofrecen los lenguajes orientados a objetos⁵.

7.1. MODELO DE REPRESENTACIÓN

Dentro de los componentes incluidos en el software se encuentra un modelo de representación que cubre varios aspectos importantes en el proceso en cuestión y que corresponden directamente a los pasos del proceso:

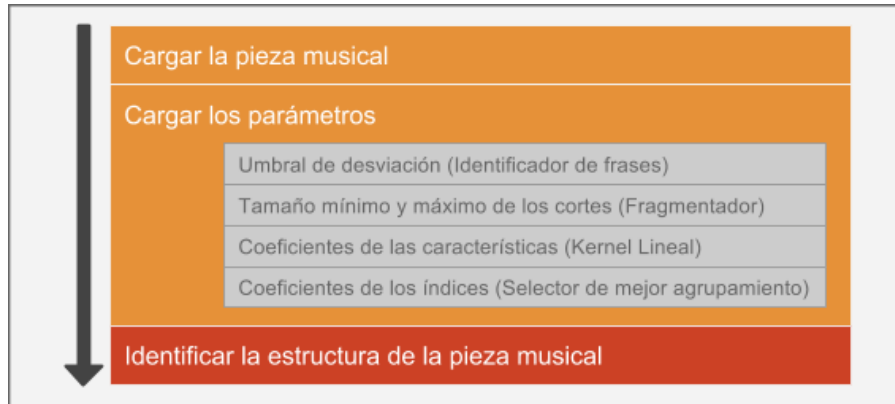
- a. Musical: incluye los elementos de primer nivel para la representación musical: *pieza* y *nota musical*.
- b. Fragmentación: elementos necesarios para representar las posibles configuraciones de cortes en la pieza musical: *corte*, *fragmento* y *fragmentación*. En este punto es importante recordar que un *fragmento*, al igual que una pieza musical, es una secuencia de *notas*.
- c. Representación vectorial: define las características de los fragmentos abstraídas de la relación entre sus notas musicales contenidas: *duración total*, *densidad rítmica*, *proporción sonido/silencio*, *línea rítmica*, *distribución tonal*, *índice potencial de tensión* y *línea melódica*.
- d. Agrupamiento: la *matriz de distancia* es el elemento necesario para el proceso de agrupamiento automático, la cual representa las distancias medidas entre todos los fragmentos en observación. La tarea de agrupamiento da como resultado un *grupo jerarquizado* que representa la organización abstraída a partir de los fragmentos.
- e. Selección: el *índice de agrupamiento* es el elemento primordial para el proceso de selección, este incluye las mediciones de *similitud intra-grupo*, *similitud inter-grupo*, *índice de grupos* e *índice de generalización* de un agrupamiento específico (expuestas en el capítulo anterior).

7.2. ALGORITMOS

Para el desarrollo de los algoritmos se siguieron las recomendaciones enunciadas en la programación dinámica, esto para garantizar la reducción del tiempo de ejecución y un mejor uso de los recursos computacionales. El esquema del algoritmo general implementado se encuentra descrito en la Figura 24.

⁵ Ver código fuente en el CD anexo

Figura 24. Esquema del algoritmo general.



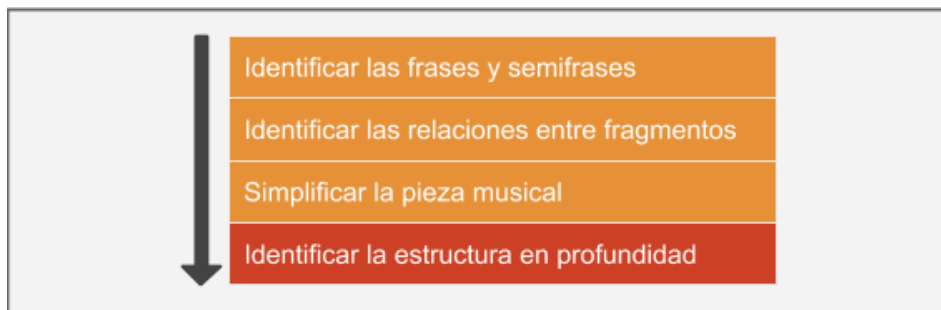
Enseguida se ofrece una descripción de los parámetros mencionados en el algoritmo.

- Umbral de desviación: es un valor decimal y positivo utilizado por el identificador de frases y semifrases.
- Tamaño mínimo y máximo de los cortes: son dos valores enteros positivos que se utilizaron en el fragmentador para construir las fragmentaciones candidatas.
- Coeficientes de las características: son constantes de ponderación para el Kernel Lineal, este Kernel es utilizado como característica de los fragmentos.
- Coeficientes de los índices: son constantes de ponderación para el cálculo del Índice de Agrupamiento utilizado en el algoritmo de selección de mejor fragmentación.

Como primera instancia se cargó la pieza musical y los parámetros, los cuales son enviados a la funcionalidad de identificación estructural de la pieza, expuesta en la Figura 25.

7.2.1. Identificación estructural

Figura 25. Esquema del algoritmo de identificación estructural.



Esta funcionalidad inicia con la identificación de las frases y semifrases, para ello se escribió un algoritmo que responde a la lógica expuesta en el inciso “Identificación de Frases y Semifrases” del Capítulo 4, recibiendo como entrada la *pieza musical* y el parámetro *Umbral de desviación* y entregando como resultado los puntos de la pieza donde se detectaron pausas.

Cortando la pieza musical en dichos puntos de pausa se obtuvo una secuencia de fragmentos que representan un primer nivel de alta jerarquía. Con estos fragmentos se aplicó un algoritmo de identificación de relaciones que se explica enseguida.

- Algoritmo de identificación de relaciones

Esta funcionalidad hace uso de la representación vectorial de los fragmentos y entrega como resultado un árbol de agrupamiento jerárquico para cada una de las características en análisis. Adicionalmente, cada árbol tiene señalado el punto en que se encuentra la máxima relacionalidad. Este algoritmo también entrega un esquema único de agrupamiento superior, obtenido de la relación existente entre los agrupamientos menores.

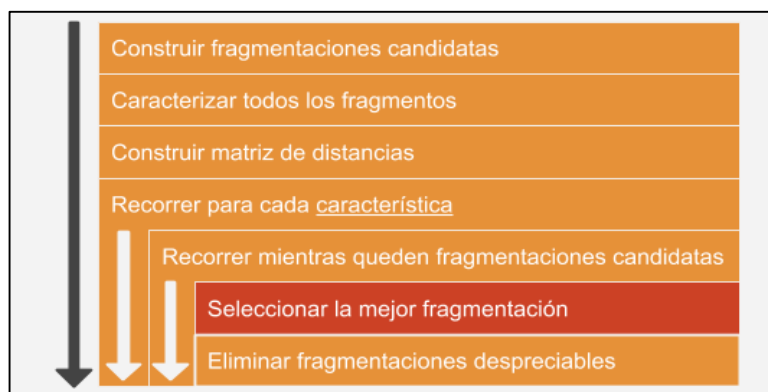
Para el procesamiento e identificación de los agrupamientos jerárquicos se utilizó la librería LingPipe [31] que cuenta con una versión de uso libre.

A continuación se simplificó la pieza musical, descartando para los procesos siguientes las repeticiones de fragmentos como se sugiere en el Capítulo 4 en pro de la reducción del espacio de búsqueda.

Es en este momento cuando se lanza el proceso de identificación estructural en profundidad, presentado gráficamente en la Figura 26.

7.2.2. Identificación estructural en profundidad

Figura 26. Esquema del algoritmo de identificación estructural en profundidad.



Para empezar, se construyeron todas las posibles fragmentaciones que serían candidatas en el posterior proceso de selección. Esta labor se realizó utilizando un algoritmo que recibe como entrada la pieza musical ya particionada en frases y los parámetros de mínimo y máximo tamaño de cortes. El algoritmo realiza un proceso combinatorio para encontrar todas las posibles formas de fragmentar la pieza musical, respetando los cortes de frase previos.

Todos los posibles fragmentos encontrados en el proceso anterior son caracterizados y llevados a la representación definida en el modelo; para ello se implementó una rutina de cálculo que recibe como entrada el fragmento y entregan en la salida la representación vectorial. Esta rutina hace uso interno de sub-rutinas que realizan los cálculos parciales para cada característica.

Teniendo la representación vectorial de todos los fragmentos, y por medio de unas rutinas de cálculo de distancia entre pares, se construyó una matriz en la que se esquematizó la relación de distancia entre todos los fragmentos. Tal como lo sugiere la programación dinámica, solo se realiza el cálculo de las distancias entre fragmentos incluidos en la misma fragmentación candidata, las demás distancias se pueden despreciar. Por lo tanto se obtuvo una matriz donde solo algunas de sus entradas tienen valores y las demás se encuentran vacías.

Para el cálculo de distancia en variables compuestas, fue necesario utilizar rutinas especiales descritas a continuación:

- Distribución de alturas: la distancia para esta variable se calculó empleando un promedio de las restas parciales en cada uno de los 12 valores de frecuencia.
- Distancia de Levenshtein: para el algoritmo utilizado en el cálculo de la distancia entre variables de tipo cadena se tuvieron que realizar algunas modificaciones en pro de su ajuste a un mejor funcionamiento sobre variables musicales dado que en este caso, la distancia entre dos líneas rítmicas o melódicas, definida especialmente en términos numéricos (altura de las notas o duración de los eventos) no solo se ve afectada por la cantidad de ediciones necesarias para transformar una cadena en la otra (como lo hace el algoritmo clásico), sino también por la distancia abarcada por los elementos en edición. Fue necesario escribir un nuevo algoritmo que implementara explícitamente esta medición.
- Kernel lineal: utilizando las características del par de fragmentos y los coeficientes ingresados como parámetro al algoritmo general, se procedió a calcular esta distancia complementaria. Esta medida se puede definir como el producto punto de las representaciones vectoriales de dos fragmentos.

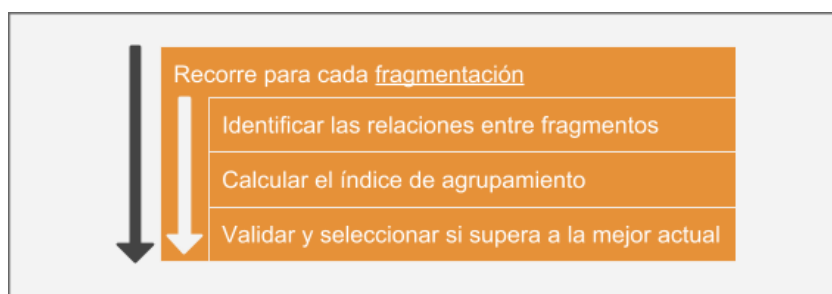
Es importante aclarar que la matriz calculada tiene tres dimensiones, donde las filas y columnas representan la totalidad de fragmentos en análisis, la tercera dimensión representa la característica del fragmento en comparación y las entradas (los valores de la matriz) representan la distancia calculada entre cada par de fragmentos para una característica específica.

Esta particularidad presente en la matriz abre un espacio de análisis, permitiendo determinar para dos fragmentos que, por ejemplo, se encuentran cerca en alguna de sus características, pero lejos en otra.

Por tal razón, el paso siguiente en el proceso fue recorrer de forma independiente cada una de las características de la representación vectorial.

Ya dentro de cada característica en específico, y mientras queden fragmentaciones candidatas por analizar, se realizó el proceso de selección de la mejor fragmentación, mostrado en la Figura 27.

Figura 27. Esquema del algoritmo de selección de mejor fragmentación.



Se recorrió entonces el conjunto de fragmentaciones candidatas y para cada una de ellas se identificó la relación entre fragmentos utilizando el mismo algoritmo que se expuso con anterioridad para el nivel de mayor jerarquía, la diferencia en este caso es que para el esquema de agrupamiento resultante y con ayuda de otra rutina programada, se calculó el índice de agrupamiento, siguiendo los lineamientos enunciados en el Capítulo 6.

Teniendo esta medida concreta para las diferentes fragmentaciones candidatas, se aplicó una rutina que entregó como resultado la fragmentación que presentó un agrupamiento de mejores condiciones.

Finalmente y para continuar con la identificación en profundidad, se eliminaron de las fragmentaciones candidatas aquellas cuyos cortes ya no correspondían con los incluidos en la fragmentación seleccionada. Esto permitió en una nueva ejecución del ciclo de selección, encontrar nuevas divisiones de los fragmentos en los niveles superiores.

7.3. FORMATOS

La herramienta desarrollada puede identificar la estructura formal para piezas musicales tanto en formato *Song* como en formato *Kern*.

Song (.song) es un formato propio que se utilizó inicialmente para almacenar la información musical abstraída de las canciones en bibliotecas MIDI cuando aún no se había definido la

realización de las pruebas con archivos *Kern*. Los archivos *Song* contienen una secuencia de líneas en un archivo plano, donde cada línea corresponde a los valores altura, tiempo de ejecución y duración de una nota musical, separados por punto y coma (;).

Para los archivos *Kern*, fue necesario programar en *Java* algunas funcionalidades que interpretarían el contenido musical y la marcación de frases, ya que las aplicaciones publicadas abiertamente por los creadores del formato y por otros grupos en la comunidad investigativa fueron desarrolladas para otros lenguajes o no presentaban la información de la manera como la herramienta desarrollada en esta investigación pudiera utilizarla.

Tabla 4. Parámetros de configuración de la prueba.

Parámetro	Valor
Umbral de desviación	1,4
Tamaño mínimo de bloque	4
Tamaño máximo de bloque	7
Coefficientes de las características	
Tiempo total	0,01
Densidad rítmica	0,09
Proporción sonido/silencio	0,01
Índice potencial de tensión	0,09
Línea rítmica	0,3
Línea melódica	0,3
Distribución de alturas	0,2
Coefficientes de los índices	
Similitud intra-grupo	0,25
Distancia inter-grupo	0,25
Agrupamiento	0,25
Generalización	0,25

Configuración utilizada para la identificación estructural del caso de ejemplo.

La elección de los valores para los coeficientes de las características fue fundamentada en el grado de importancia que estas representan en los materiales desarrollados en la cultura occidental donde prima la elaboración ritmo-melódica y la tonalidad, todo esto documentado en la teoría musical [5, 6]. Estos coeficientes se utilizaron para calcular el *Kernel lineal*, el cual se comparó con los resultados de un Kernel adaptativo y los resultados del análisis por cada característica, para demostrar la riqueza e importancia del análisis en múltiples dimensiones independientes.

El cambio en los coeficientes de los índices puede afectar ampliamente los resultados. En este caso se distribuyeron de forma homogénea para obtener respuestas equilibradas.

8.2. FRASES Y SEMIFRASES

El algoritmo de identificación de fragmentos de mayor nivel de jerarquía, entrega como resultado la fragmentación expuesta en la Figura 29, donde se puede observar la relación entre los puntos de corte y las pausas de la melodía. Esta fragmentación coincide completamente con las marcaciones de frases/semifrases señaladas por un músico experto.

Figura 29. Fragmentación mayor del caso ejemplo.

The image displays a musical score in treble clef with a key signature of one sharp (F#) and a 6/8 time signature. The melody is divided into eight numbered segments (1 through 8) by vertical red lines. Segment 1 starts at the beginning and ends at the first red line. Segment 2 starts at the second red line and ends at the third. Segment 3 starts at the fourth red line and ends at the fifth. Segment 4 starts at the sixth red line and ends at the seventh. Segment 5 starts at the eighth red line and ends at the ninth. Segment 6 starts at the tenth red line and ends at the eleventh. Segment 7 starts at the twelfth red line and ends at the thirteenth. Segment 8 starts at the fourteenth red line and ends at the end of the piece. The segments are numbered 1 through 8 in red text above the staff.

Fragmentación resultante de la identificación de fragmentos de mayor nivel de jerarquía (en este caso semifrases). Los fragmentos se encuentran enumerados y separados por las líneas verticales en color rojo.

8.3. RELACIONES ENTRE FRAGMENTOS

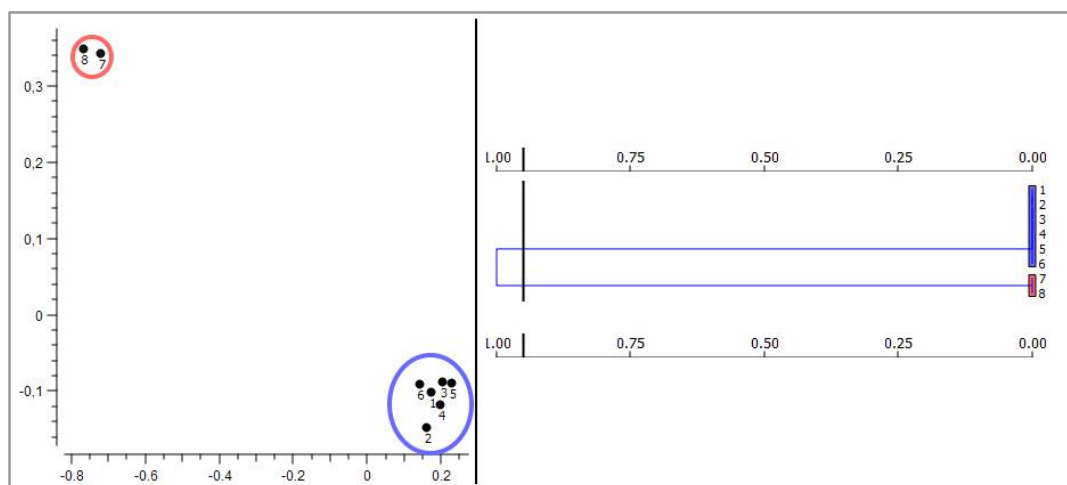
Para una mejor comprensión de las relaciones y resultados expuestos en seguida, los fragmentos de la Figura 29 han sido numerados.

Para el análisis de cada una de las características en cuestión, se presentarán los siguientes modos de visualización, según convenga:

- Visor MDS: representación gráfica en dos dimensiones que permite visualizar distancias en proporción. Las medidas de este tipo de visualización son inferidas a partir de las distancias entre pares en el espacio dimensional original de la característica evaluada y reducida a una representación de distancia bidimensional que solo tiene un significado representativo.
- Visor Jerárquico: representación gráfica en árbol que permite visualizar la familiaridad por cercanía entre fragmentos, también incluye una línea de corte que indica el mejor particionamiento por grupos identificado en el recorrido del árbol. El eje horizontal se conoce como *distancia de agrupamiento* y se refiere a la medida de relación entre los grupos identificados.

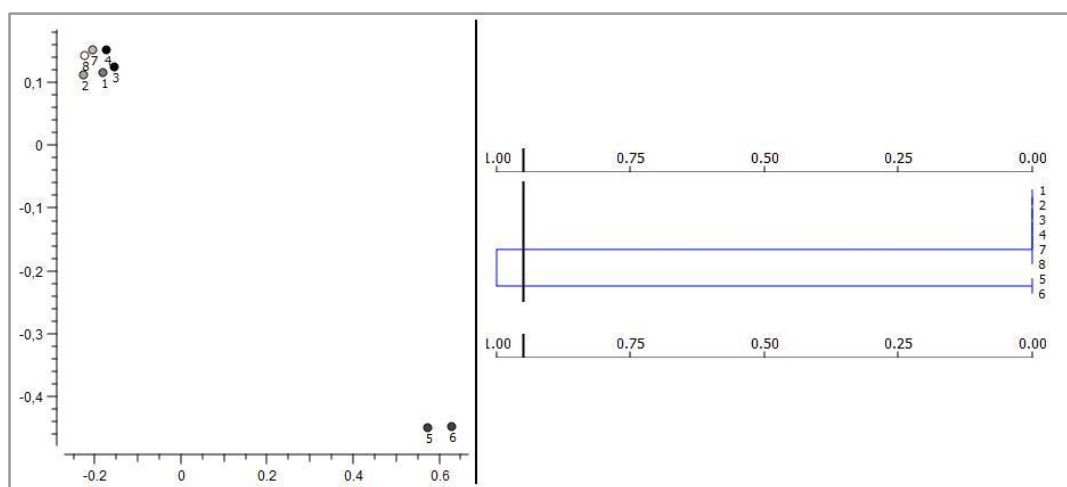
- c. Densidad rítmica: para esta característica ya encontramos dos grupos claramente diferenciados {1, 2, 3, 4, 5, 6} y {7, 8}. La razón es que los fragmentos 7 y 8, a diferencia de los demás, finalizan en negra con puntillo; esto les da la particularidad de una menor actividad rítmica hacia su parte final.

Figura 32. Visor MSD y jerárquico para la densidad rítmica.



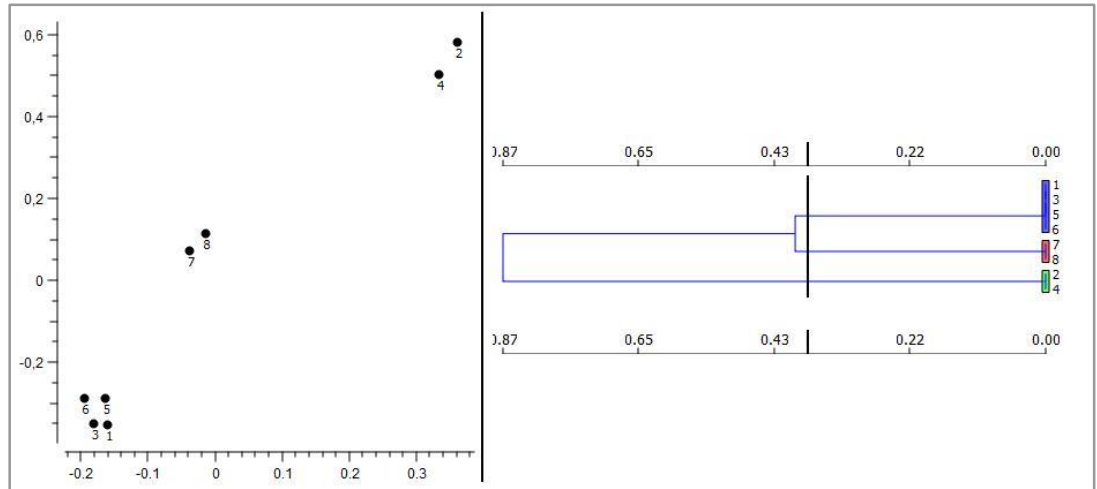
- d. Índice potencial de tensión: con esta característica sucede algo similar, dos grupos claramente diferenciados {1, 2, 3, 4, 7, 8} y {5, 6}, debido al uso de las notas Do# y Re en mayor parte del fragmento para el segundo grupo. Estas notas tienen una distancia de semitono entre ellas (la mínima posible), por tal razón, el grado de proximidad de las alturas del fragmento es elevado respecto a los demás. Algo importante para recalcar en este caso específico es el obvio agrupamiento de 5 y 6 porque son repeticiones.

Figura 33. Visor MDS y jerárquico para el Índice potencial de tensión.



- e. Línea rítmica: para esta característica, la formación de grupos empieza a lucir un poco más compleja. Se tienen dos grupos polarizados y diferenciados claramente, el primero {1, 3, 5, 6} (donde 3 y 6 son repeticiones), caracterizado por tener la misma idea rítmica finalizada en corchea-negra, y el otro {2, 4} (repetición) que presentan una idea contrastante. Adicionalmente, en el medio (un poco más cerca del primer grupo) aparece el grupo {7, 8} compartiendo una idea que se produce de una mezcla de elementos de las dos anteriores.

Figura 34. Visor MDS y jerárquico para la línea rítmica.



Para evidenciar en detalle los resultados encontrados en la línea rítmica, se señala a continuación en las figuras 35, 36 y 37 las relaciones presentes sobre la partitura.

Figura 35. Señalamiento sobre la partitura del primer grupo identificado.

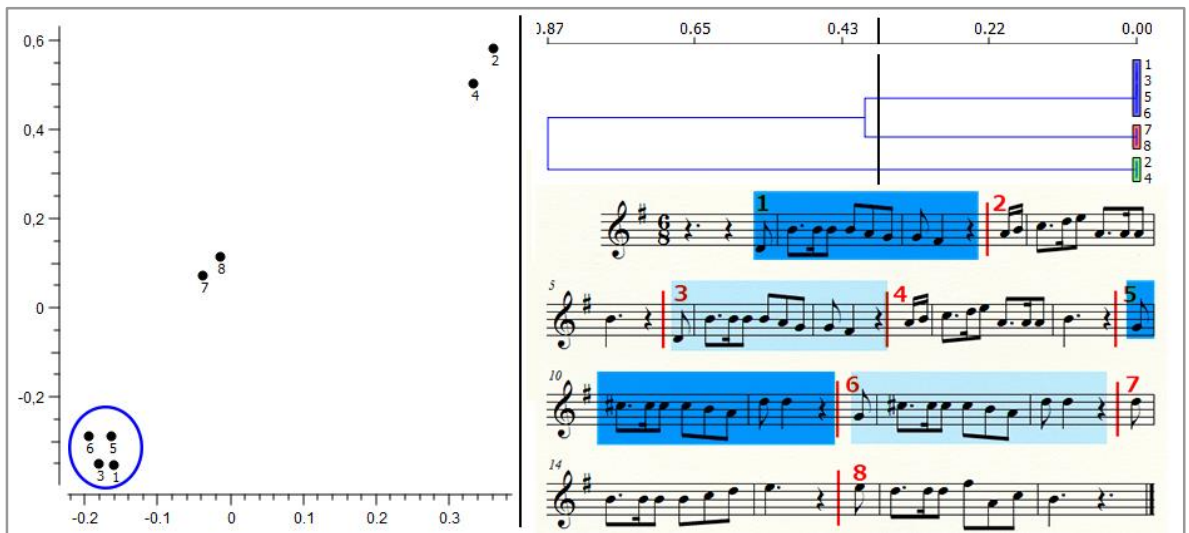


Figura 36. Señalamiento sobre la partitura del segundo grupo identificado.

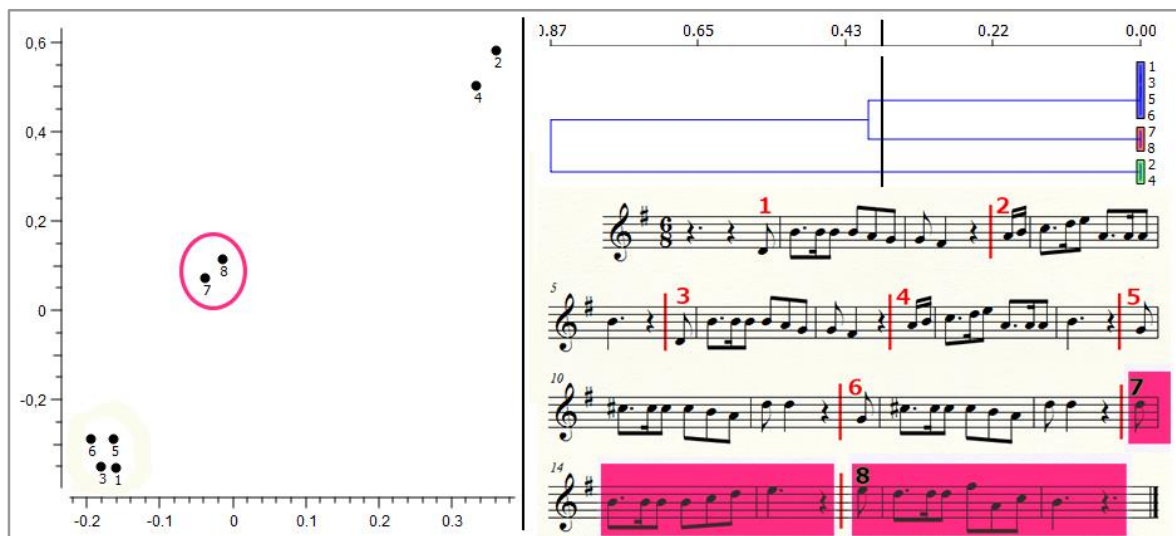
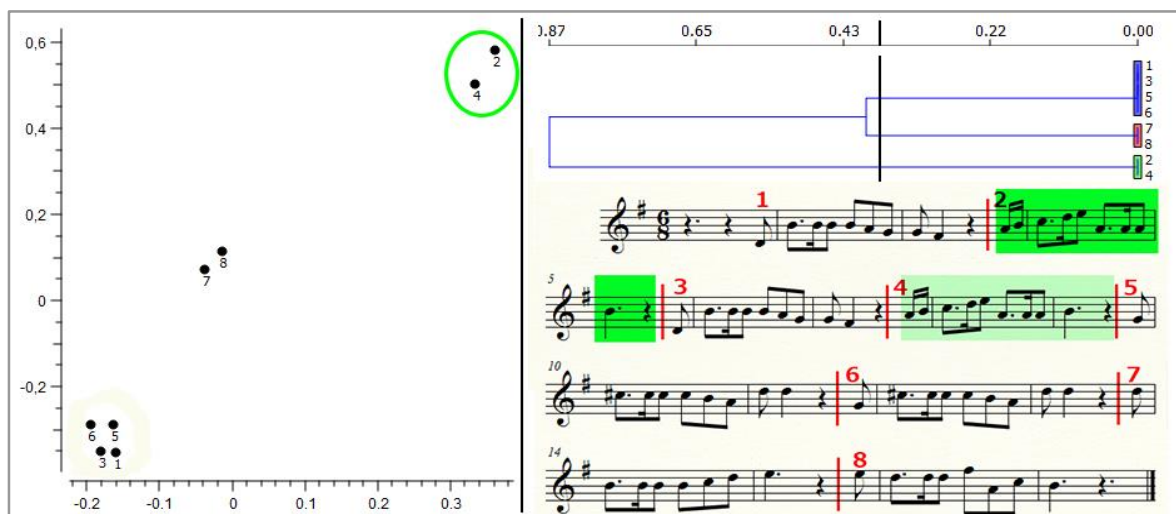
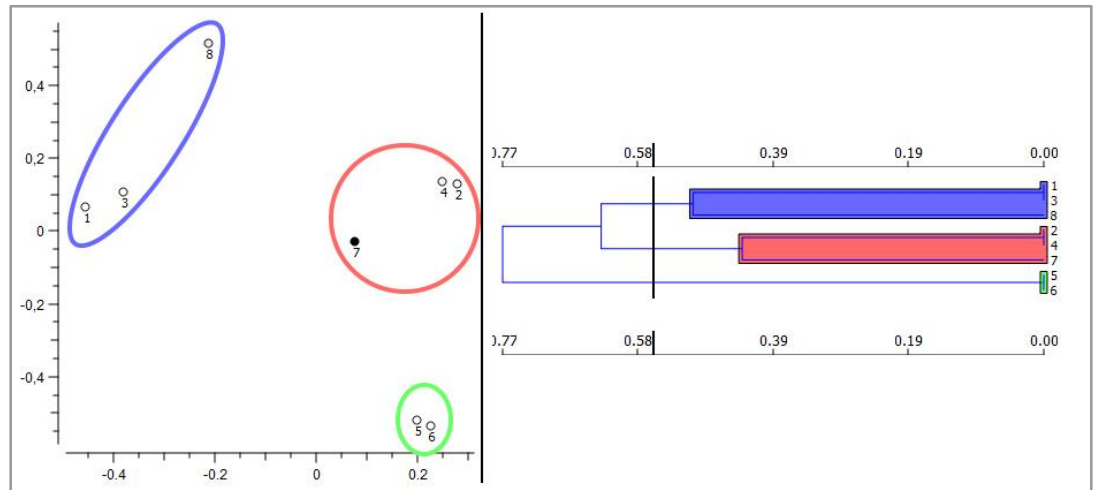


Figura 37. Señalamiento sobre la partitura del tercer grupo identificado.



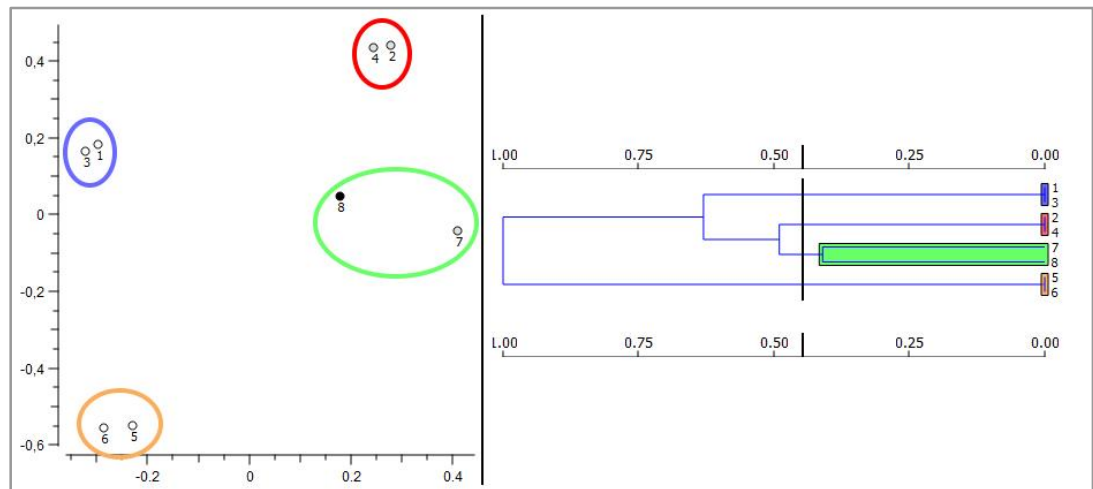
- f. Línea melódica: en esta característica la situación se vuelve aún más compleja pues se tiene una dispersión de grupos débilmente relacionados de los cuales se pueden concluir algunas ideas menores: 1) Se conforma un pseudo-grupo {2, 4, 7} cuyos fragmentos presentan un movimiento melódico predominantemente ascendente. 2) Se conforma otro pseudo-grupo aún más débil {1, 3, 8}, en cuyos fragmentos predomina un movimiento melódico descendente. 3) Si se observa con otra perspectiva el agrupamiento, dando pie a su capacidad para aislar grupos, entonces el grupo {5, 6} se destaca ya no por similitud sino por diferencia.

Figura 38. Visor MDS y jerárquico para la línea melódica.



- g. Distribución de alturas: nuevamente se observa una dispersión de grupos débilmente relacionados, pero con una característica especial que lo hace interesante; de nuevo el grupo {5, 6} se destaca por diferencia, esto por incluir en su material la nota Do# proveniente de una modulación temporal.

Figura 39. Visor MDS y jerárquico para la distribución de alturas.



Para entender la fortaleza de las relaciones encontradas, se definió una medida llamada *grado de familiaridad*, la cual suma la cantidad de características en las que unos fragmentos se encontraron relacionados. De esta manera y realizando un análisis minucioso de las

relaciones identificadas en cada característica se encontró que los fragmentos que comparten mayor familiaridad son:

- i. {1, 3}, {2, 4} y {5, 6} con 7 grados de familiaridad, lo cual es obvio, porque todas son repeticiones de su par.
- ii. {7, 8} con 6 grados de familiaridad y 1 de diferencia.

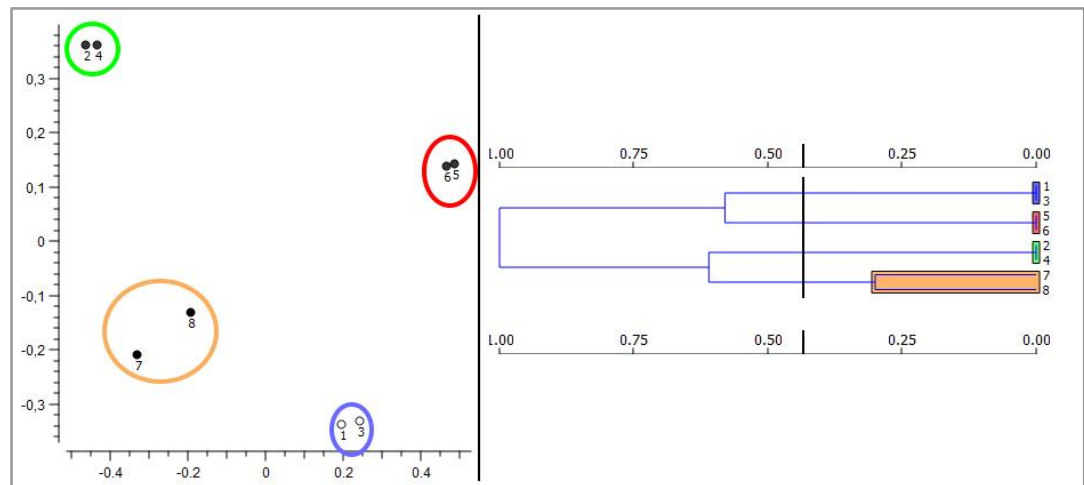
Con las relaciones descritas y asignando una letra a los grupos específicos se puede identificar la estructura general de semifrases:

- A: {1, 3}
- B: {2, 4}
- C: {5, 6}
- {D, D'}: {7, 8}

Utilizando la representación con letras y la posición de cada fragmento en la pieza musical, se obtiene: AB AB CC DD' que a un nivel de jerarquía mayor se puede entender como AABC. De esta manera se pudo dar por identificada la relación estructural de la pieza analizada.

- h. Kernel lineal: este nuevo espacio de comparación ofrece como resultado una relación general entre fragmentos, que de alguna forma describe la ponderación de las variables anteriormente presentadas, pero que pierde detalle a la hora de mencionar rasgos específicos y describir las relaciones.

Figura 40. Visor MDS y jerárquico para el Kernel lineal.



En todo caso, es importante destacar los rasgos descritos por el Kernel: 1) la relación {7, 8} apareciendo luego las relaciones de repetición, 2) las relaciones lejanas {{1, 3}, {5, 6}} y

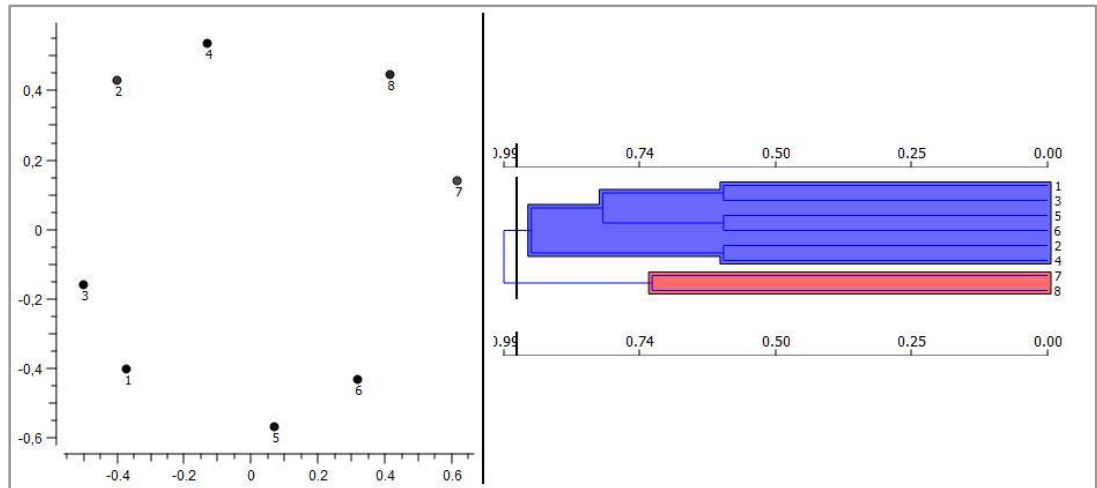
{{2, 4}, {7, 8}}. La desventaja resulta en que la descripción de dichas relaciones no es explícita en el resultado.

- i. Kernel con parametrización adaptativa: como parte de la experimentación, se realizaron pruebas con un espacio de comparación innovador que consiste en una ponderación de los parámetros del Kernel adaptado de forma automática según la importancia del agrupamiento identificado en cada característica. Este espacio fue propuesto con el fin de entregar como resultado un esquema de relación general, al igual que el Kernel lineal, pero mejorando su capacidad de adaptarse a la importancia particular de las características para la pieza musical analizada.

La parametrización adaptativa toma los valores de relación entre grupos para cada característica, identifica en cuáles de ellos se presentó una mejor relación (esto haciendo uso del *índice de agrupamiento* definido en el Capítulo 6) y aplica una compensación en la similitud de los elementos relacionados en dicha característica.

Los resultados del Kernel con parametrización adaptativa se observan en la Figura 41 donde se pueden apreciar a grandes rasgos los mismos grupos que en el Kernel lineal, pero algunas diferencias en cuanto a las jerarquías y el aislamiento de {7,8}.

Figura 41. Visor MDS y jerárquico para la parametrización adaptativa.



8.4. RELACIONES DE PROFUNDIDAD

Enseguida se muestran los resultados obtenidos con el algoritmo de identificación estructural en profundidad. Este proceso se inicia con una pieza musical simplificada en términos de las repeticiones.

En la Figura 42, se muestra la fracción de la pieza musical que continuará el proceso.

Figura 42. Pieza musical simplificada.

The image shows a musical score in G major and 6/8 time, consisting of four staves. The first staff contains measures 1 through 4. The second staff, starting at measure 5, contains measures 5 through 8, with the first two measures (5-6) and the last two measures (7-8) shaded gray. The third staff, starting at measure 10, contains measures 10 through 13, with the last two measures (12-13) shaded gray. The fourth staff, starting at measure 14, contains measures 14 through 17. Vertical red bars are placed at the beginning and end of each measure to indicate segment boundaries.

Se muestran sombreadas las repeticiones de la pieza que serán despreciadas. Los cortes que generan los fragmentos se muestran como barras verticales rojas sobre la partitura.

Para la presentación de resultados del algoritmo de identificación estructural en profundidad de este ejercicio en específico, se dividieron los fragmentos 1, 2, 5, 7 y 8 en un antecedente y un consecuente y se volvieron a enumerar, como se muestra en la partitura marcada del caso de ejemplo en la Figura 43.

Figura 43. Fragmentación en profundidad del caso ejemplo.

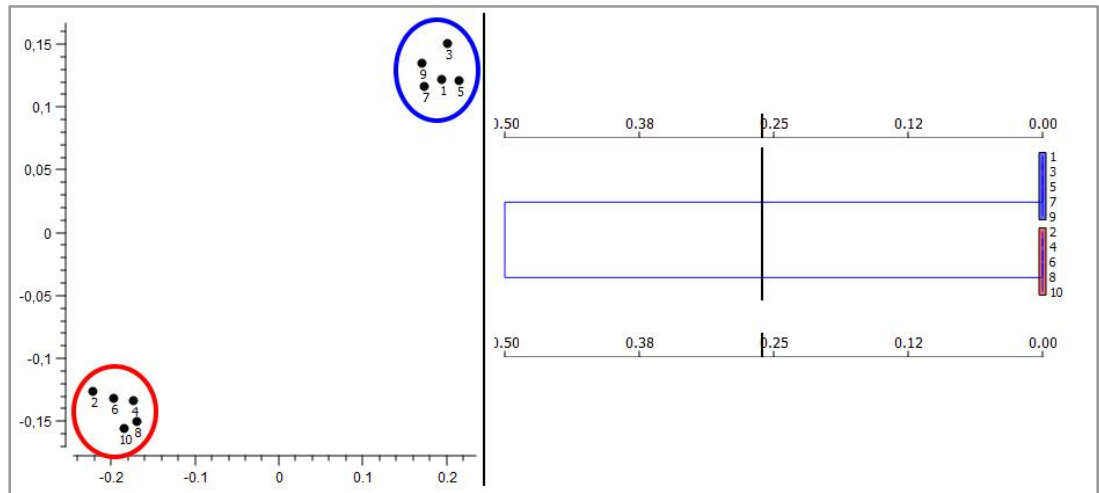
The image shows the same musical score as Figure 42, but with 10 numbered fragments (1-10) and shaded gray segments. The fragments are: 1 (measures 1-2), 2 (measures 3-4), 5 (measures 5-6), 7 (measures 7-8), 6 (measures 10-11), 8 (measures 12-13), 9 (measures 14-15), and 10 (measures 16-17). The shaded gray segments are: measures 5-6, measures 7-8, and measures 12-13. Red numbers 1-10 are placed above the notes to indicate the fragments.

Se definen la numeración para los nuevos fragmentos antecedente-consecuente.

Al ejecutar el algoritmo de identificación estructural en profundidad con los parámetros inicialmente definidos, los resultados son los siguientes:

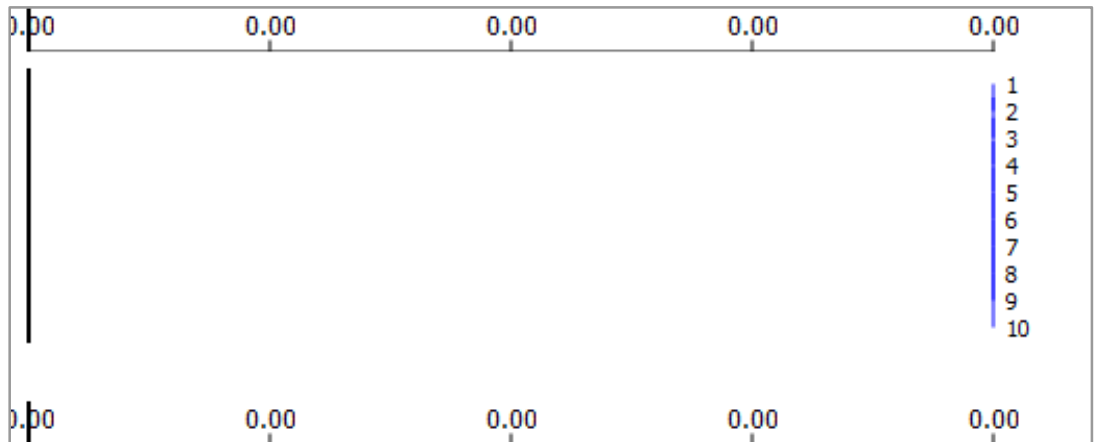
- a. Tiempo total: identifica claramente dos grupos, antecedentes {1, 3, 5, 7, 9} y consecuentes {2, 4, 6, 8, 10}, donde la duración de estos últimos es mayor.

Figura 44. Visor MDS y jerárquico para el tiempo total en profundidad.



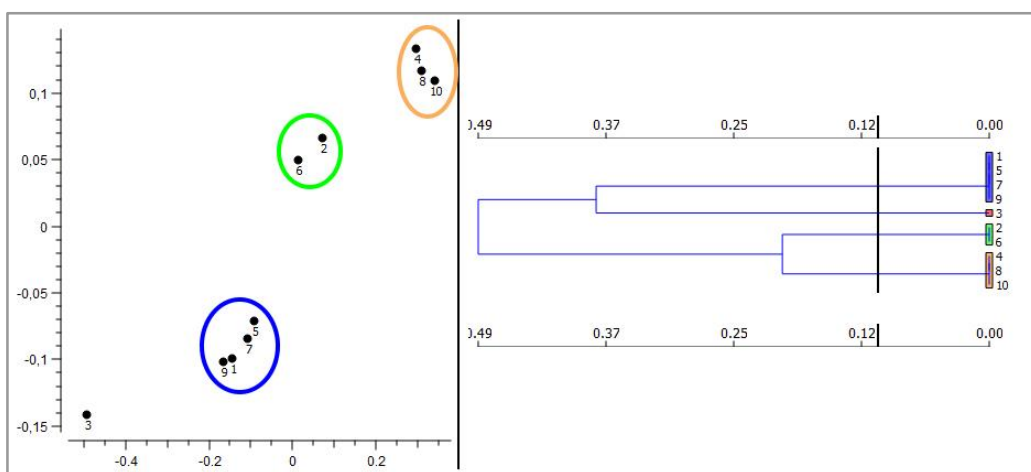
- b. Proporción sonido/silencio: como todos los fragmentos tienen un ritmo con notas continuas (sin la presencia de silencios dentro de la frase), esta característica percibe sólo un grupo {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}, no aporta en la discriminación de rasgos.

Figura 45. Visor jerárquico para la Proporción sonido/silencio en profundidad.



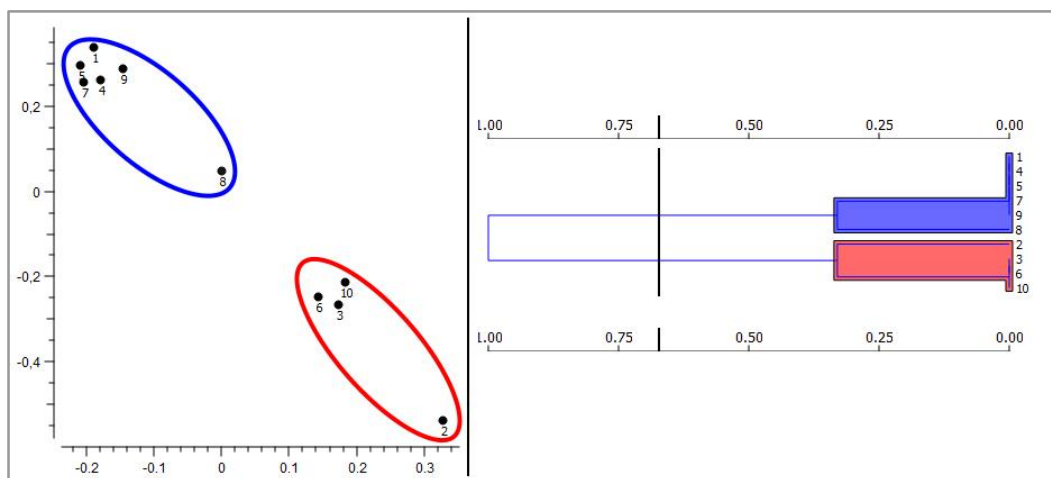
- c. Densidad rítmica: muestra dos grupos polarizados: {3} (idea aislada con la mayor actividad rítmica) y {4, 8, 10} (ideas con la menor actividad rítmica). A mitad de camino se completan los dos grupos, uno de antecedentes {{3}, {1, 5, 7, 9}} y otro de consecuentes {{2, 6}, {4, 8, 10}}, igual que en la característica Tiempo total, complementando la idea de la polaridad rítmica, pero de una proporción más débil.

Figura 46. Visor MSD y jerárquico para la densidad rítmica en profundidad.



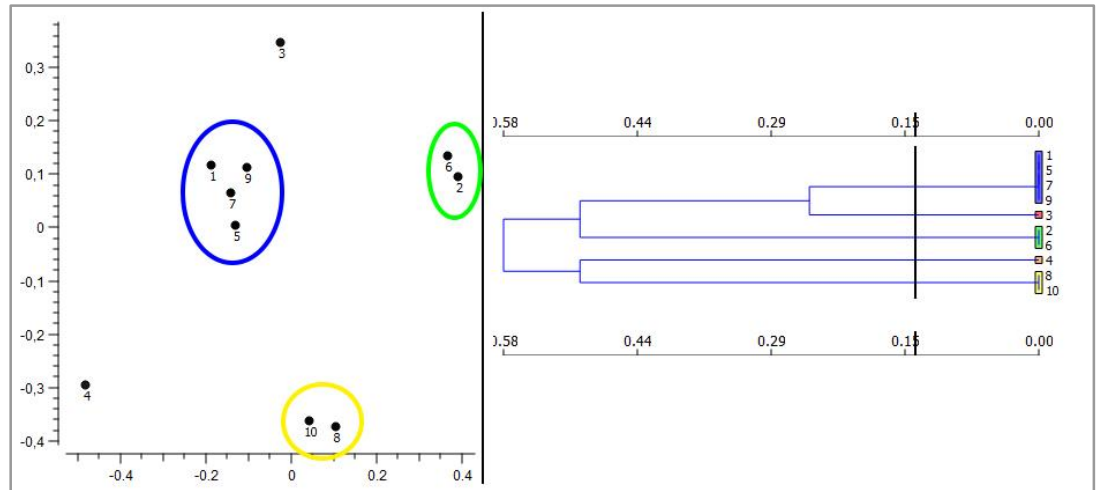
- d. Índice potencial de tensión: se presenta una cadena familiar con potenciales de tensión que van de mayor a menor de la siguiente forma: 2, quien presenta varias distancias de tono y la relación Sol-Fa# (semitono). {3, 6, 10} contienen un Do-Si (semitono) en proporción regular. 8, contiene también Si-Do pero en menor proporción. Los demás {1, 4, 5, 7, 9} con diferencias entre alturas mayores al semitono. En conclusión, se pueden agrupar en dos familias, no tan polarizadas.

Figura 47. Visor MDS y jerárquico para el índice potencial de tensión en profundidad.



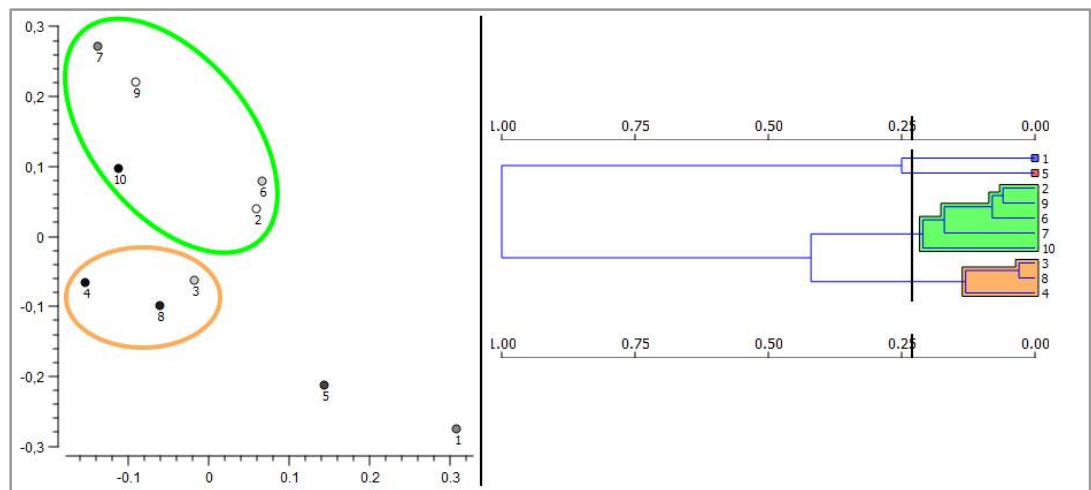
- e. Línea rítmica: relaciona todos los antecedentes {1, 3, 5, 7, 9} aunque destacando particularidades en el fragmento 3, también relaciona algunos consecuentes y aísla al fragmento 4.

Figura 48. Visor MDS y jerárquico para la línea rítmica en profundidad.



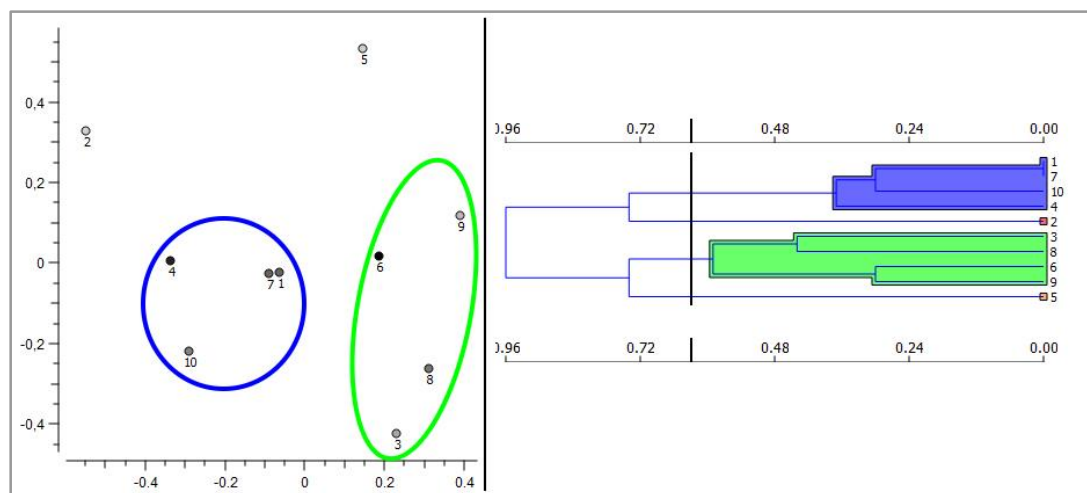
- f. Línea melódica: presenta una dispersión de grupos donde se caracterizan los de dirección mayormente descendente {2, 9, 6, 7, 10}, mayormente ascendente {3, 8, 4} y el grupo aislado y característico con salto ascendente en la melodía {1, 5}.

Figura 49. Visor MDS y jerárquico para la línea melódica en profundidad.



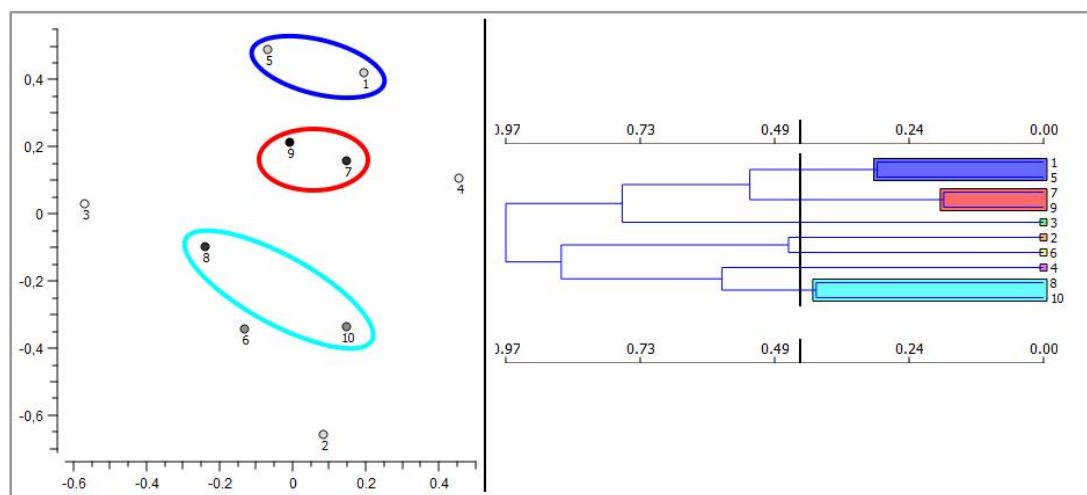
- g. Distribución de alturas: en esta dimensión se destaca el fuerte grupo {1, 7} por utilizar las mismas notas como su insumo melódico (Si y Re), y el aislamiento de los fragmentos 2 (Fa#) y 5 (Do#)

Figura 50. Visor MDS y jerárquico para la distribución de alturas en profundidad.



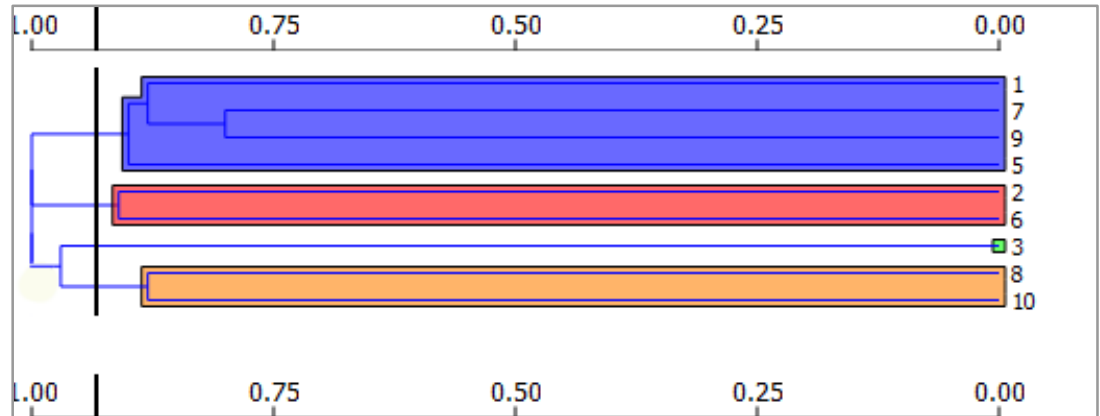
- h. Kernel lineal: a este nivel de detalle es aún más difícil leer los resultados compuestos dentro de un espacio unificado generado por el Kernel y es en este punto donde se percibe la importancia de los resultados detallados encontrados en el análisis multivista. De todas formas en el análisis general se destacan las relaciones {7, 9}, {1, 5} y {8, 10}, también {2, 6} de forma débil.

Figura 51. Visor MDS y jerárquico para el Kernel lineal en profundidad.



- i. Parametrización adaptativa: con este espacio inducido se obtienen las mismas agrupaciones del Kernel lineal pero con diferencias en la jerarquía.

Figura 52. Visor jerárquico para la parametrización adaptativa en profundidad.



Como conclusión de los resultados expuestos, se puede decir que el análisis por características revela rasgos representativos de la pieza musical, descritos en las relaciones entre fragmentos y a varias escalas de jerarquía.

9. PRUEBAS Y EVALUACIÓN

Dando cumplimiento en esta sección a lo planteado en la metodología, se tiene una fase de pruebas dividida en dos categorías: la primera de forma automática para el contraste de los cortes de primer nivel de jerarquía (frase y semifrase) y la segunda de tipo manual para los resultados de relaciones entre fragmentos y características de mayor nivel de detalle.

9.1. PRUEBAS AUTOMÁTICAS

El capítulo 4, en el numeral *Reducción del Espacio de Búsqueda*, se presenta un algoritmo para la identificación de *frases*, este algoritmo tiene referentes en la literatura y por tal razón se presenta enseguida un análisis comparativo con otros métodos.

La metodología propuesta en la literatura [4] menciona algunas definiciones y condiciones para aplicar la prueba y de esta forma hacer comparables los resultados.

- Un corte ocurre estrictamente entre dos notas.
- El modelo solo aplica para un nivel de jerarquía (frases y semifrases).

Dadas estas condiciones, el modelo necesita representar la fragmentación de la pieza musical utilizando un vector binario con un elemento por cada nota musical incluida. Cada entrada del vector indica si en esa posición se finaliza un fragmento.

Para un caso de comparación de resultado específico en una de las piezas musicales, se procedió de la siguiente manera: primero que todo, se aplicó el algoritmo de identificación de frases y su resultado se transformó en un vector binario (fragmentación candidata). Paso seguido y utilizando la representación en vector binario de la fragmentación etiquetada por los expertos [4] de forma manual para dicho ejercicio (fragmentación correcta), se aplicó un modelo de computación de distancia entre dos vectores binarios.

El modelo de distancia propuesto por la literatura [4] ha sido ampliamente utilizado en el campo de la *minería de datos*, la *recuperación de información* y la *predicción*; el modelo se fundamenta en la construcción de una matriz de confusión que se produce al poner en juicio la fragmentación candidata (predicción) respecto a la correcta (esperada):

- Verdaderos positivos (VP): Cantidad de cortes válidos (donde los dos vectores coinciden con valor 1).
- Verdaderos negativos (VN): Cantidad de posiciones sin cortes en acuerdo con la fragmentación correcta (donde los dos vectores coinciden con valor 0).
- Falsos positivos (FP): Cantidad de cortes inválidos (los vectores no coinciden pero el vector candidato tiene 1).
- Falsos negativos (FN): Cantidad de cortes válidos ausentes (los vectores no coinciden pero el vector candidato tiene 0).

Para la actual evaluación y tomando como referencia los valores de la matriz de confusión, se utilizaron las medidas *precisión*, *sensibilidad* y *media armónica* (conocidas en inglés como *precision*, *recall* y *f1* respectivamente).

En este caso específico, la *precisión* representa la proporción de cortes que coinciden con frases correctas (VP) sobre la cantidad de cortes totales en la predicción sin incluir penalización cuando son erróneos, por otra parte, la *sensibilidad* representa también la proporción de cortes que coinciden con las frases correctas, pero, sobre la cantidad de cortes correctos en la fragmentación candidata, osea, penalizando los errores, con lo cual se puede asegurar que a mayores valores de *sensibilidad* la fragmentación es más relevante; finalmente, la *media armónica* es un tipo de promedio de las otras dos medidas, esto para regularizar la proporción de mayor número de cortes correctos con una mayor relevancia.

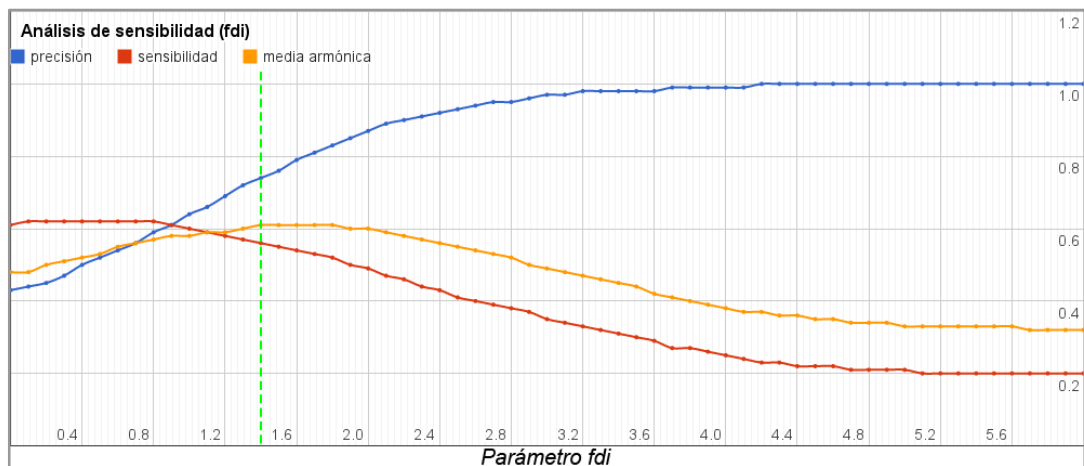
$$\text{Precisión} = \text{VP} / (\text{VP} + \text{FP})$$

$$\text{Sensibilidad} = \text{VP} / (\text{VP} + \text{FN})$$

$$\text{Media armónica} = (2 * \text{precisión} * \text{sensibilidad}) / (\text{precisión} + \text{sensibilidad})$$

Estas medidas fueron aplicadas al conjunto de datos de prueba ERK [30] (calculadas para cada pieza musical y promediadas para obtener la medición general) variando el parámetro de control *Umbral de corte (fdi)* como se expone en la Figura 53, realizando así un análisis de sensibilidad, para la búsqueda de un valor óptimo.

Figura 53. Análisis de sensibilidad para el parámetro *Umbral de corte (fdi)*.



El análisis fue realizado para las medidas *precisión*, *sensibilidad* y *media armónica* sobre el conjunto de datos ERK [30].

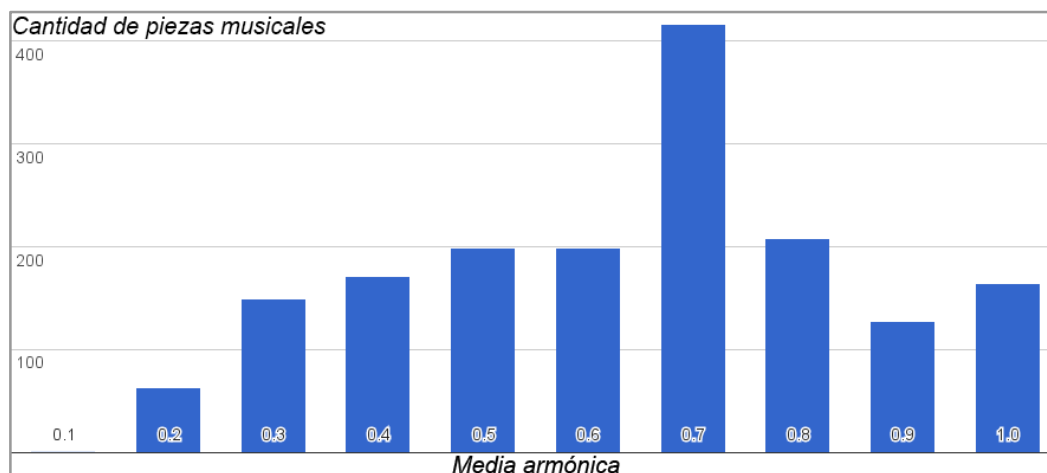
Analizando cada una de las variables por separado se obtiene:

- a. **Precisión:** cuando el parámetro está en el extremo inferior (cerca a cero), el valor de la *precisión* es mínimo, indicando que existen pocos cortes válidos, esto debido a que un umbral bajo produce pocos cortes en la pieza. A medida que el parámetro aumenta, también lo hace la *precisión* hasta llegar a su valor máximo (*Umbral de corte* = 4.2), en este punto cabe resaltar que los umbrales altos producen fragmentación con cortes en cada una de las notas de la pieza musical, razón por la cual incluye todos los cortes válidos (VP) y a la vez todos los erróneos (FP).
- b. **Sensibilidad:** para el extremo inferior, la *sensibilidad* toma valores medios, indicando que tiene una buena proporción de aciertos y errores. Esto es normal, recalcando que en umbrales bajos se tienen pocos cortes. A medida que el umbral aumenta esta regularidad empieza a perderse dado que la cantidad de cortes erróneos también empieza a crecer.
- c. **Media armónica:** encuentra su punto de equilibrio máximo al llegar a 0.61 con un *Umbral de corte* en 1.4 y con buenos valores para *sensibilidad* (0.56) y *precisión* (0.74).

Aplicando el algoritmo a las 1.700 piezas musicales del conjunto de datos de prueba, 135 (aproximadamente un 8% de las piezas) mostraron resultados excelentes alcanzando una *media armónica* con valor 1.0 y 416 (casi un 25% del conjunto) con resultados favorables alcanzando valores alrededor de 0,7.

El histograma que se muestra en la Figura 54 ejemplifica los resultados mencionados.

Figura 54. Histograma para la media armónica.



Histograma obtenido para la *media armónica* con un *umbral de corte* de 1.4.

De esta manera se evidencia que el algoritmo presentó resultados competitivos frente a los modelos en contraste expuestos en la Tabla 1 del numeral 2.2.2.3.

9.2. EVALUACIÓN POR EXPERTOS

Como se propuso en la metodología, la evaluación de resultados del algoritmo a nivel de relación de fragmentos y profundidad, se realizó con la ayuda de músicos con conocimiento en diseño formal.

De la totalidad de las piezas musicales ofrecidas por ERK se tomaron en primera instancia las 135 que presentaron mejores resultados en el proceso de evaluación automática, esto con el fin de garantizar que el análisis de relaciones y profundidad tuviera resultados coherentes y de esta forma las pruebas manuales se realizaran en condiciones ideales para una evaluación.

Teniendo en cuenta lo costoso en tiempo y dinero que puede ser un proceso de evaluación manual en cuanto a la preparación de las pruebas mismas y la evaluación por parte de los músicos, se decidió seleccionar una cantidad muy limitada de ejercicios, 3 (al azar) en total, para realizar un piloto de pruebas en condiciones mínimas. Para la lectura de resultados, los ejercicios fueron identificados con el código que aparecen en la colección ERK, de la siguiente forma:

- **0623:** Et Frad de Kenenk Genzt Dem Rein.
- **0969:** Wilhelm Komm An Meine Seite.
- **1806:** Was Hoerth Man Denn Neues Vom Kriegen.

Para cada una de las piezas se aplicó el proceso de Identificación estructural ejemplificado en el capítulo anterior, y los resultados se tradujeron a partir de los números y estructuras de jerarquía, a 10 enunciados musicales, cada uno acompañado de una imagen de la pieza misma con los fragmentos en discusión resaltados. La Figura 55, ilustra un ejemplo de enunciado de evaluación.

Figura 55. Ejemplo de enunciado de evaluación.

The image shows a musical score snippet with four staves. The first staff has a green highlight labeled 'F1' covering a phrase. The second staff has a green highlight labeled 'F1' covering a phrase and another green highlight labeled 'F2' at the end. The third staff has a green highlight labeled 'F3' covering a phrase. The fourth staff continues the melody. The score is in treble clef with a key signature of one sharp (F#) and a time signature of 6/8.

El texto del enunciado que acompaña a esta imagen es: “La primera semifrase de F1 y la semifrase presente en F2 utilizan la misma idea rítmica identificable por su finalización en corchea-negra”.

Es importante aclarar que los evaluadores no tuvieron acceso a las visualizaciones en MDS ni al Árbol de jerarquías, esto para tener unas pruebas más parecidas a la explicación escrita del análisis realizada por un músico.

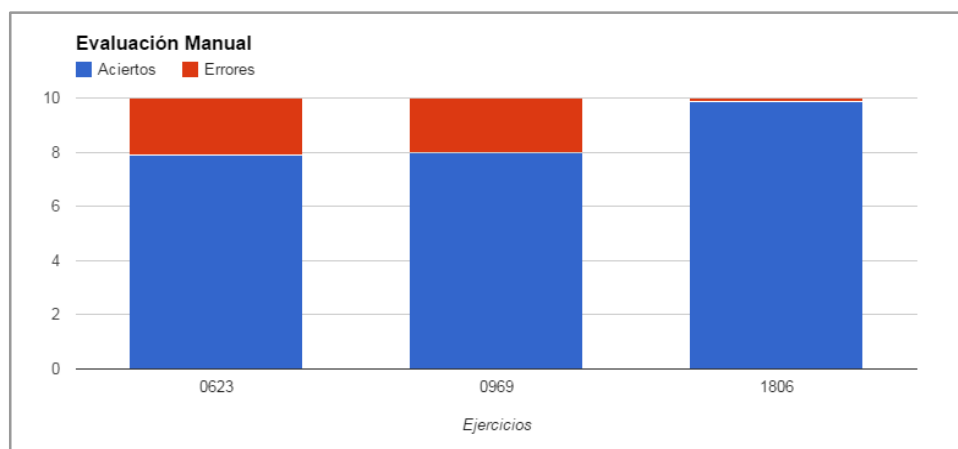
En resumen, la evaluación consta de 3 secciones (por cada pieza); cada sección inicia presentando la partitura y un audio de la pieza, seguida por los 10 enunciados, en cada uno de los cuales el evaluador debía indicar si estaba o no de acuerdo con lo expuesto. Finalmente se le solicitó al evaluador que describiera los rasgos de diseño formal característicos de la pieza musical que no se mencionaron en los enunciados.

El equipo de evaluadores fue integrado por 8 músicos de niveles variados (2 profesionales graduados, 3 estudiantes de semestres avanzados y 3 músicos empíricos), todos conocedores de la forma musical, quienes participaron voluntariamente para aplicar a las encuestas.

9.2.1. Resultados cuantitativos

Los resultados de las encuestas fueron tabulados y analizados para determinar un valor de acierto general de 8,6 sobre 10. La Figura 56 muestra los promedios de acierto y error para cada ejercicio, según las evaluaciones de los músicos.

Figura 56. Resultados de la evaluación manual.



Para cada ejercicio, se presentan los promedios de aciertos y errores determinados por los evaluadores.

9.2.2. Resultados cualitativos

En general, los resultados escritos, entregados por los evaluadores, acordaron las mismas acotaciones:

- El desacuerdo en algunos casos con los puntos de corte para los segmentos en profundidad, lo cual tiene explicación, por un lado, en la subjetividad natural que caracteriza este tipo de análisis, por otro, en la falencia actual del algoritmo por unificar dichos puntos de corte a partir de los resultados parciales en cada característica.
- La inconformidad con la exposición de algunas relaciones de familiaridad, sobre todo en los casos de agrupamientos dispersos, donde fue necesario recurrir a textos como “guarda cierta familiaridad con...” y “similar a... excepto por...” para lograr una explicación en palabras de las relaciones, que seguramente se comprenden mejor al visualizarlas en los árboles de jerarquía o los diagramas MDS.

CONCLUSIONES

La realización de este trabajo, permitió el desarrollo satisfactorio de una metodología para la identificación estructural automática de piezas musicales, utilizando una propuesta novedosa de análisis de las relaciones de similitud, desde una aproximación, no reportada en la literatura, de vistas múltiples, demostrando la riqueza y detalle descriptivo que puede tener frente a modelos más simples.

El reto de la representación vectorial de los elementos musicales fue superado mediante el enriquecimiento de las características reportadas en la literatura con características novedosas fundamentadas en conceptos teóricos musicales, demostrando su utilidad en la identificación de familiaridad y de rasgos diferenciales, necesarios en la descripción detallada del diseño formal de la pieza musical analizada.

La representación de piezas musicales planteada implica la fragmentación de la misma en frases como insumo principal para la identificación estructural automática, para lo que se desarrolló un nuevo algoritmo de particionamiento de nivel superior que muestra resultados competitivos respecto a los demás métodos computacionales reportados en la literatura.

El análisis detallado de la estructura de una pieza musical presentó un reto en cuanto al gran tamaño del espacio de búsqueda y la exploración de las posibles fragmentaciones; para superarlo, se propuso una nueva heurística de identificación de frases basada en la percepción de pausas, aplicable a piezas cortas, permitiendo de esta forma recorrer en un tiempo computacional razonable posibles particionamientos de tamaños heterogéneos y a un mayor nivel de profundidad.

En cuanto el agrupamiento de los fragmentos la aproximación multivista requirió del desarrollo de un nuevo índice que tuviera en cuenta el aporte de cada una de las dimensiones o vistas y promediarlo con el agrupamiento global, lo cual significó el desarrollo de una variante del algoritmo de Levenshtein⁶ para calcular distancia y se utilizó en la comparación de cadenas rítmicas y melódicas. También se propusieron variantes a la representación de secuencias utilizando el criterio de vista perspectiva para encontrar relaciones de similitud entre fragmentos que aplican variaciones rítmicas y melódicas.

El nuevo índice de evaluación de agrupamiento tiene en cuenta criterios teóricos de percepción musical en pro de evaluar y encontrar un agrupamiento con criterios similares a los utilizados por un músico experto.

Finalmente, la aproximación multivista propone otro reto, el de la comparación dinámica de las dimensiones generadas por cada una de las características, para ello se propuso una parametrización adaptativa para el Kernel lineal y su aplicación presentó resultados favorables encontrando relaciones dependientes del contexto de la pieza analizada.

⁶ Ver código fuente en el CD anexo

TRABAJOS FUTUROS

Identificación de notas no estructurales: dentro de las heurísticas propuestas para la reducción del espacio de búsqueda, se recomienda desarrollar un nuevo método para identificar notas no estructurales dentro de la pieza musical, estas notas en general no se encuentran en los puntos de fragmentación pero si aumentan el tamaño de la pieza en análisis.

Inclusión de nuevas características musicales: el enriquecimiento del modelo en cuanto a variables generadas para los fragmentos mejoraría los resultados obtenidos y permitiría tener una mayor dimensionalidad en los resultados que podrían aportar a una mejor discriminación de la estructura y el diseño formal.

Agrupamiento semi-supervisado: se puede determinar que las piezas musicales tienen un contexto que trasciende de las condiciones propuestas al inicio de esta investigación, donde se plantea un modelo NO supervisado para la solución del problema, fundamentada únicamente en la relación subyacente entre elementos constitutivos de la pieza musical. Adicionalmente se ha demostrado que los modelos que emplean aprendizaje computacional semi supervisado tienen mejoras en la adaptación del aprendizaje relacionado a un contexto sin perder las características de la No supervisión.

El aprendizaje semi-supervisado también sería una buena opción para la mejora en la integración de los análisis sobre las múltiples características analizadas, para mejorar no solo la respuesta totalizada que indica en qué puntos de la pieza se debe fragmentar, sino en la decisión misma de cómo agrupar los elementos bajo una característica, pero teniendo en cuenta los agrupamientos generados en las demás.

Lógica difusa: se puede inferir que la metodología propuesta puede tener problemas a la hora de seleccionar puntos únicos en los cuales fragmentar ciertas piezas musicales. Estos puntos únicos pueden ser dinámicos si la herramienta cuenta con las posibilidades ofrecidas por la lógica difusa, así mismo, los fragmentos podrán tener longitudes menos estrictas que permitirían en ciertos casos el solapamiento, algunas veces necesario para comprender las dinámicas más complejas de algunas piezas especiales.

Identificación estructural de composiciones extensas: la solución propuesta tiene problemas en cuanto a los tiempos de operación con piezas musicales de extensión mayor, dado que aún es altamente dependiente de las condiciones mismas del espacio de búsqueda. Por tal motivo se sugiere la exploración de otros métodos para el fraccionamiento en espacios de alta longitud.

Gracias a la riqueza de los resultados ofrecidos por esta metodología, se sugiere su uso para el desarrollo de varias aplicaciones en el campo de la tecnología musical, dentro de las que se pueden mencionar en primera instancia, un asistente para prácticas de análisis de forma (apoyo a la pedagogía), una mejora en la anotación de rasgos característicos (recuperación de información musical) y un asistente para composición (producción musical).

BIBLIOGRAFÍA

- [1] L. Horacio, T. Carlos, "Construcción de un Ambiente Interactivo para Asistencia en Línea al Aprendizaje de la Armonía Musical", *Experiencias Significativas en Innovación Pedagógica*, Eds. Germán Hernández y Freddy Abel Vargas Cardozo, Universidad Nacional de Colombia, <http://dis.unal.edu.co/~hernandg/esip/esip.pdf>.
- [2] Pierre-Yves, R., & Jean-Gabriel, G. (2002). Pattern Detection and Discovery: The Case of Music Data Mining. *Pattern Detection and Discovery*, 190–198. doi:10.1007/3-540-45728-3_15
- [3] R. Pierre-Yves, G. Jean-Gabriel, "Music Pattern Extraction and Similarity Assessment", *Readings in Music and Artificial Intelligence*, Volume 11, Eduardo Reck Miranda, Routledge Publisher, New York, USA, 2000.
- [4] Marcus T. Pearce, Daniel M'ullensiefen and Geraint A. Wiggins, *Melodic Grouping in Music Information Retrieval: New Methods and Applications*
- [5] M. G. Douglass, "Form in Tonal Music", Second Edition, Chapter 3 - Melodic structure of the phrase, Harcourt Brace Jovanovich College Publishers, 1993.
- [6] Piston, W., DeVoto, M., & Jannery, A. (1978). *Harmony* (pp. 440-442). London: Gollancz.
- [7] Schönberg, A. (1967). *Fundamentals of musical composition* (p. 58). G. Strang (Ed.). Faber & Faber.
- [8] Lerdahl, Fred y Jackendoff, Ray. (1983) *Grouping Structure*, En *A Generative Theory of Tonal Music* MIT Press, London.
- [9] Russell, S., Norvig, P. & Davis, E. (2010). *Artificial intelligence: a modern approach*. Upper Saddle River, NJ: Prentice Hall.
- [10] Chen, H. C., Lin, C. H., & Chen, A. L. (2004, June). Music segmentation by rhythmic features and melodic shapes. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on* (Vol. 3, pp. 1643-1646). IEEE.
- [11] Schuller, B., Dibiasi, F., Eyben, F., & Rigoll, G. (2010). Music thumbnailing incorporating harmony-and rhythm structure. In *Adaptive Multimedia Retrieval*.

- Identifying, Summarizing, and Recommending Image and Music (pp. 78-88). Springer Berlin Heidelberg.
- [12] Bimbot, F., Deruty, E., Sargent, G., & Vincent, E. (2012, October). Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. In 13th International Society for Music Information Retrieval Conference (ISMIR).
- [13] Goebel, M., & Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. *ACM SIGKDD Explorations Newsletter*, 1(1), 20-33.
- [14] Mabroukeh, N. R.; Ezeife, C. I. (2010). "A taxonomy of sequential pattern mining algorithms". *ACM Computing Surveys*
- [15] Ren, J.-M., Chen, Z.-S., & Jang, J.-S. R. (2010). On the use of sequential patterns mining as temporal features for music genre classification. *IEEE International Conference on Acoustics, Speech and Signal Processing*. doi:10.1109/icassp.2010.5495955
- [16] Darrell Conklin (2010) Discovery of distinctive patterns in music, *Intelligent Data Analysis*, Volume 14, 5, 547 – 554.
- [17] Ron Kovahi; Foster Provost (1998). "Glossary of terms". *Machine Learning* 30: 271–274
- [18] Christoph F. Eick , Nidal Zeidat , Zhenghong Zhao, Supervised Clustering – Algorithms and Benefits, *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI04) (2004)* Boca.
- [19] Dy, J. G., & Brodley, C. E. (2004). Feature selection for unsupervised learning. *The Journal of Machine Learning Research*, 5, 845-889.
- [20] Acharyya, Ranjan (2008); *A New Approach for Blind Source Separation of Convolutional Sources*, ISBN 978-3-639-07797-1.
- [21] Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4), 354-359.
- [22] Huma, Lodhi., Craig Saunders, John Shawe-Taylor, Nello Cristianini, and Chris Watkins. (2002) "Text Classification using String Kernels". *Journal of Machine Learning Research* 2: 419-444.

-
- [23] B. Asa, S. N. William (2005) "Kernel Methods for Predicting Protein-Protein Interaction" Department of Genome Sciences - Department of Computer Science and Engineering, University of Washington, Seattle, WA, USA.
- [24] Ben-Hur, a. (2003) "Remote homology detection: a motif based approach". *Bioinformatics* 19 (90001) (July 3): 26i-33. doi:10.1093/bioinformatics/btg1002. Disponible <http://www.bioinformatics.oupjournals.org/cgi/doi/10.1093/bioinformatics/btg1002>.
- [25] Cohen, P., Adams, N., & Heeringa, B. (2007). Voting experts: An unsupervised algorithm for segmenting sequences. *Intelligent Data Analysis*, 11(6), 607-625.
- [26] Bernhard Schölkopf, Koji Tsuda, Jean-Philippe Vert. (2004) "Kernel methods in computational biology". Yamanishi.
- [27] H. Jegou, M. Douze, C. Schmid, "Improving Bag-of-Features for large scale image search", *International Journal of Computer Vision*, Volume 87, Number 3, 316-336.
- [28] Haussler, David. (1999). "Convolution Kernels on Discrete Structures".
- [29] Schaffrath, H. (1995). The Essen folksong collection. In D. Huron (Ed.), *Database containing 6,255 folksong transcriptions in the Kern format and a 34-page research guide* [computer database]. Menlo Park, CA: CCARH.
- [30] *Deutscher Liederhort*. Auswahl der vorzüglicheren Deutschen Volkslieder, originally collected by Ludwig Erk; revised and edited by Franz Magnus Böhme in 3 vols., Leipzig, 1893-94.
- [31] Alias-i, LingPipe, Natural language processing software for text analytics, text data mining and search. Disponible en <http://alias-i.com/lingpipe/>, Brooklyn, NY
- [32] Demšar, J., Curk, T., & Erjavec, (2013) A. Orange: Data Mining Toolbox in Python; *Journal of Machine Learning Research* 14 (Aug), 2349–2353, 2013.