



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
POSGRADO EN CIENCIA E INGENIERÍA DE LA
COMPUTACIÓN

Evaluación de la novedad musical en sistemas co-creativos

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS (COMPUTACIÓN)

PRESENTA:

SAUL IVAN RIVAS VEGA

TUTORES:

DRA. WENDY ELIZABETH AGUILAR MARTÍNEZ
DR. CALEB ANTONIO RASCÓN ESTEBANÉ

CIUDAD DE MÉXICO, 2021

Dedico este trabajo a la intersección entre la ciencia y el arte.

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para poder realizar mis estudios de maestría (CVU: 972751).

Al Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas de la Universidad Nacional Autónoma de México por la educación recibida para realizar este trabajo de investigación. A mis tutores la Dra. Wendy Elizabeth Aguilar Martínez y el Dr. Caleb Rascon Estebané, su guía y experiencia me permitieron completar este trabajo extendiendo mi entendimiento y alimentando mi curiosidad sin dejar de lado la empatía y compasión durante los tiempos difíciles que afectan al mundo en la actualidad.

A mi familia, Ana Pimentel y mis amigos, por su apoyo incondicional y amor que me motivan día con día.

Declaracion de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

SAUL IVAN RIVAS VEGA. CIUDAD DE MÉXICO, 2021

Resumen

Medir la similitud entre melodías es una parte importante de cualquier esfuerzo que tenga como objetivo crear piezas musicales novedosas. Hay varias medidas actuales que lo llevan a cabo utilizando representaciones simbólicas de melodías. Este trabajo presenta una medida de similitud que proporciona una distancia entre melodías en un rango continuo de valores cerrados, que es capaz de discernir entre estilos. A través de experimentos controlados, mostramos que supera algunas medidas de similitud actuales al juzgar melodías que se modifican continuamente desde su versión original. Comparamos el rendimiento de la medida propuesta con otras medidas en la clasificación binaria de melodías que fueron juzgadas por expertos humanos. Y también mostramos cómo es capaz de crear un plano de similitud que es fácil de interpretar y útil de usar junto con un sistema co-creativo.

Índice general

Índice de figuras	XIII
Índice de tablas	XV
1. Introducción	1
1.1. Objetivos	2
1.2. Estructura de la tesis	2
2. Sistemas Co-Creativos	5
2.1. Características	6
2.2. Clasificación	6
2.2.1. Por tipo de interacción	6
2.2.2. Por roles	7
2.3. Evaluación	8
2.3.1. ¿Quién está evaluando?	8
2.3.2. ¿Qué se está evaluando?	8
2.3.3. ¿Cuándo ocurre la evaluación?	9
2.3.4. ¿Cómo se realiza la evaluación?	9
2.3.5. Evaluación de la Novedad	12
2.4. Compositores musicales co-creativos	13
2.4.1. Basados en división de tareas	13
2.4.1.1. MuseGAN	13
2.4.1.2. ChordAI	15
2.4.1.3. LeadsheetGAN	16
2.4.1.4. Anticipation-RNN	18
2.4.1.5. InpaintNet	20
2.4.1.6. CoCoNet	21
2.4.1.7. MusicVAE	22
2.4.1.8. Morphing GTTM	23
2.4.1.9. ChordLSTM	25
2.4.2. Basados en iniciativa mixta	26
2.4.2.1. CIM	26

2.4.2.2.	Impro-Visor	27
2.4.2.3.	JamSketch	28
2.4.3.	Basados en turnos explícitos	30
2.4.3.1.	DeepBach	30
2.4.3.2.	MidiNet	31
2.4.3.3.	LakhNES	33
2.4.3.4.	GrooVAE	34
2.4.3.5.	Folk-RNN	35
2.4.3.6.	MorpheuS	37
2.4.4.	Resumen	38
3.	Evaluación de la novedad musical	41
3.1.	Propuestas generales	41
3.1.1.	Propuesta de Alison Pease	41
3.1.2.	Criterios empíricos de Ritchie	45
3.2.	Medidas de similitud musical	49
3.2.1.	Basadas en contenido de audio	49
3.2.1.1.	Distancia Euclidiana en espacios multidimensionales usando PCA y RCA	49
3.2.1.2.	Divergencia Kullback-Leibler basada en Mezcla Gaussiana usando Coeficientes Mel	51
3.2.1.3.	Similitud basada en ritmo	52
3.2.1.4.	Similitud basada en clasificadores	53
3.2.1.5.	Similitud Híbrida Múltiple	55
3.2.1.6.	Similitud usando redes neuronales	56
3.2.1.7.	Similitud usando desviación concurrente	58
3.2.1.8.	Similitud usando descriptores melódicos y rítmicos	59
3.2.1.9.	Similitud con acercamientos probabilistas	62
3.2.1.10.	Similitud usando <i>SIMPL</i> e	63
3.2.1.11.	Similitud usando Recurrencia Cruzada	64
3.2.2.	Basadas en contenido representado simbólicamente	64
3.2.2.1.	Similitud usando distancia correlacional	65
3.2.2.2.	Similitud usando distancia de ciudad	66
3.2.2.3.	Similitud usando distancia Euclidiana	66
3.2.2.4.	Similitud usando Alineación Local	67
3.2.2.5.	Similitud usando Inducción de Estructura	68
3.2.2.6.	Similitud usando Interpolación	70
3.2.2.7.	Similitud usando Ondeletas	71
3.2.2.8.	Similitud usando redes neuronales	72
3.2.2.9.	Similitud usando grafos	73
3.2.2.10.	Similitud usando Arreglo Espiral	74
4.	RNA: Una medida de similitud musical	77
4.1.	Descripción	77

4.2. Evaluación de la medida propuesta	82
4.2.1. Evaluación con porcentajes de transformación	83
4.2.2. Evaluación con anotadores expertos	87
4.3. Caso de estudio con Impro-Visor	90
4.4. Discusión	94
5. Conclusiones	97
Bibliografía	99

Índice de figuras

4.1. Propiedades principales de comparación y orden presentes en <i>RNA</i>	78
4.2. Extracto del himno de la alegría, en su representación simbólica (arriba) y en una representación como una cadena cromática de notas musicales (abajo).	79
4.3. Ejemplo de una sub-secuencia común de longitud máxima entre un par de melodías.	81
4.4. Diferencias promedio por porcentaje de transformación.	85
4.5. Diferencias promedio por tipo de transformación.	86
4.6. Diferencias promedio por anotador.	87
4.7. Desempeño de las distintas medidas de similitud en las tareas de clasificación: a) “casi idénticas” ó “relacionadas pero variadas”, b) “relacionadas pero variadas” ó “diferente”, y c) “casi idénticas” ó “diferente”.	89
4.8. Visualización ejemplo del campo conceptual de Jazz y la posición de una melodía que no es nueva (izquierda), fundamentalmente nueva (centro) o demasiado caótica para considerarse nueva (derecha).	91
4.9. Visualización ejemplo del campo conceptual de Jazz y la cuantificación de la distancia de una melodía que no es nueva (izquierda), fundamentalmente nueva (centro) o demasiado caótica para considerarse nueva (derecha).	91
4.10. Medidas de similitud en un proceso co-creativo.	92

Indice de tablas

3.1. Características en medidas basadas en contenido de audio.	49
3.2. Conjuntos de datos evaluados.	50
3.3. Conjuntos de datos evaluados.	52
3.4. Conjuntos de datos evaluados.	53
3.5. Conjuntos de datos evaluados.	54
3.6. Resultados de cada métrica por tarea.	56
3.7. Resultados de cada métrica para similitud e inconsistencia.	56
3.8. Comparación de resultados por método.	57
3.9. Resultados de similitud entre canciones.	58
3.10. Comparación de resultados por transformación.	60
3.11. Comparación de resultados por transformación.	61
3.12. Exactitud de las métricas comparadas.	62
3.13. Resultados por conjunto de datos.	63
3.14. Tipos de Descriptores.	65
3.15. Comparación de resultados por método.	72
3.16. Exactitud por tipo de segmentos.	73
3.17. Similitud por sección.	75
4.1. Listado de métricas a comparar.	82
4.2. Comparación de diferencias promedio.	84
4.3. Comparación de diferencias promedio por porcentaje de transformación.	84
4.4. Comparación de diferencias promedio por tipo de transformación.	86

Capítulo 1

Introducción

El término creatividad ha sido históricamente difícil de definir. Por ejemplo, Taylor (1988) presenta más de 50 definiciones. Una razón importante para explicar esto es el hecho de que su significado ha cambiado a lo largo del tiempo y entre culturas. Sin embargo, parece haber cierta concordancia en reconocer a la novedad como característica fundamental de los artefactos creativos. Determinar la novedad requiere de la identificación y comparación de características clave entre un par de artefactos en donde no solamente se toma en cuenta el momento de su creación, ya que un artefacto recién creado puede ser simplemente una copia de algún otro creado con anterioridad. Estas características clave varían según el medio y el dominio en el cual se trabaje, y podemos realizar su comparación a través de una medida de similitud. Utilizar medidas de similitud para evaluar la novedad de los artefactos es un proceso independiente de su origen y por lo tanto incluye tanto a los creados por humanos como a los creados por sistemas computacionales autónomos y colaborativos, más aún, estas medidas pueden ayudar a guiar parte del proceso creativo de los sistemas.

Múltiples medidas de similitud han sido propuestas para comparar melodías representándolas simbólicamente (por ejemplo, en formato MIDI), las cuales se pueden usar para evaluar qué tan diferentes son las melodías creadas por un sistema computacional (sólo o en colaboración con algún humano) de aquellas pertenecientes a un corpus relevante. Sin embargo, la mayoría de las métricas existentes dependen de un umbral personalizado para discriminar únicamente entre melodías similares y diferentes, lo cual no aporta detalles sobre un grado de similitud.

Superar este obstáculo podría beneficiar el desarrollo y actualización de distintas herramientas de generación y análisis musical. Con este propósito, se propone en este trabajo de investigación desarrollar una medida que considere un rango continuo cerrado de valores que permitan una comparación entre ellos independientemente de las melodías analizadas. Al ser continuo puede dar una perspectiva de exploración creativa con mayor fluidez y que se pueda utilizar para visualizar intuitivamente en un plano conceptual de distintos estilos.

1. INTRODUCCIÓN

La medida de similitud puede apoyar en el avance en tareas como: identificación de *covers*, consultas de búsqueda por contenido, clasificación por emoción y género musical y obtención de huella digital musical. Adicionalmente, puede ayudar a medir el avance en nuevos trabajos propuestos.

1.1. Objetivos

Objetivo General: Proponer una nueva medida de similitud entre melodías que permita cuantificar sus diferencias de manera continua y en un rango acotado de valores.

Objetivos específicos:

- Compararla con otras medidas de similitud del estado del arte.
- Compararla con respecto a la similitud asignada por expertos humanos.
- Evaluar la novedad de las melodías generadas en un proceso co-creativo humano-computadora.

1.2. Estructura de la tesis

En el capítulo 2 se detallan brevemente las bases teóricas de la creatividad computacional con el propósito de introducir el concepto de sistema co-creativo, sus características y su clasificación dependiendo de su tipo de interacción y el rol que puede desempeñar en un proceso creativo. De igual forma se describen las distintas perspectivas que se pueden tomar en la evaluación de los sistemas creativos, considerando elementos como observadores, el entorno, los productos y el momento en el tiempo en el que se realiza la evaluación. Posteriormente se incluye una lista, que no pretende ser exhaustiva, de compositores musicales co-creativos explicando detalles detrás de su funcionamiento y evaluación. Al final de este capítulo se sintetizan los conceptos más importantes en una sección de resumen.

El capítulo 3 comienza describiendo distintas metodologías de evaluación de la novedad en el contexto de la creatividad computacional. En seguida, se describen las principales medidas de similitud musical que existen hasta el momento, organizadas por el tipo de representación que utilizan.

Utilizando una representación simbólica, el capítulo 4 define la medida de similitud propuesta, sus propiedades y pasos a seguir para llevarla a cabo. Para evaluar su desempeño, la medida propuesta se comparó con una selección de métricas revisadas en el capítulo 3 cuyas implementaciones se encuentran disponibles públicamente.

La evaluación incluye la descripción y resultados de una evaluación objetiva en conjuntos de datos anotados y de igual forma una evaluación subjetiva con anotaciones hechas por músicos expertos humanos. Se incluye al final del capítulo una sección de discusión sobre los resultados y comportamiento de las medidas.

Finalmente, el capítulo 5 engloba las conclusiones a partir de los resultados obtenidos y presenta algunas ideas del trabajo a futuro para extender esta investigación.

Capítulo 2

Sistemas Co-Creativos

La idea de construir computadoras capaces de mostrar cierto grado de creatividad se remonta hasta un escrito de Ada Lovelace en donde considera que, a las máquinas, aunque capaces de generar sonatas, nunca se les podrá atribuir la característica de ser creativas puesto que solo hacen lo que se les ordenó. En la escuela de verano de Dartmouth de 1956 se pretendía definir en su totalidad a la inteligencia artificial, incluyendo a la creatividad como parte de de las características a modelar computacionalmente. Sin embargo, a pesar de que existieron trabajos pioneros en el área, no fue hasta los años 90's que comenzó a captar más la atención de la comunidad científica y artística, y fue hasta el año 2010 que se organizó la primera conferencia internacional en creatividad computacional (ICCC). Desde entonces, ésta se lleva a cabo anualmente, convirtiéndose en el principal evento de esta área de investigación. En esta conferencia definen a la creatividad computacional como:

La creatividad computacional es un esfuerzo interdisciplinario localizado en la intersección de los campos de la inteligencia artificial, psicología cognitiva, filosofía, ciencias sociales y las artes. La meta de la creatividad computacional es modelar, simular o replicar la creatividad utilizando una computadora, cumpliendo uno de los siguientes objetivos:

- *Construir un programa o computadora capaz de mostrar algún grado de creatividad.*
- *Mejorar el entendimiento de la creatividad humana y formular una perspectiva algorítmica sobre el comportamiento en humanos.*
- *Diseñar programas que aumenten a la creatividad humana sin ser creativos por sí solos.*

Es así, que los sistemas computacionales pueden jugar diferentes roles en el proceso creativo Lubart (2005):

2. SISTEMAS CO-CREATIVOS

- *Niñera*: El sistema computacional mantiene un ambiente controlado de acción creativa con el propósito de incentivar y mantener motivado a un inexperto en una actividad creativa en particular.
- *Facilitador*: Herramientas que permitan una comunicación efectiva entre los miembros de un grupo al realizar alguna actividad creativa.
- *Coach*: Sistema experto que motiva a una persona al realizar alguna actividad creativa guiándolo y mostrando sugerencias que potencien su creatividad.
- *Colega*: El sistema colabora con un humano en la creación de un artefacto, generando sus ideas basadas en el trabajo en conjunto. A este tipo de sistemas se les llama co-creativos.
- *Creador autonomo*: El sistema genera sus propios productos creativos de manera independiente.

En este trabajo de investigación estamos interesados en evaluar la novedad de las composiciones musicales generadas en ambientes co-creativos, por lo cual el resto del capítulo se enfocará en este tipo de sistemas.

2.1. Características

Tomando en cuenta la definición propuesta en Karimi et al. (2018), “Un sistema co-creativo es aquel sistema en el que interactúan por lo menos un agente artificial y un agente humano donde toman acción con base en la respuesta de su compañero y en su propio conocimiento durante la tarea co-creativa”. Al tipo de colaboración donde los diferentes participantes contribuyen al producto final y sintetizan sus aportaciones durante la interacción es a lo que nos referimos como co-creatividad.

2.2. Clasificación

Los sistemas co-creativos se pueden clasificar de acuerdo con el tipo de interacción que se lleva a cabo entre los colaboradores o por el papel que juegan en el proceso creativo.

2.2.1. Por tipo de interacción

- *División de tareas*: En este tipo de interacción los distintos participantes trabajan de manera independiente en su aportación considerando las ideas de los demás participantes y cumpliendo con sus tareas asignadas. Davis et al. (2015) definen incluso la participación en tiempo real como requerimiento, por ejemplo, la que se da una improvisación.

- **Iniciativa Mixta:** Yannakakis et al. (2014), describen que en este tipo de interacción tanto el humano como el agente artificial tienen iniciativa propia de participar activamente en un mismo artefacto compartido. Donde “participar activamente” significa que el agente artificial es proactivo en el sentido de que no responde a una petición o reacción del usuario. Ambos agentes no requieren contribuir al mismo grado ni de manera simétrica.
- **Turnos explícitos:** Los turnos en los cuales cada agente realiza su aportación está definido. En algunos casos el turno del agente artificial se percibe como una interface “mágica” que responde a las peticiones del usuario siguiendo pasos específicos. Otros, son dominados por el agente humano donde el agente artificial tiene una menor participación con sugerencias u observaciones críticas.

2.2.2. Por roles

Durante el proceso creativo existen distintas tareas que pueden ser distribuidas entre el equipo y han servido para categorizar los roles en la Co-Creatividad de la siguiente manera:

- *Generar:* En esta categoría los sistemas tienen un rol generativo produciendo su aporte al producto final.
- *Evaluar:* No todos los agentes generadores tienen la capacidad de evaluar los productos o el producto en conjunto, sin embargo, se busca que el tener el rol de generador sea emparejado con el de evaluador para mejorar.
- *Encontrar problemas:* En esta categoría los roles buscan definir los conceptos a explorar en la actividad creativa, su estética y la abstracción del porqué de la solución o producto generado.
- *Controlar la iniciativa:* El agente puede generar un impacto en el grupo que termine afectando al producto final dependiendo de su intención, por ejemplo, un ente que tenga como rol complacer la intención del grupo realizando acciones que estén acorde, o al contrario tener el rol de provocar al grupo al realizar acciones en desacuerdo con el fin de explorar el dominio creativo de la tarea.

En general, las categorizaciones como la anterior y las clasificaciones como la inicial son idealizaciones de un proceso creativo general, no se ignora que en actividades en específico dichas categorizaciones ya no son aplicables y se deben buscar mejores categorizaciones tomando en cuenta las características individuales en el estudio de un dominio creativo en específico, lo cual sigue siendo un latente problema de investigación.

2.3. Evaluación

El marco de evaluación propuesto en Karimi et al. (2018), presenta cuatro preguntas que guían la evaluación realizada a sistemas co-creativos.

2.3.1. >Quién esta evaluando?

- **La inteligencia artificial:** el agente o agentes artificiales son capaces de auto-evaluar sus creaciones y son conscientes de sus procesos.
- **El usuario:** el agente o agentes humanos evalúan los resultados o la creatividad del sistema, especialmente durante el proceso creativo.
- **Un tercero:** ni el sistema, ni los agentes humanos evalúan al sistema, este tercer caso toma la forma de una evaluación realizada por expertos en el dominio quienes evalúan la creatividad y calidad de los productos resultantes.

En el dominio de la música, Stowell et al. (2009) proponen realizar la evaluación usando una modificación a la prueba de Turing, en donde un baterista artificial, un músico humano y un metrónomo realizan el seguimiento y acompañamiento de una línea base. La idea de incluir a un metrónomo es que este sea un punto de referencia así el baterista artificial pretendería no ser confundido con el metrónomo y a su vez ser confundido por el humano. En este caso, la evaluación es realizada por músicos expertos que escuchan segmentos donde solo suena alguno de los 3 y requieren identificar al que este generando el acompañamiento.

2.3.2. >Que se esta evaluando?

- **Producto:** resultado del esfuerzo conjunto del usuario y el sistema. Usualmente ambos trabajan con un mismo objetivo, sin embargo, existen tareas abiertas donde la exploración se da con objetivos distintos y entonces existe una creatividad emergente.
- **Proceso:** se realizan pruebas al sistema en las que se evalúa, por ejemplo, su habilidad y apreciación.
- **Creatividad del usuario:** apoyada en trabajos referentes a la psicología, por ejemplo.
- **Interacción máquina-humano:** hay un dinamismo en la interacción entre el usuario y el sistema innato de toda colaboración creativa y requiere un conjunto de métodos de evaluación distintos del producto o del proceso. Es el enfoque final de lo que puede ser evaluado en sistemas co-creativos, evaluando aspectos tales como: expresividad, efectividad o el afecto que produce.

Kantosalo et al. (2015) utilizaron el marco de evaluación D.E.C.I.D.E. con el fin de llegar a definir su proceso de evaluación en el desarrollo de un sistema interactivo de creación de poemas. En este trabajo se evalúa al sistema creativo en usabilidad, utilidad y que tan agradable es la interacción.

2.3.3. >Cuándo ocurre la evaluación?

- **Durante el proceso creativo:** evaluación *Formativa*, la cual provee retroalimentación activa de un agente (artificial o humano) a otro, con el fin de guiar sus acciones en contextos a tiempo real, toma de turnos, iniciativa mezclada, etc.
- **Después del proceso creativo:** evaluación *Sumativa*, el usuario provee retroalimentación que puede influir en futuras tareas del sistema, normalmente se realiza en contextos de experimentación con el sistema.

El objetivo de Hsu y Sosnick (2009) fue desarrollar una metodología de evaluación para sistemas musicales interactivos en la improvisación. Las pruebas deben capturar las experiencias de los músicos interactuando con los sistemas, así como las experiencias de la audiencia que observa las interpretaciones. Por su simplicidad se escogieron cuestionarios y grabaciones de conversaciones con los músicos participantes después de las sesiones de pruebas.

2.3.4. >Como se realiza la evaluación?

- **Métodos:** el principal método de evaluación es el estudio de usuario y se ha realizado de distintas maneras. Algunos de los métodos son:
 - Observación
 - Análisis de protocolo
 - Encuestas
 - Entrevistas
 - Prueba de algoritmos
 - Análisis de datos biométricos
- **Métricas:** en diversos estudios se han realizado evaluaciones tomando en cuenta el valor cultural y la complejidad buscando una independencia de dominio. Algunas de las métricas utilizadas son:
 - Novedad
 - Calidad
 - Usabilidad

- Valor
- Utilidad
- Sorpresa
- Exactitud
- Participación

En el trabajo propuesto por Guzdial y Riedl (2019) se define un marco de referencia capaz de representar las categorías más utilizadas para la clasificación de sistemas que realicen tareas en contextos creativos en una interacción por turnos. El marco de referencia tiene los siguientes componentes para la reflexión en el diseño de sistemas creativos:

1. **Comienzo:** ¿Cómo empieza la interacción? ¿Quién tiene el primer turno?
2. **Acciones:** ¿Qué acciones son capaces de realizar los agentes dada la implementación y el diseño de la interfaz?
 - a) **Acciones en el artefacto:** En cada turno, ¿Cuáles son las acciones que los agentes pueden tomar que directamente afecten al artefacto?
 - b) **Otras acciones:** En cada turno, ¿Qué acciones no afectan directamente al artefacto? Por ejemplo, Un agente explica su razonamiento detrás de su acción.
 - c) **Acciones Fuera de turno:** ¿Qué acciones pueden llevar a cabo los agentes fuera de su turno? Por ejemplo, aprender de las acciones que tome el otro agente en su turno o explorar el artefacto.
3. **Inteligencia Artificial:** ¿Qué algoritmo o técnica es utilizada para entrenar o diseñar al agente artificial? Este agente, ¿Es estático, escogiendo de entre una serie de comportamientos existentes, o aprende durante la interacción?
4. **Usuario:** ¿Quién es el usuario objetivo para este sistema?
5. **Turnos:** ¿Cómo comienzan y terminan los turnos? ¿Qué agente tiene el control de esas decisiones?
6. **Final:** ¿Cómo termina la interacción? ¿Por cuál agente o cuales condiciones?

Distintas clasificaciones de sistemas creativos pueden ser representadas en este marco de referencia. Clases como las mencionadas en Karimi et al. (2018):

- Herramientas de soporte creativo: El agente artificial no tiene *Acciones en el artefacto*.
- Sistemas creativos autónomos: El agente humano no tiene *Acciones en el artefacto* y solo una *Otra acción* que es comenzar con el único turno del agente artificial.

- Sistemas co-creativos: Involucra que ambos agentes tengan *Acciones en el artefacto* tomando turnos.

Así como las categorías propuestas en Lubart (2005):

- Nana: El agente artificial no tiene *Acciones en el artefacto* y sus *Otras acciones* o *Acciones fuera de turno* son tareas como monitorizar, apoyar o supervisar al usuario.
- Facilitador: Ningún agente tiene *Acciones en el artefacto* y sus *Otras acciones* o *Acciones fuera de turno* son tareas de reflexión creativa entre los agentes.
- Coach: El agente artificial no tiene *Acciones en el artefacto* y sus *Otras acciones* o *Acciones fuera de turno* son tareas de enseñanza y retroalimentación constructiva con el usuario.
- Colega: El agente artificial no tiene *Acciones en el artefacto* y sus *Otras acciones* o *Acciones fuera de turno* pueden contribuir a interacciones complementadas.

Kantosalo y Riihiahho (2019) definieron su proceso de evaluación después de haber realizado un análisis de la literatura previa sobre perspectivas de experiencia de usuario, sistemas de soporte creativo y creatividad computacional. Obtuvieron 59 métricas en común y redujeron su proceso a considerar 7 de ellas:

1. Diversión
2. Disfrutabilidad
3. Expresividad
4. Satisfacción del resultado
5. Facilidad de realización
6. Colaboración
7. Propiedad

El trabajo expone una perspectiva humana al evaluar las interacciones en la co-creatividad revelando áreas de oportunidad, redundancia en las métricas y refuerzo en la robustez de la metodología.

Ravikumar et al. (2018) llevaron a cabo un estudio donde junto a un grupo de expertos, en la improvisación musical, definieron los aspectos esenciales en una interpretación en grupo para ser considerada una co-experiencia. Una co-experiencia es la experiencia subjetiva que evoca una improvisación en grupo a un músico que participa en ella. Para los agentes humanos la co-experiencia es la experiencia donde se miden aspectos de los sentimientos en conjunto durante una interacción y para el sistema computacional su intención es ser una construcción formal que guíe las decisiones del sistema.

Los aspectos de una improvisación en conjunto que fueron revisados son:

- **Interacción social:** según las respuestas de los entrevistados, no hay una comunicación no-verbal directa en interpretaciones con otros humanos durante una improvisación. Por lo tanto, no es una característica esencial en una co-experiencia.
- **Fisicalidad:** algo que consideraron importante durante una improvisación con otros humanos fue la *sicalidad* que, en lugar de una interacción social, únicamente se percibe visualmente a los otros músicos. Por lo tanto, no se consideró esencial este aspecto.
- **Entornos:** se realizó una comparación de sus experiencias en los siguientes entornos:
 - Interpretaciones con una computadora, los músicos detallaron en general que no requieren comunicarse directamente con movimientos físicos. Tocar con alguien sin presencia, como una computadora, no supone la pérdida de la co-experiencia, solo es más complicado.
 - Interpretaciones en ambientes computacionales, los ambientes computacionales son por ejemplo vídeo conferencias, las cuales introducen una menor fiabilidad en el sonido y aberraciones acústicas o de sincronización. Estos detalles, según los músicos, suelen ser asumidos como inherentes del ambiente y por lo tanto son considerados como elementos que usan para determinar sus siguientes acciones.
 - Músicos novatos, algunos de los entrevistados son maestros de música y mencionaron que, en este ambiente por la falta de experiencia de sus compañeros, requieren enfatizar sus instrucciones musicales para poder recibir una respuesta de los compañeros y coordinar sus acciones en conjunto. Esto no causa que se pierda la co-experiencia.
- **Coordinación de cambios:** la variación entre las expresiones musicales de seguir a otro músico o liderar la pieza musical es lo que distingue principalmente una co-experiencia. Los músicos describieron que su atención se concentra en las estructuras musicales que utilizan para liderar y seguir a los demás mientras se espera una respuesta y decidir su siguiente acción. La habilidad de coordinar diferencias en los entendimientos compartidos entre liderar y seguir parece ser importante en la co-experiencia.

2.3.5. Evaluación de la Novedad

En este trabajo se exploran propuestas que evalúan a los artefactos resultantes de un proceso creativo. Estas propuestas tienen fija la respuesta a “¿Qué es lo que se está evaluando?” puesto que se están evaluando los productos, quién lo realiza puede variar, de igual forma el cómo y el cuando. Como se muestra en este capítulo no existe un conjunto de características universales para evaluar artefactos creativos, pero si se encuentra a la novedad como característica común en distintos estudios, es por ello

que será el enfoque principal de este trabajo de investigación. Múltiples compositores musicales co-creativos han sido propuestos aún sin tener una metodología estandarizada de evaluación no solo de la novedad sino de los productos en general. En la siguiente sección se realiza una compilación que no pretende ser exhaustiva sino de muestreo en técnicas de evaluación. Posteriormente en el capítulo 3 se describe formalmente como definiremos novedad para este trabajo, tomando en cuenta lo revisado en marcos de evaluación generales de creatividad, el compendio de muestra de compositores musicales co-creativos y metodologías de evaluación pertenecientes al dominio musical.

2.4. Compositores musicales co-creativos

En este capítulo se presenta una lista de compositores artificiales donde se consideran sus principales características de la arquitectura de sistemas creativos y su forma de colaboración con el usuario para clasificarlos. Se detallan los agentes que participan en un caso de uso del sistema, así como la interacción entre ellos. Adicionalmente, se detalla las evaluaciones objetivas y subjetivas, si es que las tuvieron.

2.4.1. Basados en division de tareas

2.4.1.1. MuseGAN

MuseGAN (Dong et al., 2017), es un sistema de composición capaz de generar composiciones en múltiples pistas generando los acompañamientos para una melodía introducida por el usuario. Por ejemplo, una melodía dada por el usuario puede ser reproducida como la correspondiente a la guitarra mientras que MuseGAN genera la pista para el bajo y la batería. Es capaz de igual forma de generar las múltiples pistas de forma autónoma.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** 45,129 archivos del Dataset Lakh Pianoroll Dataset emparejados con su correspondiente archivo en el Million Song Dataset.
- **Estética:** Su estética se infiere a partir de la base de datos con el objetivo de que se generen salidas lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato Piano Roll.
- **Evaluación de la representación interna:** Al usar un modelo GAN el sistema evalúa sus productos en la etapa de entrenamiento con un discriminador, el cual lo clasificará como real o falso de acuerdo a que tan similar es con respecto a los que se encuentran en la base de conocimiento.

- **Conceptualización:** Un modelo GAN como parte del entrenamiento o aprendizaje extrae las características de los ejemplos para crear uno nuevo a través de su generador. El nuevo ejemplo es ahora clasificado por el discriminador como real o falso, si acierta entonces el generador ajusta sus parámetros para poder engañar al discriminador y en caso contrario el discriminador ajusta sus parámetros para clasificar correctamente el siguiente ejemplo.
- **Generación:** En MuseGAN el generador consiste de una red neuronal convolucional (CNN) que toma como entrada un vector de condiciones inferido de lo que ingreso el usuario o generadas aleatoriamente. La salida de la red será la nueva composición al ser decodificada.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Se realizaron pruebas de escucha con usuarios a los que se les pidió evaluar composiciones generadas por el sistema en las siguientes métricas:
 - Armonía placentera.
 - Ritmo unificado.
 - Estructura Musical clara.
 - Coherencia.
 - Calidad en general.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una pista inicial con el requerimiento mínimo de una estructura temporal definida.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una melodía en formato MIDI.
- El sistema decodifica la melodía y reconoce la estructura temporal presente.
- El sistema genera las pistas de los acompañamientos siendo similares a su base de conocimientos utilizada en el entrenamiento.
- El usuario entonces puede realizar modificaciones a las pistas en algún software de composición musical.

- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.2. ChordAI

ChordAI (Hao, 2019), es un sistema de composición capaz de generar melodías a partir de una progresión de acordes dada por el usuario. De igual forma puede generar nuevas progresiones de acordes que continúen a las dadas por el usuario.

Componentes del compositor musical co-creativo artificial

- **Bases de Conocimiento:**
 - **Nottingham Dataset:** Conjunto de anotaciones de acordes y melodías.
 - **McGill-Billboard Chord Annotations:** Un aproximado de 1000 canciones con anotaciones de acordes.
 - **CSV leadsheet database:** 2200 Canciones Occidentales de distintos géneros utilizada para la anotación de de acordes y melodías.
- **Estética:** Su estética se infiere a partir de la base de datos con el objetivo de que se generen salidas lo más similar posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato Piano Roll.
- **Evaluación de la representación interna:** Al usar un modelo LSTM el sistema evalúa sus productos en la etapa de validación del modelo, es decir que tan similar son sus predicciones con respecto a la base de conocimiento.
- **Conceptualización:** Usando un modelo de redes neuronales, la conceptualización se lleva a cabo en el entrenamiento cuando se determinan las distribuciones de probabilidad con base en el análisis de las características de los ejemplos de la base de conocimiento.
- **Generación:** Al utilizar una red neuronal que toma como entrada un vector de condiciones inferido de lo que ingreso el usuario o generadas aleatoriamente. La salida de la red será la nueva composición al ser decodificada.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.

- **Evaluación de la representación externa:** Se realizaron pruebas de escucha con usuarios a los que se les pidió evaluar composiciones generadas por el sistema en las siguientes métricas:
 - Armonía.
 - Ritmo.
 - Estructura.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una progresión de acordes que será utilizada en la composición.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una progresión de acordes en la notación occidental, por ejemplo: “F#:maj” y “A:min”.
- El sistema decodifica la progresión de acordes.
- El sistema genera la pista de la melodía siendo similar a su base de conocimientos utilizada en el entrenamiento.
- El usuario entonces puede realizar modificaciones a la pista en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas progresiones en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.3. LeadsheetGAN

LeadsheetGAN (Liu y Yang, 2018), es un sistema de composición capaz de generar composiciones en múltiples pistas generando los acompañamientos para una melodía introducida por el usuario. Por ejemplo, una melodía dada por el usuario puede ser reproducida como la correspondiente a la mano derecha mientras que LeadsheetGAN genera la pista para la mano izquierda. Es capaz también de generar las múltiples pistas de forma autónoma.

Componentes del compositor musical co-creativo artificial

- **Bases de Conocimiento:**
 - **Dataset Lakh Pianoroll Dataset:** Usando el subconjunto de 21,425 archivos multipista en formato Pianoroll.
 - **TheoryTab:** Alrededor de 16 mil segmentos con anotaciones multipista en formato XML.
- **Estética:** Su estética se infiere a partir de la base de datos con el objetivo de que se generen salidas lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato Piano Roll.
- **Evaluación de la representación interna:** Al usar un modelo GAN el sistema evalúa sus productos en la etapa de entrenamiento con un discriminador, el cual lo clasificará como real o falso de acuerdo a que tan similar es con respecto a los que se encuentran en la base de conocimiento.
- **Conceptualización:** Un modelo GAN como parte del entrenamiento o aprendizaje extrae las características de los ejemplos para crear uno nuevo a través de su generador. El nuevo ejemplo es ahora clasificado por el discriminador como real o falso, si acierta entonces el generador ajusta sus parámetros para poder engañar al discriminador y en caso contrario el discriminador ajusta sus parámetros para clasificar correctamente el siguiente ejemplo.
- **Generación:** En LeadsheetGAN el generador consiste de una red neuronal convolucional (CNN) que toma como entrada un vector de condiciones inferido de lo que ingreso el usuario o generadas aleatoriamente. La salida de la red será la nueva composición al ser decodificada.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Se realizaron pruebas de escucha con usuarios a los que se les pidió evaluar composiciones generadas por el sistema en las siguientes métricas:
 - Armonía.
 - Ritmo.
 - Sentimiento en general.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una progresión de acordes que será utilizada en la composición.

2. SISTEMAS CO-CREATIVOS

- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista inicial con la melodía en formato MIDI.
- El sistema decodifica las características de armonía, ritmo y tonalidad del archivo.
- El sistema genera las pistas de los acompañamientos: cuerdas, bajo, percusiones, guitarra y piano. Siendo similar a su base de conocimientos utilizada en el entrenamiento.
- El usuario entonces puede realizar modificaciones a las pistas en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.4. Anticipation-RNN

Anticipation-RNN (Hadjeres y Nielsen, 2020), es un sistema capaz de generar melodías en una sola pista de forma no necesariamente secuencial (es decir que puede componer para completar la composición dada por el usuario), algunos de los requerimientos que se pueden imponer por el usuario son:

- *Requerimiento 1*: Dar el inicio y el final de la melodía para que el sistema componga la parte media.
- *Requerimiento 2*: Dar el inicio y el final, aunque estos se encuentren en otra clave y requiera hacer una transición tonal.
- *Requerimiento 3*: Dejar huecos grandes en la entrada del usuario.
- *Requerimiento 4*: Utilizar solo una escala ascendente.
- *Requerimientos 5 y 6*: Usar notas aleatorias cada cuarto y octavo de beat respectivamente.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento**: Voz soprano de 402 composiciones en tiempo 4/4 de la librería music21.

- **Estética:** Su estética se infiere a partir de la base de datos con el objetivo de que se generen salidas lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en un formato especial codificando propiedades de la melodía y el ritmo.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una red LSTM.
- **Conceptualización:** Como parte de un modelo LSTM existe una red neuronal recurrente el cual sigue un proceso de entrenamiento donde define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica de la similitud y convicción al adjudicar algunas de las composiciones a Bach.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una composición en una sola pista a completar por el sistema.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista inicial con la melodía en formato XML desde algún software de composición.
- El sistema decodifica las condiciones impuestas por el archivo.
- El sistema genera la pista de la melodía incluyendo a las notas proporcionadas por el usuario, siempre y cuando logren mantener coherencia con la base conocimientos.
- El usuario entonces puede realizar modificaciones a la pista resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.

- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.5. InpaintNet

InpaintNet (Pati et al., 2019), es un sistema capaz de generar melodías en una sola pista, las cuales pueden ser condicionadas por el usuario al imponer notas al inicio y al final para que el sistema genere una melodía intermedia.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Melodías folclóricas monofónicas escocesas e irlandesas del sitio web "The Session".
- **Estética:** Su estética se infiere a partir de la base de datos con el objetivo de que se generen salidas lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en un formato especial codificando propiedades de la melodía y el ritmo.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que elegían que ejemplos de melodías intermedias eran más adecuadas.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una composición en una sola pista especificando el inicio y el final.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía definiendo solo el inicio y el final en formato XML desde algún software de composición.
- El sistema decodifica la características melódico-rítmicas del archivo.
- El sistema genera la pista de la melodía conectando las partes de inicio y el final.
- El usuario entonces puede realizar modificaciones a la pista resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.6. CoCoNet

CoCoNet (Huang et al., 2019), es un sistema capaz de generar armonizaciones a 4 voces al estilo de bach, las cuales pueden ser condicionadas por el usuario al imponer una melodía inicial, de no estar condicionada, la generación será autónoma.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** 382 Armonizaciones de Bach.
- **Estética:** Su estética se infiere a partir de la base de datos, representando una armonización lo más similar posible a una realizada por Bach.
- **Representación interna:** Los ejemplos de armonizaciones de la base conocimiento se encuentran representados en formato MIDI el cual se decodifica posteriormente a un formato Piano-Roll.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Convolutacional.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que elegían que ejemplos de armonizaciones eran más parecidas al estilo de Bach.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía inicial.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato MIDI o Piano-Roll desde algún software de composición.
- El sistema decodifica las condiciones de notas del archivo.
- El sistema genera la pista de la melodía manteniendo las notas impuestas por el usuario.
- El usuario entonces puede realizar modificaciones a la pista resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.7. MusicVAE

MusicVAE (Roberts et al., 2019), es un sistema capaz de generar pequeñas melodías de forma autónoma, realizar interpolaciones, es decir una transformación continua de una primera melodía en una segunda melodía, ambas impuestas por el usuario y finalmente generar los acompañamientos de una melodía a 3 voces: melodía, bajo y percusiones.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** 1 millón de archivos MIDI públicos en internet.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato MIDI.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.

- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que elegían que ejemplos de composiciones les parecían más musicales.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía inicial para generar un acompañamiento o dos melodías para generar una interpolación.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato MIDI desde algún software de composición o un par de melodías para la interpolación.
- El sistema decodifica las condiciones de melodías impuestas por el archivo.
- El sistema genera las pistas de acompañamientos o una sola pista que contiene la interpolación de las melodías.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.8. Morphing GTTM

Morphing GTTM (Hamanaka et al., 2015), es un sistema capaz de generar melodías de forma autónoma y realizar interpolaciones entre dos melodías, ambas impuestas por el usuario.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Conjunto de reglas basadas en la teoría de generación de música tonal (GTTM) con las cuales se realizan transformaciones, generaciones y extensiones a composiciones musicales.
- **Estética:** La estética está alineada con la teoría GTTM y se evalúa con el cumplimiento de dicho conjunto de reglas.
- **Representación interna:** Las composiciones a transformar o generar están en formato XML.
- **Evaluación de la representación interna:** El producto es evaluado en el cumplimiento del conjunto de reglas.
- **Conceptualización:** El conjunto de reglas define como construir una estructura de datos generativa a partir de una melodía existente, durante dicha construcción se internaliza las características rítmico-temporales de la composición.
- **Generación:** Se realiza un recorrido de la estructura resultante de la melodía aplicando aún otras reglas del conjunto original.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con un experimentado musicólogo el cual corregía las composiciones resultantes de ser necesario.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía inicial para generar una continuación o dos melodías para generar una interpolación.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato XML desde algún software de composición o un par de melodías para la interpolación.
- El sistema construye una estructura de datos generativa a partir de las melodías.
- El sistema la pista que contiene la interpolación de las melodías.
- El usuario entonces puede realizar modificaciones a la pista resultante en algún software de composición musical.

- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.1.9. ChordLSTM

ChordLSTM (Lim et al., 2017), es un sistema capaz de generar acordes que incluyan a las notas presentes en una melodía proporcionada por el usuario.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** 2,252 composiciones en formato XML del sitio Wikifonia.org.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato XML.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un proceso de inferencia en el modelo similar al realizar una predicción en un clasificador.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Una encuesta en internet donde a 25 usuarios sin entrenamiento musical (12 mujeres y 13 hombres) juzgaron que acordes eran más adecuados para una melodía dada, entre los cuales se encontraban los generados por el sistema y los que pertenecen a la canción en el conjunto de datos original.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía inicial para generar una progresión de acordes.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato MIDI desde algún software de composición.
- El sistema decodifica la condición de melodía impuesta por el archivo.
- El sistema genera la pista con los acordes.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.2. Basados en iniciativa mixta

En esta sección se encuentran los sistemas que pro-activamente contribuyen a la creación del artefacto no necesariamente en la misma proporción o de manera simétrica. Un caso particular es la contribución por turnos explícitos en el cual se describirán otros sistemas más adelante.

2.4.2.1. CIM

CIM (Brown, 2018), es un sistema capaz de generar las pistas correspondientes a la pareja en una improvisación a dueto de piano.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Cuenta con reglas específicas que regulan su tempo y la distribución con la que realiza muestreos de su repertorio dinámico.
- **Estética:** Su estética está implícita en su repertorio dinámico pues este depende del gusto personal del otro miembro del dúo.
- **Representación interna y Representación externa:** Las piezas musicales del otro miembro del dúo y el repertorio dinámico están en un formato ABC, el cual permite una renderización inmediata y la evaluación se realiza directamente en la renderización.
- **Evaluación de la representación interna y la representación externa:** 6 músicos expertos en improvisación exploraron al sistema por horas en acompañamiento con la guía de los autores con la finalidad de obtener su retroalimentación.

- **Conceptualización:** El sistema con las reglas de muestreo y obtención del repertorio dinámico, realiza grabaciones de la pareja en el dueto mientras interpreta las piezas generadas. El repertorio dinámico en un principio está vacío y durante la improvisación se añaden frases y estructuras extraídas de las grabaciones de su pareja, las cuales son reproducidas acorde a sus reglas iniciales.
- **Generación:** Se realiza un proceso donde se alterna un muestreo aleatorio y reflexivo sobre el repertorio dinámico.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar las reglas iniciales de tempo y muestreo del repertorio dinámico.
- Realizar una improvisación en piano.

Interacción co-creativa

- El usuario realiza una improvisación en piano.
- El sistema graba extractos de la improvisación del usuario mientras interpreta una selección de su repertorio.
- El sistema incluye las grabaciones a su repertorio.
- El usuario concluye su improvisación.
- El sistema concluye su improvisación.

2.4.2.2. Impro-Visor

Impro-Visor (Johnson et al., 2017), es un sistema capaz de generar melodías a manera de improvisación con las condiciones impuestas por una progresión de acordes dada por el usuario.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Propio conjunto de 200 composiciones de Jazz en formato LeadSheet.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible a las presentes en la base de conocimiento.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato LeadSheet.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.

- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo de las distribuciones de probabilidad obtenidas del entrenamiento generando una parte a la vez y utilizando las últimas generaciones para condicionar las siguientes generaciones.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con un experimentado Jazzista para obtener su retroalimentación durante improvisaciones.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar progresiones de acordes en los cuales se realizará la improvisación.
- Definir los parámetros que determinan el nivel de riesgo, descansos, notas sostenidas, etc.
- Tocar durante la improvisación el tiempo requerido.

Interacción co-creativa

- El usuario proporciona las condiciones iniciales para el sistema.
- El sistema incorpora los parámetros especificados para condicionar las salidas que genere.
- El usuario entonces puede comenzar y mantener una improvisación.
- El sistema podrá entonces mantener de manera alternada un proceso de grabación del usuario y de interpretación.
- Finalmente ambos concluyen la improvisación.

2.4.2.3. JamSketch

JamSketch (Kitahara et al., 2018), es un sistema capaz de generar una melodía usando algoritmos genéticos que siga el contorno dibujado en la representación gráfica de Piano-Roll, donde el eje X es el tiempo y el eje Y es la nota musical.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Base de datos Weimar de Jazz, consiste en 456 composiciones guardadas como descripciones de eventos en formato SQL.

- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible a las presentes en la base de conocimiento.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados como eventos descritos en formato SQL, de igual forma se realiza una re-interpretación en formato Piano-Roll.
- **Evaluación de la representación interna:** Se realiza la evaluación de un modelo probabilista N-gram como función de fitness en el algoritmo genético.
- **Conceptualización:** Como parte de un modelo probabilista N-gram se sigue un proceso de entrenamiento que define la distribución de probabilidad al analizar los ejemplos en la base de conocimiento, además de contar con los componentes elementales para generar las composiciones que fungirán como miembros de la población.
- **Generación:** Se realiza un proceso de evolución iterativa en el marco de los algoritmos genéticos con el fin de que en cada iteración la nueva población mejore su puntaje en la función de fitness hasta llegar a la convergencia.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Se les pidió a 12 investigadores y estudiantes en áreas relacionadas a la música evaluar ejemplos de composiciones del sistema evaluando las siguientes métricas del 0 al 10:
 - Calidad en general.
 - Notas adecuadas.
 - Ritmo adecuado.
 - Variedad melódica.
 - Similitud al Blues.
 - Muestra de habilidad.
 - Similitud humana.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar un contorno en tiempo real a seguir por el sistema para generar la melodía.
- Realizar ediciones pertinentes a la pieza final.
- Establecer la condición de paro.

Interacción co-creativa

- El usuario dibuja un contorno en la pantalla táctil del sistema.
- El sistema decodifica la estructura temporal y armónica del dibujo.
- El sistema genera las notas correspondientes que se adecuen al contorno y las reproduce en tiempo real.
- Finalmente si el usuario decide cuando terminar y obtener la composición final en un formato editable.

2.4.3. Basados en turnos explícitos

En este caso particular de la iniciativa mixta la contribución de los agentes se realiza a través de turnos definidos.

2.4.3.1. DeepBach

DeepBach (Hadjeres et al., 2017), es un sistema capaz de generar composiciones a 4 voces al estilo de Bach de forma autónoma o una armonización a una melodía proporcionada por el usuario.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Conjunto de armonizaciones “BWV 250-438” que consiste en 188 composiciones de Bach y están presentes en la paquetería de software music21.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible a las de Bach.
- **Representación interna:** Los ejemplos de composiciones de la base de conocimiento se encuentran representados en un formato especial que codifica las características melódico-rítmicas.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.

- **Evaluación de la representación externa:** Evaluación empírica con usuarios que elegían que ejemplos de composiciones les parecían más similares a Bach.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía inicial para generar una armonización.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato XML desde algún software de composición.
- El sistema decodifica las condiciones de melodías impuestas por el archivo y las características melódico-rítmicas.
- El sistema genera las pistas del resto de las voces para completar una armonización a 4 voces.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas melodías en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.3.2. MidiNet

MidiNet (Yang et al., 2017), es un sistema capaz de generar melodías que continúen una melodía previa dada por el usuario y es capaz también de generar 2 pistas para ambas manos de piano condicionadas por la melodía previa y una progresión de acordes. Esta última puede ser arbitrariamente modificada en que tan estricto ser con respecto a la estructura de acordes pudiendo explorar otras opciones con el riesgo de perder coherencia.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Conjunto de 1,022 archivos MIDI de música pop del sitio TheoryTab.

- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible a las presentes en la base de conocimiento.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato MIDI.
- **Evaluación de la representación interna:** Al usar un modelo GAN el sistema evalúa sus productos en la etapa de entrenamiento con un discriminador, el cual lo clasificará como real o falso de acuerdo a que tan similar es con respecto a los que se encuentran en la base de conocimiento.
- **Conceptualización:** Un modelo GAN como parte del entrenamiento o aprendizaje extrae las características de los ejemplos para crear uno nuevo a través de su generador. El nuevo ejemplo es ahora clasificado por el discriminador como real o falso, si acierta entonces el generador ajusta sus parámetros para poder engañar al discriminador y en caso contrario el discriminador ajusta sus parámetros para clasificar correctamente el siguiente ejemplo.
- **Generación:** El generador consiste de una red neuronal convolucional (CNN) que toma como entrada un vector de condiciones inferido de lo que ingreso el usuario o generadas aleatoriamente. La salida de la red será la nueva composición al ser decodificada.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que evaluaron los ejemplos en que tan placenteros, reales e interesantes eran.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía para que el sistema la continúe y opcionalmente una progresión de acordes para generar las 2 pistas para el piano.
- Definir el nivel de libertad a errores que le permitan al sistema tomar riesgos para explorar el espacio de composiciones.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato MIDI desde algún software de composición y opcionalmente una progresión de acordes.
- El sistema decodifica las condiciones de melodías impuestas por el archivo y opcionalmente la progresión de acordes.

- El sistema genera las pistas correspondientes a ambas manos del piano si fue proporcionada la progresión de acordes o una melodía simple en caso contrario.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir esa última melodía en el sistema o solicitar generaciones autónomas para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.3.3. LakhNES

LakhNES (Donahue et al., 2019), es un sistema capaz de generar múltiples pistas que continúen una melodía previa dada por el usuario al estilo de composiciones musicales de videojuegos del sistema NES.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Conjunto de 46 horas de música para el sistema NES que pertenecen al “NES Music Database” (NES-MDB) y 775mil archivos MIDI del Lakh MIDI dataset.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible a las presentes en la base de conocimiento.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato MIDI.
- **Evaluación de la representación interna:** Se realiza la evaluación de la verosimilitud de los datos con respecto al análisis probabilístico en la arquitectura de redes neuronales “Transformer”.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo de las distribuciones de probabilidad obtenidas del entrenamiento generando una parte a la vez y utilizando las últimas generaciones para condicionar las siguientes generaciones.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Prueba de Turing al pedirle a usuarios encuestados determinar si las composiciones fueron hechas por un humano o una computadora, así como preguntar cuál de las dos prefieren.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una melodía para que el sistema la continúe.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía en formato MIDI desde algún software de composición.
- El sistema decodifica las condiciones de melodías impuestas por el archivo.
- El sistema genera las pistas correspondientes a 4 voces al estilo NES.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir esa última melodía en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.3.4. GrooVAE

GrooVAE (Gillick et al., 2019), es un sistema capaz de generar una pista de percusiones dada una pista que describa las propiedades rítmicas deseadas. De igual forma el sistema es capaz de rellenar una pista de percusiones, incluyendo más componentes de las percusiones como platillos, toms, caja y bombos.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Groove MIDI Dataset (GMD), contiene 13.6 horas, 1,150 archivos MIDI y más de 22,000 compases de interpretaciones con percusiones.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato MIDI.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.

- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que elegían que ejemplos de composiciones les parecían más humanas.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una pista con la estructura rítmica deseada o una pista de percusiones ya hecha en algún software de composición.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la estructura rítmica o la pista de percusiones a rellenar.
- El sistema decodifica las condiciones rítmicas impuestas por el archivo independientemente de cual tipo sea.
- El sistema genera la pista de percusiones básicas si es que no están presentes en la pista original y en caso contrario rellena la pista con el resto de los instrumentos de percusión.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas pistas en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.3.5. Folk-RNN

Folk-RNN (Sturm et al., 2016), es un sistema capaz de generar una pista con una melodía la cual continúe a la melodía proporcionada por el usuario.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** 23,636 composiciones descargadas del sitio “The session” en formato ABC.
- **Estética:** Su estética se infiere a partir de la base de datos, representando composiciones lo más similares posible.
- **Representación interna:** Los ejemplos de composiciones de la base conocimiento se encuentran representados en formato MIDI.
- **Evaluación de la representación interna:** Evaluación de verosimilitud en una Red Neuronal Recurrente.
- **Conceptualización:** Como parte de un modelo de aprendizaje profundo se sigue un proceso de entrenamiento que define las distribuciones de probabilidad al analizar los ejemplos en la base de conocimiento.
- **Generación:** Se realiza un muestreo a las distribuciones de probabilidad obtenidas del entrenamiento.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con usuarios que expresaron sorpresa y diversión en su retroalimentación después de usar el sistema.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar una pista con la melodía que continuará el sistema.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.
- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona una pista con la melodía a continuar por el sistema.
- El sistema decodifica las condiciones tonales impuestas por el archivo.
- El sistema genera la pista de la melodía condicionada por la entrada del usuario.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas pistas en el sistema para continuar el desarrollo del producto.

- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.3.6. MorpheuS

MorpheuS (Herremans y Chew, 2019), es un sistema capaz de generar composiciones musicales a múltiples voces dado un perfil de tensiones o una plantilla basada en la teoría musical occidental.

Componentes del compositor musical co-creativo artificial

- **Base de Conocimiento:** Reglas de generación con base en plantillas y perfiles de tensión.
- **Estética:** Su estética está implícita en las reglas que definen el “gusto” del sistema dada una plantilla o perfil de tensiones, así como también está definida por la función de optimización en el algoritmo de generación.
- **Representación interna:** Sus composiciones pueden ser posteriormente editadas en algún software de edición musical al estar codificadas en formato MIDI.
- **Evaluación de la representación interna:** Al utilizar una función de optimización el puntaje que esta otorgue a cada producto será su correspondiente evaluación.
- **Conceptualización:** Dependiendo de la definición de las reglas se desarrollará la función de optimización la cual a través de las iteraciones que se lleven a cabo hasta mejorar su puntaje, integrará la estética implícita en las reglas.
- **Generación:** Se realiza un proceso de iteraciones en la que cada generación buscará como mejorar el puntaje máximo de la función de optimización.
- **Representación externa:** Es el renderizado de la composición, es decir la reproducción sonora de la composición.
- **Evaluación de la representación externa:** Evaluación empírica con varios expertos marcando el humor detrás de la ingenuidad del sistema al cometer violaciones a las convenciones musicales en las que fue construido, con ocasionales sorpresas y resultados viables resultado de una búsqueda arriesgada de cosas nuevas.

Responsabilidades del agente humano en el proceso co-creativo

- Proporcionar el perfil de tensiones o una plantilla de composición.
- Realizar ediciones pertinentes a la pieza en general y al resultado de interactuar con el sistema.

- Establecer alguna condición de paro u objetivo que completar.

Interacción co-creativa

- El usuario proporciona el perfil de tensiones o la plantilla a seguir por el sistema.
- El sistema incorpora las condiciones del perfil o la plantilla en su función de optimización.
- El sistema genera las pistas correspondientes a ambas manos en una composición para piano después de mejorar su puntaje iterativamente.
- El usuario entonces puede realizar modificaciones a las pistas resultante en algún software de composición musical.
- De las modificaciones realizadas el usuario puede volver a introducir nuevas pistas en el sistema para continuar el desarrollo del producto.
- Finalmente si el usuario no realiza modificaciones puede determinar el punto de paro de la interacción con el producto terminado.

2.4.4. Resumen

En este capítulo se realizó una breve introducción a los sistemas computacionales co-creativos, inicialmente se mencionó la definición acuñada en la conferencia internacional en creatividad computacional como un esfuerzo multidisciplinario entre las ciencias y las artes con 3 objetivos, simular la creatividad en un sistema computacional, mejorar nuestro entendimiento de la creatividad humana y finalmente aumentar la misma. Poder cuantificar el trabajo realizado en esta área es un requerimiento fundamental para comparar y avanzar efectivamente en nuevos retos científicos del área. Se incluyó también una organización de los sistemas co-creativos dependiendo de sus características que busca direccionar dichos esfuerzos de mejora. La evaluación de los sistemas creativos es incluida por los autores sin embargo no hay un consenso general entre la comunidad de investigación por lo que se muestran algunos ejemplos de las preguntas que surgen para los sistemas creativos, “¿Quién está evaluando?”, “¿Qué se está evaluando?”, “¿Cuándo ocurre la evaluación?”, “¿Cómo se realiza la evaluación?”. De las características que pueden ser evaluadas se eligió la novedad al ser incluida como característica fundamental de manera común en trabajos de revisión interna en el campo de la creatividad computacional.

El capítulo termina con una muestra no exhaustiva de compositores musicales co-creativos, mencionando a grandes rasgos su clasificación, su proceso creativo, su interacción con el usuario y sus métodos de evaluación. Aunque los métodos de evaluación son muy variados entre los distintos compositores documentados, podemos observar que sus metodologías de evaluación son elegidas en su mayoría en el análisis estadístico de datos abstractos que han demostrado ser eficientes para generar productos similares a su conjunto de inspiración. Sin embargo, estos métodos ignoran detalles del dominio musical al ser de propósito general con el objetivo de poder usarse en otros dominios como el procesamiento de lenguaje natural o reconocimiento de imagen. Los sistemas presentados generalmente carecen de una evaluación extendida con detalles propios del dominio musical o de juicio de su creatividad. Esta carencia sugiere una accesible y extensa área de oportunidad de traer métodos de evaluación musical a estos sistemas co-creativos. Dichos métodos de evaluación musical serán explorados en el siguiente capítulo.

Capítulo 3

Evaluación de la novedad musical

Podemos utilizar una definición de novedad que dependa de una métrica de similitud entre piezas musicales para poder determinar la distancia de productos nuevos con respecto a su espacio conceptual. En este capítulo se presentan las medidas utilizadas basadas en el contenido musical, divididas por la representación que utilizan del contenido, ya sea una representación auditiva como un archivo wav o mp3, o una representación simbólica como es un archivo MIDI que codifica una partitura. A continuación, se muestran las medidas más relevantes en cada representación, así como en su correspondiente categorización interna.

3.1. Propuestas generales

3.1.1. Propuesta de Alison Pease

Novedad De acuerdo con Alison Pease Pease et al. (2001), el conocimiento explícito de un programa puede ser la base de datos que utiliza, y el conocimiento implícito es, por ejemplo, el conocimiento del programador con el cual definió reglas de comportamiento. Se hará referencia al conjunto del conocimiento explícito e implícito cómo el conjunto de inspiración I . De igual forma se hará referencia al conjunto de productos generados por el programa como el conjunto R . Un elemento x para ser considerado nuevo requiere existir en $R \cap I$, sin embargo, no es suficiente.

- **Novedad relativa a un campo de conocimiento**, si el elemento x esta en una zona del campo de conocimiento bien explorada *no es nuevo*, si esta en una zona poco explorada pero dentro de los límites es *meramente nuevo*, si esta fuera, pero cercano a los límites es *fundamentalmente nuevo* y si esta fuera y demasiado lejos entonces es *demasiado caótico para ser considerado nuevo*.

Formalmente, se define al programa P como el conjunto de procedimientos k , que

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

son procedimientos para generar artefactos, procedimientos para generar y evaluar procesos de generación y evaluación de artefactos, estrategias de búsqueda, datos de entrada, etc. Realicemos las siguientes definiciones de los conjuntos O_x , M_x , y el saco H_x (Un saco a diferencia de un conjunto puede tener repeticiones).

$$H_x = [k \geq P: k \text{ fue utilizado para generar al artefacto } x]$$

H_x no es único, puesto que pueden existir diversas formas de generar al mismo elemento x .

$$O_x = \{k \geq P: k \text{ es usado directamente en la generación o evaluación de } x\}$$

$$M_x = \{k \geq P: k \text{ es usado para generar o evaluar un elemento de } O_x\}$$

Si un elemento x fue creado por un saco H_x y en dicho saco hay procedimientos k que también se encuentran en M_x , y no existe otro saco H_x^0 que no comparta ningún procedimiento k con M_x , entonces el elemento x es *fundamentalmente nuevo*. Esto es porque al forzosamente requerir el uso de un procedimiento k en M_x y los elementos de M_x generan o evalúan procesos de generación de artefactos, quiere decir que se tuvo que realizar alguna exploración de forma externa del campo de conocimiento. Formalmente:

$$\text{novedad}(x) = \text{fundamental si } H_x \setminus M_x \neq \emptyset \text{ ; y } \nexists H_x^0 \text{ tal que } H_x^0 \setminus M_x = \emptyset \text{ ;}$$

Si un elemento x fue creado por un saco H_x y en dicho saco no hay procedimientos k que se encuentren en M_x , entonces el elemento es *meramente nuevo*, es decir, fue generado usando únicamente procedimientos a nivel objeto. Formalmente:

$$\text{novedad}(x) = \text{meramente nuevo si } H_x \setminus M_x = \emptyset \text{ ;}$$

Finalmente, la novedad de x es *nula* en cualquier otro caso.

- **Novedad relativa a la complejidad**, un elemento es considerado creativo si **(i)** el espacio conceptual en el que yace es grande y complejo y **(ii)** es generado de un área poco explorada.

- **(i)** Dado un programa P , definimos a su espacio conceptual como todos los artefactos x que, en principio, puede generar P .

$$CS_P = \{x: x \text{ puede ser generado por } P\}$$

Así una primera medida de complejidad (i) es:

$$\text{complejidad}_{i1} = |CS_P|$$

Otra medida es usar el *espacio conceptual restringido* el cual corresponde a todos los artefactos x^0 que pueden ser generados con menor o igual cantidad de conocimiento utilizado para generar a un artefacto x .

$$RCS_1(x) = \{x^0 \geq CS_P: |H_{x^0}| \leq |H_x|\}$$

Entonces:

$$\mathbf{complejidad}_{i2} = jRCS_1(x)j$$

Alternativamente se puede definir el conjunto $RCS_2(x)$ de todos los artefactos x^j que pueden ser generados con un subconjunto del conocimiento utilizado para generar a un artefacto x .

$$RCS_1(x) = f x^j \supseteq CS_P : H_{x^j} \quad H_x g$$

Entonces:

$$\mathbf{complejidad}_{i3} = jRCS_2(x)j$$

- (ii) Se revisan a continuación medidas que capturan exactamente que artefactos yacen en un área poco explorada. Definimos al $saco_R$ y al $conjunto_R$ como el saco y el conjunto de todas las piezas de conocimiento k usadas para generar a todos los artefactos x en R :

$$\begin{aligned} saco_R &= B_{x \supseteq R} H_x \\ conjunto_R &= \bigcup_{x \supseteq R} H_x \end{aligned}$$

Ahora definimos la medida de frecuencia $f(k)$ para cada $k \supseteq saco_R$

$$f(k) = \text{número de ocurrencias de } k \text{ en } saco_R$$

Con lo anterior definimos dos medidas, la *frecuencia promedio*($saco_R$) = $af(saco_R) = \frac{jsaco_Rj}{jconjunto_Rj}$ (la cual obtiene el número promedio de veces que un elemento k es usado para la generación de todos los artefactos x en R) y la *frecuencia relativa*(k) = $rf(k) = \frac{f(k)}{af(saco_R)}$ (el número de veces que un k específico se usó para la generación de todos los artefactos x en R comparado con el promedio).

Si $rf(k_i)$ devuelve un valor menor a 1 entonces k_i es usado en menor frecuencia que el promedio, y si es mayor a 1 entonces es usado con mayor frecuencia.

Para medir la complejidad definamos el conjunto que contiene a las piezas de conocimiento k que no son usualmente usadas: $Raro = fk : rf(k) < 1g$,

Entonces:

$$\mathbf{complejidad}_{ii1}(x) = jH_x \setminus Raroj$$

El número de piezas k usadas para generar a x y son raras.

De igual forma, sumamos las frecuencias relativas de todas las k usadas para generar a x (H_x), donde mientras más grande sea la suma, menor será la complejidad de x .

$$\mathbf{complejidad}_{ii2}(x) = \frac{1}{rf(x)} H_x$$

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Una tercera medida para la complejidad (ii) es tomar la pieza k con la menor frecuencia relativa de entre todas las $k \in H_x$.

$$\text{complejidad}_{ii3}(x) = \frac{1}{\min(\text{rf}(H_x))}$$

Finalmente, la **novedad** puede medirse de manera binaria en donde tanto la complejidad (i) y la complejidad (ii) sean mayores a valores arbitrarios α y β :

novedad1(x) = 1 si $\text{complejidad}_i(x) > \alpha$ y $\text{complejidad}_{ii}(x) > \beta$ de otra forma.

Alternativamente se puede usar una función que use ambas:

$$\text{novedad2}(x) = \text{complejidad}_i(x) + \text{complejidad}_{ii}(x)$$

A la cual se le puede integrar la cantidad de conocimiento como factor $\int H_x$.

- **Novedad relativa a un arquetipo**, el conjunto de *tipicalidad* T se basa en un vector ponderado de propiedades que caracterizan a los artefactos en el dominio, pueden ser cuantitativas, cualitativas, subjetivas u objetivas. Tomemos ahora el conjunto resultante de $R \cap T$. Sea $R \cap T(x)$ el grado con el que x pertenece a $R \cap T$. $R \cap T(x) \in [0; 1]$ donde 0 = *completamente familiar* y 1 = *completamente nuevo*. Así definimos la primera medida de novedad:

$$\text{novedad1}(x) = R \cap T(x)$$

Definamos al conjunto $A = \{a : a \text{ es un arquetipo de un tipo en particular}\}$, entonces:

$$\text{novedad2}(x) = \text{mínima}(\text{distancia}(x; a)) \text{ sobre } a \in A$$

- **Novedad como sorpresa**, un artefacto fundamentalmente nuevo en algún momento de su generación sorprendió a su creador, por lo que es relevante en mostrar cuando un artefacto *no* es fundamentalmente nuevo (puesto que no es sorprendente), sin embargo, no puede usarse como único factor para determinar su novedad fundamental. La medida de sorpresa junto con alguna medida de novedad previa puede ser una herramienta poderosa.

La medida de sorpresa es dependiente de las probabilidades individuales de un evento que es clasificado como similar a x . Siendo un evento menos probable más sorprendente al suceder. Se define como:

$$\text{sorpresa}(x) = 1 - \frac{\sum_{y \in Y} \text{sim}(x; y) \cdot \text{prob}(y)}{\sum_{y \in Y} \text{sim}(x; y)}$$

Donde Y es el conjunto de eventos comparables y $\text{sim}(x; y)$ es la similaridad entre x y y .

- **Novedad percibida**, Es complicado formalizar una medida de novedad, sin embargo, los humanos son jueces capaces de distinguir algo nuevo, aburrido o caótico. Se pueden realizar pruebas experimentales donde se les pida a los participantes:
 - (i) Distinguir elementos que pertenecen a I de los que pertenecen a R .
 - (ii) Cuantificar el grado de variación entre todos los elementos que pertenecen a I y de igual forma de entre los elementos que pertenecen a R .
 - (iii) Cuantificar el grado de variación entre un par de elementos, ambos pertenecientes a I ; y también de un par de elementos donde uno pertenezca a I y el otro a R .

Los elementos de R son considerados nuevos si:

- (i) Los participantes son incapaces de identificarlos con seguridad como pertenecientes a R . La proporción media de los elementos clasificados como de R es menor a 0.5.
- (ii) No hay diferencia significativa en la media percibida de variación entre los subconjuntos de I y los subconjuntos de R .
- (iii) No hay diferencia significativa en la media percibida de variación entre las parejas de I y las parejas de elementos combinados de I y R .

3.1.2. Criterios empíricos de Ritchie

Ritchie (2007) propone 18 métricas con las que se pretenden capturar las propiedades con las cuales se puede emitir un juicio metodológicamente en la atribución de la creatividad a un programa computacional. Son factores empíricamente observables buscando ser lo más similares posible a los factores que observaríamos en un humano, por ejemplo, el peso que tiene el proceso de generación es reducido ya que en el proceso creativo ejercido por una persona no es algo que podamos observar tan fácilmente. Los programas contemplados poseen las siguientes características:

- Hay una clase de artefactos (usualmente definida por la cultura) en la que el sistema genera sus salidas. Esta clase existe antes que el programa y no se define por los procesos realizados por el programa.
- La clase es, en principio, extremadamente grande, posiblemente infinita.
- Dado un elemento (ya sea generado por un humano o un programa), no hay definición precisa sobre si es perteneciente a la clase ya sea objetiva o subjetiva o discrepante entre jueces.
- Dado un elemento un juez humano puede subjetivamente otorgar un puntaje de calidad.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Ahora, de igual forma se toma en cuenta los dos componentes esenciales para la atribución de creatividad, acorde con la literatura revisada, siendo *novedad* y *calidad*. Sin embargo, sugiere el uso de la *tipicalidad* puesto que es de mayor relevancia en los programas computacionales el poder determinar el grado con el que un artefacto es perteneciente a la clase e inferir la novedad asociada con la baja tipicalidad de algún elemento. La evaluación de tipicalidad y calidad será de manera subjetiva por jueces humanos.

Las definiciones básicas para los criterios son:

- La *clase* $(B; typ; val)$, cual es una tripleta de un conjunto de elementos básicos (B) , y esquemas de evaluación para la tipicalidad y valor respectivamente typ y val .
- El conjunto de inspiración I , el cual tiene los artefactos que de manera explícita o implícita influyeron o influyen en la capacidad de creación del programa, se toma en cuenta el caso cuando el conjunto es vacío.
- El programa a evaluar P .
- Los resultados R , que son las salidas del programa durante la ejecución en la que será evaluado.

Describamos la notación utilizada a lo largo de los criterios, dado un conjunto de elementos X definimos 3 conjuntos:

- $T; (X) = \{x \in X \mid typ(x) \in [g, h]\}$, el subconjunto de X que esta en un rango de tipicalidad entre dos valores arbitrarios g y h .
- $V; (X) = \{x \in X \mid val(x) \in [g, h]\}$, el subconjunto de X que esta en un rango de calidad entre dos valores arbitrarios g y h .
- $AV(F; X) = \frac{\sum_{x \in X} F(x)}{|X|}$, el valor promedio de una función F en el conjunto finito X .
- $ratio(X; Y) = \frac{|X \cap Y|}{|Y|}$, el tamaño relativo de dos conjuntos finitos X, Y , donde $Y \subseteq X$.

Los criterios son los siguientes:

1. Los productos son razonablemente típicos. $AV(typ; R) > \alpha$, aquí α es un valor arbitrario indicando a qué punto se considera que los elementos han alcanzado una tipicalidad aceptable.
2. Una proporción razonable de los resultados es típica. $ratio(T_{\alpha}(R); R) > \beta$. En este criterio se definen α y β . La expresión $T_{\alpha}(R)$ se toma como “artefactos que cumplen con la definición establecida o las normas de la clase”.
3. En promedio los productos son de calidad. $AV(val; R) > \gamma$, aquí γ es un valor arbitrario indicando a qué punto se considera que los elementos han alcanzado una calidad aceptable.

4. Una proporción razonable de los resultados son de calidad. $ratio(V_{.1}(R); R) > .$
Proponiendo valores adecuados para α y β .
5. Una proporción razonable de los resultados típicos, es también de calidad.
 $ratio(V_{.1}(R) \setminus T_{.1}(R); T_{.1}(R)) > .$ Proponiendo valores adecuados para la
calidad α , la tipicalidad β y la expresión en general γ .
6. Una proporción razonable de todos los resultados es atípica y de calidad.
 $ratio(V_{.1}(R) \setminus T_0; (R); R) > .$, proponiendo valores adecuados para la tipicalidad
 β , la calidad α y la expresión en general γ .
7. Una proporción razonable de los resultados atípicos es también de calidad.
 $ratio(V_{.1}(R) \setminus T_0; ; T_0; (R)) > .$, proponiendo valores adecuados para la tipica-
lidad β , la calidad α y la expresión en general γ .
8. Una proporción razonable de los elementos de calidad, es también atípica.
 $ratio(V_{.1}(R) \setminus T_0; (R); V_{.1}(R))$, proponiendo valores adecuados para la tipi-
calidad β , la calidad α y la expresión en general γ .
9. Una proporción razonable del conjunto de inspiración aparece en los resultados.
 $ratio(R; I \setminus R) > .$, proporcionando un valor adecuado para α .
10. Una proporción razonable de los resultados no está en el conjunto de inspiración.
 $1 - ratio(I \setminus R; R) > .$, proporcionando un valor adecuado para α .
11. En promedio los resultados nuevos son típicos.
 $AV(typ; (R \setminus I)) > .$, para un valor adecuado de α .
12. En promedio los resultados nuevos son de calidad.
 $AV(val; (R \setminus I)) > .$, para un valor adecuado de α .
13. Una proporción razonable de los resultados es nueva y típica.
 $ratio(T_{.1}(R \setminus I); R) > .$, para valores adecuados de tipicalidad (β) y de de la
expresión en general γ .
14. Una proporción razonable de los resultados es nueva y de calidad.
 $ratio(V_{.1}(R \setminus I); R) > .$, para valores adecuados de calidad (α) y de de la
expresión en general γ .
15. Una proporción razonable de los resultados nuevos es también típica.
 $ratio(T_{.1}(R \setminus I); R \setminus I) > .$, para valores adecuados de tipicalidad (β) y de de
la expresión en general γ .

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

16. Una proporción razonable de los resultados nuevos es también de calidad.

$ratio(V_{,1}(R \setminus I); R \setminus I) > \alpha$, para valores adecuados de calidad (α) y de de la expresión en general β .

17. Una proporción razonable de los resultados nuevos es típica y de calidad.

$ratio(V_{,1}(R \setminus I) \setminus T_{,1}(R \setminus I); (R \setminus I)) > \alpha$, para valores adecuados de tipicidad (α), de calidad (β) y de de la expresión en general γ .

18. Una proporción razonable de los resultados nuevos es atípica y de calidad.

$ratio(V_{,1}(R \setminus I) \setminus T_0; (R \setminus I); (R \setminus I)) > \alpha$, para valores adecuados de tipicidad (α), de calidad (β) y de de la expresión en general γ .

El uso de los criterios no debe ser el de componentes de un juicio estricto y rígido para atribuir creatividad al cumplir un subconjunto de ellos. Es más bien un bosquejo de un marco de referencia prematuro, el cual requiere extensión y especificación de acuerdo a la tarea que en la que se desenvuelve el programa. De manera abstracta se busca dar una dirección al evaluar las siguientes propiedades:

- Éxito básico, criterios: 1, 2, 9
- Calidad en general, criterios: 3, 4
- Habilidad convencional, criterio: 5
- Habilidad no convencional, criterios: 6, 7, 8
- Evasión de replicación, criterio: 10
- Éxito básico (evitando replicación), criterios: 11, 13, 15
- Calidad en general (evitando replicación), criterios: 12, 14, 16
- Habilidad convencional (evitando replicación), criterio: 17
- Habilidad no convencional (evitando replicación), criterios: 18

3.2. Medidas de similitud musical

Es importante aclarar la distinción entre la “ distancia ” entre melodías y una medida de similitud, ya que la primera implica que los valores más bajos indican una mayor semejanza y las medidas de similitud son las inversas. Si bien en esta sección se presenta el cálculo de cada medida en su forma original, por consistencia en el resto del trabajo se modifican las medidas de similitud para que presenten sus resultados en un estilo tipo distancia (a menor valor, mayor semejanza).

Sea \mathbf{A} una melodía, donde $\mathbf{A}[i]$ es la i ésima nota en esa melodía. La similitud entre dos melodías \mathbf{A} y \mathbf{B} , de longitud n , se puede obtener a partir de la distancia entre estas dos secuencias de notas. A continuación describimos algunas medidas de distancia utilizadas popularmente en el ámbito musical.

3.2.1. Basadas en contenido de audio

Las principales características que se toman en cuenta en este tipo de medidas son las siguientes:

Grupo	Descriptorios
Timbre	Escala Bark, Tono, Análisis Espectral y Contenido frecuencial.
Ritmo	Beats por minuto y detección de eventos.
Tonalidad	Temperamentos, disonancia y armonía.
Otros	Amplitud, intensidad y cruces por cero.

Tabla 3.1: Características en medidas basadas en contenido de audio.

3.2.1.1. Distancia Euclidiana en espacios multidimensionales usando PCA y RCA

Descripción: En el trabajo de Bogdanov et al. (2011), se lleva a cabo una comparativa de distintas medidas de similitud musical. Además de las medidas que proponen se incorporan 2 medidas utilizando métodos de aprendizaje máquina como lo son *Principal Component Analysis* (PCA) y *Relevant Component Analysis* (RCA).

Características utilizadas: Bandas de Bark, tono, coeficiente de tonalidad, contenido de altas frecuencias, bandas de energía, inarmonicidad, intensidad por beats, perfiles de clase tonal sin transponer, intensidad promedio y cruce por cero. Además de

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

incluir, con respecto al análisis espectral, su centroide, dispersión, curtosis, rolloff, disminución, asimetría, cresta, complejidad, flujo, energía, pico pronunciado, triestímulo y relación de energía armónica.

Metodología: Se forma un espacio multidimensional donde en cada eje es el valor de cada una de las características presentes. Se realiza una estandarización para que cada eje tenga un rango de [0, 1]. Posteriormente se realiza PCA y RCA para disminuir el número de dimensiones desde 201 hasta 25, obteniendo un 78 % de la varianza original con muchas menos dimensiones. Finalmente, cada pieza musical se ubica en este espacio multidimensional y su similitud es medida calculando su distancia Euclidiana a otra pieza musical. De manera general la distancia euclidiana en n dimensiones se expresa de la siguiente manera:

$$D_{ij}^2 = \sum_{v=1}^n (X_{vi} - X_{vj})^2 \quad (3.1)$$

Evaluación Objetiva: A través de distintas colecciones de piezas musicales organizadas por género musical, emoción, etc. Se realizaron pruebas de relevancia. Es decir, con una pieza musical de consulta se obtienen de la colección las piezas que resulten ser relevantes (similares). Se calcula la precisión media con un rango de 0 a 1, mientras más cercano a 1 es mejor, esto para cada conjunto y se comparó con un selector aleatorio, que al azar seleccionaba las piezas relevantes.

Colección	Categoría de la colección	Selección Aleatoria	PCA	RCA
G2	Género y Cultura	0.16	0.39	0.34
MHA	Emoción	0.53	0.58	0.53
MVI	Instrumentos	0.51	0.56	0.60
ART	Artista	0.02	0.08	0.10
ALB	Album	0.02	0.11	0.16
RBL	Ritmo y tempo	0.15	0.24	0.21

Tabla 3.2: Conjuntos de datos evaluados.

Evaluación Subjetiva: Para la evaluación subjetiva se utilizó una colección privada de piezas musicales de distintos géneros, emociones, instrumentos, artistas y culturas. Usando cada método de similitud se tomó una misma canción semilla para cada uno y se creó una lista de reproducción de 5 canciones. Hubo 12 participantes donde se les cuestionó sobre su experiencia musical y su experiencia en escucha musical. En cada experimento se le proporciona una lista de reproducción al voluntario, este califica

del 0 al 5 que tan adecuada es la lista de reproducción con respecto a la canción semilla. Además, se le pide contestar con si es que considera que hay inconsistencia o no. Posteriormente a los resultados se les agrupa usando el método de análisis ANOVA, en el cual PCA quedó agrupado con la selección aleatoria como los de peor rendimiento y mayor inconsistencia. En la escala de similitud de 0 a 5, en promedio obtuvieron 0.7 para PCA y 0.5 para selección aleatoria. En inconsistencia obtuvieron 0.8 para PCA y 0.9 para la selección aleatoria. RCA no fue utilizado en la evaluación subjetiva al ser demasiado similar a PCA entre las demás metodologías evaluadas y así evitar un gran número de experimentos.

3.2.1.2. Divergencia Kullback-Leibler basada en Mezcla Gaussiana usando Coeficientes Mel

Descripción: Siguiendo con el trabajo de Bogdanov et al. (2011), se lleva a cabo una comparativa de esta medida de similitud musical, la cual mide una diferencia entre dos distribuciones de probabilidad.

Características utilizadas: Los primeros 13 *Mel Frequency Cepstral Coefficients* (Coeficientes Cepstrales en las Frecuencias de Mel) o MFCCs.

Metodología: Se hace una aproximación geométrica cerrada a la divergencia Kullback-Leibler, donde las dos distribuciones de probabilidad a comparar son modelos de mezcla Gaussiana. Dicha aproximación queda definida de la siguiente forma para medir la diferencia entre una canción X y una canción Y :

$$d(X; Y) = Tr\left(\frac{1}{N_X} \sum_{x \in X} x x^T\right) + Tr\left(\frac{1}{N_Y} \sum_{y \in Y} y y^T\right) + Tr\left(\left(\frac{1}{N_X} \sum_{x \in X} x x^T + \frac{1}{N_Y} \sum_{y \in Y} y y^T\right)^{-1} \left(\frac{1}{N_X} \sum_{x \in X} x x^T - \frac{1}{N_Y} \sum_{y \in Y} y y^T\right)\right) \quad (3.2)$$

Donde $Tr()$ es la traza de una matriz, \bar{x} y \bar{y} son las medias de los coeficientes MFCC, Σ_X y Σ_Y son las matrices de covarianza de los coeficientes MFCC y N_{MFCC} es la dimensionalidad de los coeficientes, en este caso son 13 los coeficientes que se tomaron en cuenta.

Evaluación Objetiva: A través de distintas colecciones de piezas musicales organizadas por género musical, emoción, etc. se realizaron pruebas de relevancia. Es decir, con una pieza musical de consulta se obtienen de la colección las piezas que resulten ser relevantes (similares). Se calcula la precisión media con un rango de 0 a 1, mientras más cercano a 1 es mejor, esto para cada conjunto y se comparó con un selector aleatorio, que al azar seleccionaba las piezas relevantes.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Colección	Categoría de la colección	Selección Aleatoria	MFCC
G2	Género y Cultura	0.16	0.43
MHA	Emoción	0.53	0.58
MVI	Instrumentos	0.51	0.61
ART	Artista	0.02	0.15
ALB	Album	0.02	0.24
RBL	Ritmo y tempo	0.15	0.25

Tabla 3.3: Conjuntos de datos evaluados.

Evaluación Subjetiva: Para la evaluación subjetiva se utilizó la serie de experimentos explicados previamente en la metodología anterior. Los resultados se les agrupa usando el método de análisis ANOVA, en el cual MFCC quedó como una de las mejores en rendimiento y con menor inconsistencia. En la escala de similitud de 0 a 5, en promedio obtuvo 3.0. En inconsistencia obtuvo 0.2.

3.2.1.3. Similitud basada en ritmo

Descripción: Siguiendo con el trabajo de Bogdanov et al. (2011), se lleva a cabo una comparativa de una medida que mide la distancia rítmica entre dos canciones usando una combinación lineal ponderada.

Características utilizadas: Beats por minuto (BPM) y *Onset Rate* (OR) o Eventos por segundo.

Metodología: La distancia entre una canción X y una canción Y está definida como una combinación lineal ponderada entre su distancia en BPM y OR:

$$d(X; Y) = w_{BPM}d_{BPM}(X; Y) + w_{OR}d_{OR}(X; Y) \quad (3.3)$$

La distancia en BPM (d_{BPM}) está definida como:

$$d_{BPM}(X; Y) = \min_{i \in N} \left(\frac{1}{BPM} \frac{\max(X_{BPM}; Y_{BPM})}{\min(X_{BPM}; Y_{BPM})} \right) \quad i \quad (3.4)$$

La distancia en OR (d_{OR}) está definida como:

$$d_{OR}(X; Y) = \min_{i \in N} \left(\frac{1}{OR} \frac{\max(X_{OR}; Y_{OR})}{\min(X_{OR}; Y_{OR})} \right) \quad i \quad (3.5)$$

Donde $X_{BPM}, Y_{BPM}, X_{OR}, Y_{OR} > 0$, $w_{BPM}, w_{OR} \in [0, 1]$. Los parámetros w_{BPM}, w_{OR} son el peso que tienen cada distancia. Las dos ecuaciones de distancia están basadas en la suposición de que canciones que tienen múltiplos de OR o BPM son más similares que las que no. El parámetro α influye en que si es igual a 1 todos los múltiplos son considerados por igual, sin embargo, mientras incrementa, se prefiere múltiplos más pequeños. Con la afinación de parámetros utilizando una búsqueda de malla en una colección anotada de piezas musicales se obtuvieron los mejores resultados con $w_{BPM} = 0.5$, $w_{OR} = 0.5$ y $\alpha_{BPM} = \alpha_{OR} = 30$, revelando entonces que tanto BPM como OR son relevantes en igual proporción para determinar la distancia entre dos canciones.

Evaluación Objetiva: A través de distintas colecciones de piezas musicales organizadas por género musical, emoción, etc. Se realizaron pruebas de relevancia. Se calcula la precisión media con un rango de 0 a 1, mientras más cercano a 1 es mejor, esto para cada conjunto y se comparó con un selector aleatorio, que al azar seleccionaba las piezas relevantes.

Colección	Categoría de la colección	Selección Aleatoria	Tempo
G2	Género y Cultura	0.16	0.36
MHA	Emoción	0.53	0.56
MVI	Instrumentos	0.51	0.52
ART	Artista	0.02	0.03
ALB	Album	0.02	0.02
RBL	Ritmo y tempo	0.15	0.44

Tabla 3.4: Conjuntos de datos evaluados.

Evaluación Subjetiva: Para la evaluación subjetiva se utilizó la serie de experimentos explicados previamente. Los resultados se les agrupa usando el método de análisis ANOVA, en la escala de similitud de 0 a 5, en promedio obtuvo 1.2 y en inconsistencia obtuvo 0.7.

3.2.1.4. Similitud basada en clasificadores

Descripción: Siguiendo con el trabajo de Bogdanov et al. (2011), se lleva a cabo una comparativa de una medida que mide la distancia usando modelos previamente entrenados para obtener descriptores de alto nivel.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Características utilizadas: Género musical, cultura, emoción, instrumentación, ritmo y tempo.

Metodología: Contando con modelos entrenados se obtienen los valores de alto nivel correspondientes a cada uno de los siguientes 3 grupos: 1) Género y cultura musical, 2) emoción e instrumentación y 3) ritmo y tempo. La distancia entre una canción X y una canción Y está definida usando la correlación ponderada de Pearson:

$$\text{corr}(X; Y; w) = \rho \frac{\text{cov}(x; y; w)}{\sqrt{\text{cov}(x; x; w) \text{cov}(y; y; w)}} \quad (3.6)$$

La covarianza ponderada está definida como:

$$\text{cov}(x; y; w) = \frac{\sum_i^n w_i (x_i - m(x; w))(y_i - m(y; w))}{\sum_i^n w_i} \quad (3.7)$$

La media ponderada está definida como:

$$m(x; w) = \frac{\sum_i^n w_i x_i}{\sum_i^n w_i} \quad (3.8)$$

Donde $n = 3$ y basándose en estudios de percepción musical determinaron los pesos para los 3 grupos de la siguiente forma: $w_1 = 0.5$, $w_2 = 0.3$ y $w_3 = 0.2$.

Evaluación Objetiva: A través de distintas colecciones de piezas musicales organizadas por género musical, emoción, etc. Se realizaron pruebas de relevancia. Se calcula la precisión media con un rango de 0 a 1, mientras más cercano a 1 es mejor, esto para cada conjunto y se comparó con un selector aleatorio, que al azar seleccionaba las piezas relevantes.

Colección	Categoría de la colección	Selección Aleatoria	CLAS-Pears-W
G2	Genero y Cultura	0.16	0.67
MHA	Emoción	0.53	0.68
MVI	Instrumentos	0.51	0.59
ART	Artista	0.02	0.06
ALB	Album	0.02	0.06
RBL	Ritmo y tempo	0.15	0.35

Tabla 3.5: Conjuntos de datos evaluados.

Evaluación Subjetiva: Para la evaluación subjetiva se utilizó la serie de experimentos explicados previamente. Los resultados se les agrupa usando el método de análisis ANOVA, en la escala de similitud de 0 a 5, en promedio obtuvo 3.0 y en inconsistencia obtuvo 0.3.

3.2.1.5. Similitud Híbrida Múltiple

Descripción: Siguiendo con el trabajo de Bogdanov et al. (2011), se lleva a cabo una comparativa de una medida que mide la distancia usando una combinación lineal de distintos descriptores.

Características utilizadas: Ocupa las mismas características que las métricas usando PCA, MFCCs, ritmo y clasificadores.

Metodología: La distancia entre una canción X y una canción Y está definida como una combinación lineal ponderada entre su distancia en las métricas discutidas en las secciones previas:

$$d(X; Y) = w_{PCA}d_{PCA}(X; Y) + w_{MFCC}d_{MFCC}(X; Y) + w_{TEMPO}d_{TEMPO}(X; Y) + w_{CLAS}d_{CLAS}(X; Y) \quad (3.9)$$

Donde las distancias se normalizaron para estar uniformemente distribuidas entre 0 y 1. Los pesos se determinaron según su resultado en la evaluación subjetiva de cada uno, terminando con los siguientes valores: $w_{PCA} = 0.7$, $w_{MFCC} = 3.0$, $w_{TEMPO} = 1.2$ y $w_{CLAS} = 3.0$.

Evaluación Objetiva: A través de distintas colecciones de piezas musicales organizadas por género musical, emoción, etc. Se agruparon para formar 2 colecciones generales de piezas musicales y se realizaron pruebas de relevancia. Se calcula la precisión media con un rango de 0 a 1, mientras más cercano a 1 es mejor, esto para consultas de una de las colecciones generales a la otra y se comparó con cada una de las métricas individuales que la conforman y con un selector aleatorio, que al azar seleccionaba las piezas relevantes.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Métrica	Colección 1 a Colección 2	Colección 2 a Colección 1
RANDOM	0.07	0.08
PCA	0.09	0.11
MFCC	0.23	0.22
TEMPO	0.11	0.12
CLAS	0.21	0.23
Híbrido	0.25	0.28

Tabla 3.6: Resultados de cada métrica por tarea.

Evaluación Subjetiva: Para la evaluación subjetiva se utilizó la serie de experimentos explicados previamente con la diferencia de que utiliza las colecciones generales combinadas para formar las listas de reproducción. Los resultados se les agrupa usando el método de análisis ANOVA obteniendo los siguientes resultados en promedio comparado con las demás métricas:

Métrica	similitud (0-5)	Inconsistencia (0-1)
PCA	0.7	0.6
MFCC	2.6	0.3
TEMPO	0.9	0.6
CLAS	2.6	0.2
Híbrido	2.9	0.2

Tabla 3.7: Resultados de cada métrica para similitud e inconsistencia.

3.2.1.6. Similitud usando redes neuronales

Descripción: En el trabajo de Cleveland et al. (2020), se lleva a cabo una comparativa de una medida que determina la similitud de una pieza musical con respecto a una pieza musical origen. Se utiliza una red neuronal ligera de tripletas, en donde para una canción origen se ajustan los pesos de la red usando una segunda canción de relevancia positiva y una tercera canción de relevancia negativa.

Características utilizadas: Cromagrama de las transformadas de Fourier y constante Q, *cromagrama* normalizado, coeficientes de las frecuencias Mel, cruce por cero,

media cuadrática de energía, centroide espectral, ancho de banda espectral, contraste espectral, caída espectral y curtosis espectral.

Metodología: La red neuronal consta de dos solo 2 capas fuertemente conexas, con el mismo tamaño que la entrada. Su salida son las características iniciales modificadas con los pesos de la red, con el fin de formar un nuevo espacio multidimensional en el cual se puede medir su distancia Euclidiana l_2 . Sus pesos son ajustados usando tripletas. Una tripleta consta de una canción ancla a , una canción de relevancia positiva con respecto a la canción ancla p , y una canción de relevancia negativa n . La función de pérdida con tripletas está definida de la siguiente forma:

$$L(a; p; n) = \sum_{i=1}^N [jf(a_i) - f(p_i)]^2 - [jf(a_i) - f(n_i)]^2 + \gamma \quad (3.10)$$

Donde γ es el margen esperado entre canciones positivas y negativas. La pérdida se minimiza cuando la distancia entre a y p ($f(a)$ y $f(p)$) es menor que la distancia entre a y n ($f(a)$ y $f(n)$). Usando la base de datos *Free Music Archive* se entrenó la red neuronal usando tripletas donde las canciones positivas eran seleccionadas si eran del mismo artista. Se entrenó con 200 épocas y 102,400 tripletas usando el 70 % para entrenamiento y el 30 % para prueba, el hiperparámetro γ es igual a 1.

Evaluación Objetiva: Se toma en cuenta 2,529 consultas al conjunto de prueba usando la red entrenada, una consulta por cada artista obteniendo una clasificación para las canciones como relevantes o no relevantes. Se calcula la curva ROC para explorar el umbral con el que realiza la clasificación, reportando el área bajo la curva ROC. Un valor ideal para el área bajo la curva ROC es de 1. Se compara la red neuronal con el uso de las características originales sin ser modificadas y una versión modificada de las entradas siendo estandarizadas.

Modelo	Área bajo la curva ROC
Características originales	0.800
Características estandarizadas	0.825
Red neuronal de tripletas	0.900

Tabla 3.8: Comparación de resultados por método.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.1.7. Similitud usando desviación concurrente

Descripción: En el trabajo de Chakrabarty et al. (2019), se lleva a cabo la evaluación de una medida que determina la similitud entre piezas musicales calculando la desviación concurrente en las frecuencias fundamentales.

Características utilizadas: Frecuencias fundamentales.

Metodología: Se obtienen las frecuencias fundamentales presentes en las piezas musicales, de igual forma se obtiene la cantidad con la que se repiten dichas frecuencias en las ventanas de muestreo. Posteriormente se calcula la desviación concurrente:

1. Obtener las N frecuencias que más se repiten. En este caso $N = 16$.
2. Multiplicar cada frecuencia f_j por el número de repeticiones.
3. Recorrer los N valores en orden desde el que tiene el mayor número de repeticiones. Calculando la desviación con respecto al valor anterior (excepto el primer elemento). La desviación se determina como $+1$ si incrementó o -1 si disminuyó.
4. Repetimos los pasos anteriores para la segunda canción.
5. Realizamos un producto punto entre ambos vectores de desviación, obteniendo el valor C .
6. Finalmente la desviación concurrente R se define como:

$$R = \frac{C}{N} \quad (3.11)$$

Donde el resultado determina una mayor similitud entre las canciones si es que es más cercano a 1.

Evaluación Objetiva: Se tomaron 4 canciones, las canciones 1,3 y 4 provienen de una misma colección con características similares, y la canción 2 es de una colección diferente. Se obtuvieron los siguientes resultados en las comparaciones realizadas:

Comparación	Resultado
Canción 1 y Canción 2	0.258
Canción 1 y Canción 3	0.930
Canción 1 y Canción 4	1.064

Tabla 3.9: Resultados de similitud entre canciones.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.1.8. Similitud usando descriptores melódicos y rítmicos

Descripción: En el trabajo de Panteli y Dixon (2016), se definen nuevos descriptores que utilizan propiedades de la melodía y el ritmo. Dichos descriptores son utilizados para entrenar una serie de clasificadores, así como usarlos directamente con métricas ya existentes de similitud. De igual forma evalúan su robustez al realizar transformaciones al conjunto de canciones.

Características utilizadas:

- Ritmo:
 - Patrones de detección de eventos (OP), posterior a obtener los coeficientes cepstrales de las Frecuencias Mel (MFCCs) se calcula la diferencia de primer orden en cada banda de frecuencias y se subtrae el promedio, rectificando a media onda el resultado.
 - Fluctuación de patrones (FP), es similar a OP sin embargo, utiliza una ponderación en las frecuencias con respecto a su intensidad.
 - Transformación de escala en eventos (ST), realiza una transformación de la señal que sea invariante a la escala y por lo tanto al tiempo.
- Melodía:
 - Bihistograma tonal (PB), es una matriz de $b \times n$ donde n es el número de tonos a utilizar, y describe que tan frecuentes son cada par de notas en una ventana de tiempo d .
 - Magnitudes de la transformada de Fourier en 2 dimensiones (FTM), describe la periodicidad en la que se encuentran presentes las frecuencias a lo largo del tiempo, siendo independientes de la clave, pero incluyendo información tonal de manera cromática.
 - Intervalogramas (IG), es una representación vectorial de la representación cromática tonal de la pieza musical. Se obtienen los vectores cromáticos por cada medio segundo y se calcula la correlación cruzada con cada vector subsecuente.

Metodología: Se extraen las características mencionadas del sonido, las cuales se utilizarán para entrenar los siguientes clasificadores:

- K Vecinos más cercanos (KNN).
- Máquina de soporte Vectorial (SVM).
- Clasificador Bayesiano Ingenuo (NB).
- Análisis Discriminatorio Lineal (LDA).

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

De igual forma las características se usan para definir un espacio multidimensional donde se usan las siguientes métricas de similitud y distancia:

- Distancia Euclidiana.
- Similitud Coseno.
- Correlación.
- Distancia Mahalanobis.

Para probar la robustez de los descriptores se realizaron transformaciones en las canciones con el fin de observar la variabilidad de los resultados con respecto a ellas.

- Modificación del Timbre (Timb).
- Cambio de la calidad de grabación (RecQ).
- Cambios globales de tempo (GTemp).
- Transposición de clave musical y tempo local (LTemp).

Evaluación Objetiva: Los experimentos consisten en probar la exactitud de cada clasificador y métrica para determinar canciones relevantes para una canción origen.

La primera prueba utiliza la colección de canciones sin transformaciones. La siguiente tabla compara los 6 descriptores propuestos en los clasificadores y las métricas:

Clasificador	ST	OP	FP	PB	IG	FTM
KNN	0.86	0.71	0.68	0.88	0.83	0.86
LDA	0.82	0.66	0.59	0.83	0.82	0.82
NB	0.80	0.62	0.58	0.84	0.76	0.81
SVM	0.87	0.66	0.59	0.86	0.86	0.87
Métrica						
Distancia Euclidiana	0.65	0.47	0.42	0.56	0.83	0.67
Similitud Coseno	0.66	0.47	0.42	0.80	0.55	0.68
Correlación	0.66	0.47	0.42	0.80	0.54	0.67
Distancia Mahalanobis	0.61	0.48	0.40	0.81	0.60	0.72

Tabla 3.10: Comparación de resultados por transformación.

La segunda prueba utiliza la colección de canciones con las 4 transformaciones. La siguiente tabla compara los 6 descriptores propuestos en el clasificador KNN y en la métrica de distancia Mahalanobis:

	Descriptor	Transformaciones			
		Timb	GTemp	RecQ	LTemp
Clasificador KNN					
	ST	0.98	0.90	0.93	0.62
	OP	0.97	0.20	0.92	0.75
	FP	0.91	0.18	0.92	0.71
	PB	0.97	0.99	0.78	0.76
	IG	0.95	0.99	0.62	0.77
	FTM	0.98	0.96	0.71	0.79
Distancia Mahalanobis					
	ST	0.95	0.36	0.91	0.25
	OP	0.94	0.00	0.88	0.13
	FP	0.62	0.01	0.87	0.09
	PB	0.94	0.98	0.78	0.53
	IG	0.70	0.91	0.62	0.46
	FTM	0.87	0.88	0.71	0.57

Tabla 3.11: Comparación de resultados por transformación.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Se concluye que el descriptor de transformación de escala es el más robusto por parte del ritmo, y por parte de la melodía el más robusto ante los cambios fue el bihistograma tonal.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.1.9. Similitud con acercamientos probabilistas

Descripción: En el trabajo de Tran et al. (2020), se lleva a cabo la evaluación de máquinas restringidas de Boltzmann, entrenadas para determinar la distancia relativa de una tripleta de canciones.

Características utilizadas: Clave, Modo, Intensidad, Beats por minuto, Ritmo, Energía, Media tonal, Desviación estándar tonal, Media tímbrica, Desviación estándar tímbrica, Género.

Metodología: En la tripleta hay una canción ancla x y dos canciones adicionales y y z , donde la tarea es descubrir la relación de similitud en la que se responda si x es más similar a y que a z . Una máquina restringida de Boltzmann optimiza los pesos de sus parámetros con el fin de optimizar la siguiente tarea de clasificación en un conjunto de entrenamiento:

$$I(x; y; z) = \begin{cases} 1; & \text{si } x \text{ es más similar a } y \text{ que a } z \\ 0; & \text{si } x \text{ es menos similar a } y \text{ que a } z \end{cases} \quad (3.12)$$

Evaluación Objetiva: En la colección de canciones *MagnaTagATune* se llevó a cabo una serie de experimentos probando la tarea de clasificación de las tripletas midiendo su distancia relativa. Sus resultados comparados con una distancia Euclidiana simple y máquinas de soporte vectorial (SVM) son reportados en la siguiente tabla:

Métrica	Exactitud
Máquina restringida de Boltzmann	71.93 %
Distancia Euclidiana	56.71 %
Máquina de soporte vectorial	71.20 %

Tabla 3.12: Exactitud de las métricas comparadas.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.1.10. Similitud usando *SiMPLe*

Descripción: En el trabajo de Silva y Yeh (2016), se lleva a cabo la evaluación de una medida de Distancia basada en un perfil de matriz de similitud conocida como SiMPLe (Similarity Matrix Profile).

Características utilizadas: Vectores cromáticos de 12 tonos por cada medio segundo.

Metodología: Se extrae la energía por cada tono presente en la escala de 12, independientemente de la octava, en la música occidental.

Posteriormente se define un valor arbitrario m que será el tamaño de cada subsecuencia a utilizar. una subsecuencia es un conjunto contiguo de elementos en el vector. La matriz SiMPLe consiste en una matriz que contiene para cada subsecuencia de tamaño m en una serie de tiempo A , su vecino más cercano en la serie de tiempo B . Se utiliza la siguiente fórmula para obtener la distancia entre una subsecuencia Q de A con respecto a la subsecuencia T de B :

$$D = \frac{1}{2m-1} \sqrt{\frac{QT}{m} + \frac{m_Q M_T}{T}} \quad (3.13)$$

Donde QT es el producto punto, Q es la media de Q , M_T es la media de T , σ_Q es la desviación estándar de Q y σ_T es la desviación estándar de T . Finalmente, después de obtener todas medidas de distancia para cada par de subsecuencias, se define la distancia global como la mediana de todas ellas.

Evaluación Objetiva: En una colección de canciones obtenidas de YouTube se llevó a cabo una serie de experimentos probando la tarea de identificación de covers de una canción original, asumiendo que entonces dichos covers tendrían una menor distancia a la original que canciones completamente distintas. Los resultados se reportan con respecto a su exactitud promedio, con un rango de 0 a 1. Obtuvo un resultado de 0.591 que en su comparación superó a los métodos existentes que fueron probados en el mismo conjunto de datos.

Colección de canciones	Exactitud	Área bajo la curva ROC
DB160	0.7069	0.9733
DB400	0.9600	0.9911
DB801	0.5982	0.9157

Tabla 3.13: Resultados por conjunto de datos.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.1.11. Similitud usando Recurrencia Cruzada

Descripción: En el trabajo de Fan y Chen (2019), se lleva a cabo la evaluación de una medida de Distancia basada en la fusión de gráficas de recurrencia cruzada, siendo integradas con multi-kernels.

Características utilizadas: Perfil de clase tonal y harmónica y evolución melódica.

Metodología: Se extraen las características mencionadas. De las características extraídas se entrena por cada uno un nuevo modelo basado en gráficas de recurrencia. Las matrices de recurrencia se concatenan usando una fusión de recurrencia cruzada. Finalmente, el valor de la inferencia producida por el modelo pasa por una integración con multi-kernels que eliminan el ruido y la alta variación en los resultados a lo largo de la gamma de canciones en las colecciones que fue evaluado.

Evaluación Objetiva: Se evaluó con respecto a 3 colecciones de canciones en una tarea de identificación de covers, teniendo un umbral de distancia en la medida de similitud para ser clasificada como cover. Los resultados fueron superiores a los métodos previos en el estado del arte.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.2. Basadas en contenido representado simbólicamente

En el dominio de la música, las características clave que podemos distinguir en el mundo real han sido codificadas con el propósito de ser analizadas y reproducidas digitalmente. Un ejemplo pueden ser las grabaciones de audio de los cuales podemos identificar características como el timbre, ornamentos, variaciones tonales y de ritmo como lo son el vibrato y el tremolo. Una desventaja de los archivos de audio es la dificultad presente en los programas de computadora para poder analizarlos puesto que solo cuentan con una señal mezclada de la cual extraer todos los datos. Otra representación popular utilizada incluso en instrumentos digitales es la representación MIDI, una codificación para los sistemas computacionales similar a lo que para nosotros son las partituras. Esta codificación presenta sus propias limitaciones de igual forma, carece de información de la señal de audio puesto que solo incluye las notas que pueden ser interpretadas por distintos instrumentos, tampoco se registran datos avanzados como ornamentos. Una ventaja que tienen los archivos MIDI sobre los archivos de audio que es explorada en este trabajo es la facilidad con la que algunas de las características clave pueden ser extraídas con gran facilidad, en un archivo MIDI se encuentra la secuencia de notas presentes junto con su duración.

Aunque pueden calcularse de igual forma los descriptores de la sección anterior en este caso al omitir un análisis previo para la obtención de las notas presentes en la composición musical (puesto que en una representación simbólica estas se encuentran explícitamente detalladas) entonces se puede hacer uso de características más abstractas provenientes de la teoría musical.

Grupo	Descriptores
Ritmo	Beats por minuto y Patrones rítmicos.
Tonalidad	Notas musicales, acordes, intervalos, modos y progresiones.
Otros	Emoción, Género, Distancia de edición y comparación secuencial.

Tabla 3.14: Tipos de Descriptores.

3.2.2.1. Similitud usando distancia correlacional

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de distintas medidas de similitud con distintos acercamientos. Una de ellas es usar la distancia correlacional, la cual consiste en calcular una diferencia independiente de la transposición y rango de una melodía con respecto a otra.

Características utilizadas: Secuencias de notas tomando en cuenta la duración de cada una.

Metodología: Se define en la propuesta original de Scherrer y Scherrer (1971) una longitud arbitraria m , se obtienen las subsecuencias de notas de la melodía A , con el fin de compararlas con las subsecuencias de tamaño m en B . La distancia entre una subsecuencia de notas q y otra subsecuencia p está dada por:

$$dist(q;p) = 1 - \frac{\sum_{i=1}^n (q_i - \bar{q})(p_i - \bar{p})}{\sqrt{\sum_{i=1}^n (q_i - \bar{q})^2} \sqrt{\sum_{i=1}^n (p_i - \bar{p})^2}} \quad (3.14)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.797.

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.503$.

3.2.2.2. Similitud usando distancia de ciudad

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de distintas medidas de similitud con distintos acercamientos. Una de ellas es usar la distancia ciudad, City Block Distance (CBD) o distancia *Manhattan* la cual consiste en calcular la diferencia absoluta en cada dimensión o elemento de una secuencia.

Características utilizadas: Secuencias de notas tomando en cuenta la duración de cada una.

Metodología: En la propuesta original por Steinbeck (1982), se define una longitud arbitraria m , se obtienen las subsecuencias de notas de la melodía A , con el fin de compararlas con las subsecuencias de tamaño m en B . La distancia entre una subsecuencia de notas q y otra subsecuencia p está dada por:

$$dist(q; p) = \sum_{i=1}^m |q_i - p_i| \quad (3.15)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.774.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.499$.

3.2.2.3. Similitud usando distancia Euclidiana

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de distintas medidas de similitud con distintos acercamientos. Una de ellas es usar la distancia Euclidiana entre cada valor de nota MIDI entre las subsecuencias.

Características utilizadas: Secuencias de notas tomando en cuenta la duración de cada una.

Metodología: Se toma de referencia la metodología usada por Steinbeck (1982), basada a su vez en el trabajo de Mongeau y Sankoff (1990), Se define una longitud arbitraria m , se obtienen las subsecuencias de notas de la melodía A , con el fin de compararlas con las subsecuencias de tamaño m en B . La distancia Euclidiana entre una subsecuencia de notas q y otra subsecuencia p está dada por:

$$dist(q;p) = \sqrt{\sum_{i=1}^m (q_i - p_i)^2} \quad (3.16)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.764.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.468$.

3.2.2.4. Similitud usando Alineación Local

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de distintas medidas de similitud con distintos acercamientos. Una de ellas es calcular la alineación local óptima por cada subsecuencia de notas en la melodía. Algunas de las aplicaciones de esta medida de similitud son tutores automatizados o compañeros de composición (Duggan y O'Shea (2011) y Dannenberg y Raphael (2006)).

Características utilizadas: Secuencias de notas sin tomar en cuenta la duración.

Metodología: En la propuesta utilizada de referencia originalmente propuesta por van Kranenburg et al. (2013), se define que dadas dos melodías q y s , siendo secuencias de notas de tamaño m y n respectivamente, se define una función que describe los pasos de inserción, reemplazo y eliminación necesarios para obtener el mejor alineamiento de ambas melodías. Se recorre con dos índices las posiciones en las melodías, cuando las notas son iguales se premia, en otro caso se penaliza. Sin embargo, puede minimizarse

3. EVALUACIÓN DE LA NOVEDAD MUSICAL

la penalización eliminando la nota que no coincide y pasar a la siguiente posición, o insertando la nota necesaria para que coincida. Una matriz con los resultados locales A se inicializa con 0 en la primera columna y en la primera fila. La función para llenar cada posición restante de la matriz está definida como:

$$A(i;j) = \max \begin{cases} A(i-1;j-1) + \text{subs}(q_i; s_j) \\ A(i;j-1) + W_{\text{insercion}} \\ A(i-1;j) + W_{\text{eliminacion}} \\ 0 \end{cases} \quad (3.17)$$

Donde $W_{\text{eliminacion}} = W_{\text{insercion}} = 0.5$ y subs se define como:

$$\text{subs}(q_i; s_j) = \begin{cases} 1; & \text{si } q_i = s_j \\ 0; & \text{en otro caso} \end{cases} \quad (3.18)$$

Después de llenar cada celda de la matriz queda registrada la posición con el mayor puntaje de alineación y finalmente se obtiene una medida de similitud global definida como:

$$\text{sim}(q; s) = \frac{1}{n} \max_{i,j} (A(i;j)) \quad (3.19)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.859.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.621$.

3.2.2.5. Similitud usando Inducción de Estructura

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de un algoritmo de inducción de estructura el cual consiste en calcular realizar una transformación de las secuencias de notas usando una representación de puntos usando su nota y posición en la melodía originalmente propuesta por Meredith (2006). Posteriormente el autor realizó una serie de modificaciones Meredith (2016) que no se incluyen en este estudio.

Características utilizadas: Secuencias de notas tomando en cuenta la posición en el tiempo de cada una.

Metodología: Se obtienen la representación como vectores de puntos de cada secuencia.

Sea \mathbf{A} una melodía donde $\mathbf{A}[i]$ es un punto en el que su valor en el eje horizontal ($\mathbf{A}[i](x)$) es el tono de la i -ésima nota en esa melodía y su valor en el eje vertical ($\mathbf{A}[i](y)$) es el punto de inicio en el tiempo de la i -ésima nota. El vector de traducción (\mathbf{T}) es calculado de cada i -ésimo par de notas entre dos melodías \mathbf{A} and \mathbf{B} de la siguiente forma:

$$\mathbf{T}_i = \begin{matrix} \mathbf{A}[i](x) & \mathbf{B}[i](x) \\ \mathbf{A}[i](y) & \mathbf{B}[i](y) \end{matrix} \quad (3.20)$$

Así $\mathbf{A}[i]$ puede transformarse en $\mathbf{B}[i]$ usando su respectivo vector de traducción \mathbf{T}_i . El vector de traducción \mathbf{T}_i no es único para $(\mathbf{A}[i]; \mathbf{B}[i])$, puesto que si se aplica \mathbf{T}_i a otro punto en \mathbf{A} se puede obtener un valor distinto a $\mathbf{B}[i]$,

Para un vector de traducción \mathbf{T} se define el patrón traducible de longitud máxima o *maximally translatable pattern* (MTP), como el conjunto de puntos en \mathbf{A} tal que cuando se les aplica \mathbf{T} se obtiene un valor en \mathbf{B} , es decir:

$$MTP(\mathbf{A}; \mathbf{B}; \mathbf{T}_i) = \{ \mathbf{A}[i] \mid \mathbf{A}[i] + \mathbf{T}_i \in \mathbf{B} \} \quad (3.21)$$

Finalmente, esta medida (*SIAM*) se calcula obteniendo el MTP más grande posible entre \mathbf{A} y \mathbf{B} , normalizado entre la longitud de la melodía (n):

$$SIAM(\mathbf{A}; \mathbf{B}) = \frac{1}{n} \max_i |MTP(\mathbf{A}; \mathbf{B}; \mathbf{T}_i)| \quad (3.22)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.870.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.665$.

3.2.2.6. Similitud usando Interpolación

Descripción: En el trabajo de Janssen et al. (2015), se lleva a cabo la evaluación de un algoritmo de interpolación utilizando curvas Spline. Fue originalmente propuesta por Urbano (2012).

Características utilizadas: Secuencias de n-gramas de notas tomando en cuenta el tono, punto de inicio en el tiempo y duración.

Metodología: En el trabajo del autor se propuso usar curvas para computar la similitud entre melodías, nos referiremos a esta medida como *BSPLINE*. En este caso las melodías son representadas como secuencias de curvas en un plano tonal-temporal, el cual es aproximado usando curvas Spline uniformes de tipo B. De esta forma la melodía se representa como una curva dividida en splines, donde cada división es un n-grama. Así, es posible usar *Alineacion Local* para medir la similitud entre dos curvas Spline, realizando las siguientes modificaciones a la ecuación para la subsecuencia de longitud máxima (3.17).

$$\begin{aligned}
 W_{ins} &= (1 - f(m)) \\
 W_{rem} &= (1 - f(m)) \\
 subs(i:j) &= \begin{cases} 1 - f(m); & \text{si } \mathbf{A}[i] = \mathbf{B}[j] \\ deriv(\mathbf{A}[i]; \mathbf{B}[j]); & \text{en otro caso} \end{cases}
 \end{aligned} \tag{3.23}$$

Donde $f(m)$ indica la frecuencia de la extensión de la curva spline m en la colección de datos, y $deriv$ tiene las siguientes condiciones.

- Si las dos curvas tienen el mismo signo de derivación, tanto al inicio como al final de la división, la penalización es pequeña.
- Si las dos curvas tienen distinto signo de derivación, en ambos al inicio y al final de la división, la penalización es grande.
- Si las dos curvas tienen el mismo signo de derivación en el inicio o al final, pero no en el otro, la penalización es promedio.

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.776.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Sin embargo, no se realizaron pruebas de correlación en adición a la mencionada anteriormente.

3.2.2.7. Similitud usando Ondeletas

Descripción: En el trabajo de Janssen et al. (2017), se lleva a cabo la evaluación de distintas medidas de similitud con distintos acercamientos. Una de ellas es calcular el contorno de la melodía y determinar la distancia de dicho contorno con respecto al de otra melodía.

Características utilizadas: Secuencias de notas sin tomar en cuenta la duración.

Metodología: La descripción de la propuesta original por Velarde et al. (2013) menciona que dadas dos melodías q y s , siendo secuencias de notas de tamaño m y n respectivamente, se obtiene una señal donde se grafiquen los tiempos y tonos presentes en la melodía. Posteriormente, se calculan los coeficientes de la transformada de Ondeleta de las señales. La Ondeleta origen que se utilizó fue la Ondeleta de Haar. Los coeficientes tienen un rango de -15 a 15, con la propiedad de que se puede identificar los momentos en que la melodía *sube* o *baja*.

La distancia de similitud se mide finalmente entre la distancia ciudad entre ambas transformadas de las melodías.

$$dist(q;p) = \sum_{i=1}^n |q_i - p_i| \quad (3.24)$$

Evaluación Objetiva: Se usó la colección de datos MTC-ANN que contiene 360 melodías. El objetivo es obtener subsecuencias de notas que describan frases melódicas que comparten cada par de melodías. La base de datos tiene anotados los segmentos que debería considerar como compartidos. Se define un umbral arbitrario para el cual si la medida de distancia es menor a ese umbral entonces la subsecuencia será considerada compartida. Una evaluación para definir dicho umbral es la curva ROC, donde se prueba un rango para el umbral, un valor ideal del área bajo la curva ROC es igual a 1. Esta medida de similitud tuvo un área de 0.731.

Evaluación Subjetiva: La evaluación de precisión tomando en cuenta los verdaderos y falsos positivos se comparó con las anotaciones de expertos. Su factor de correlación es ideal si es igual a 1. El factor que obtuvo fue de $r = 0.459$.

3.2.2.8. Similitud usando redes neuronales

Descripción: En el trabajo de Bretan y Heck (2019), se lleva a cabo la evaluación de una medida de similitud que usa un espacio latente inferido de una colección de canciones a través de una red neuronal autocodificadora variacional.

Características utilizadas: Parejas de nota y duración en formato MIDI.

Metodología: En la arquitectura de un Modelo Profundo Semánticamente Estructurado (DSSM) tiene la particularidad, a diferencia de un autocodificador variacional convencional, de ser más flexible en las restricciones e integraciones del espacio latente una vez entrenado el modelo. Una vez optimizado el modelo en la tarea de agrupamiento y clasificación del conjunto de datos por artista, se puede hacer uso de los pesos del espacio latente de características que la red usa para inferir una categoría. Dicho espacio latente se observa como un espacio multidimensional en el cual se pueden codificar las melodías. Una vez codificadas las melodías en este espacio multidimensional, la similitud se puede medir con la distancia euclidiana en este espacio:

$$dist(q; p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.25)$$

Evaluación Objetiva: Se usó la colección de datos *Piano Music Dataset*, se entrenaron dos modelos. Un modelo corresponde a su arquitectura propuesta que infiere el espacio latente, mientras que un segundo modelo utiliza las características resultantes como datos de entrenamiento y poder categorizar las canciones como relevantes o no a una canción origen. Ambos modelos fueron utilizados para poder realizar agrupamientos de canciones que fueran las más similares a una canción consulta, usando grupos de 101 canciones, 1 canción es la canción consulta, otra canción es la canción que esta categorizada en la base de conocimiento como similar a la canción consulta y las 99 canciones restantes son elegidas al azar. De forma ideal la canción similar debería tener el rango 1, es decir la más similar a la canción consulta. El rango promedio obtenido por cada modelo es reportado:

Modelo	Rango Promedio
DSSM	8.8
AdjD	5.1

Tabla 3.15: Comparación de resultados por método.

Evaluación Subjetiva: No se realizó una evaluación subjetiva.

3.2.2.9. Similitud usando grafos

Descripción: En el trabajo de Simonetta et al. (2018), se lleva a cabo la evaluación de una medida de similitud usando grafos de reducción de melodías.

Características utilizadas: Melodías en formato MIDI.

Metodología: Aplicando reglas de teoría musical una melodía, en puntos seleccionados, se segmenta y reduce en elementos atómicos por segmento, es decir, se abstrae de una melodía una secuencia de notas que se asemeje a la original, pero usando un solo elemento por segmento. Los pasos que sigue la reducción se almacenan y se genera un grafo conectando cada estado de la melodía con su sucesor y predecesor. En una misma pieza musical una reducción puede pertenecer a distintos puntos origen, generando una estructura compleja interconectada. La distancia de similitud entre dos melodías puede verse como la distancia en aristas entre ambas melodías. De no tener un ancestro común en el grafo entonces se recurre a calcular la distancia ya sea directamente en su valor MIDI o tomando en cuenta los pesos de disonancia en la diferencia de acordes.

Evaluación Objetiva: Se usó la colección de datos *Exhanced Wikifonia Leadsheet Dataset* (EWLD), la cual contiene 5000 partituras. Se realizaron experimentos donde se representaban en formato de grafos las melodías, se tomaba una melodía origen y su correspondiente canción origen. En los experimentos después de representar las melodías en forma de grafos, se tomaban las 100 canciones más similares a la melodía original según la distancia en el grafo. Se promedió la exactitud con la que la canción origen aparecía en las primeras 20. La segmentación puede hacerse de manera indexada para todas las canciones, es decir que se compartan el proceso de reducción con segmentos similares, o que los segmentos fueran reducidos de manera aleatoria. Los resultados se reportan en la siguiente tabla:

Distancia	Exactitud con segmentos indexados	Exactitud con segmentos Aleatorios
MIDI	0.54	0.06
Acordes	0.56	0.07

Tabla 3.16: Exactitud por tipo de segmentos.

Los resultados demuestran que el método no es efectivo en comparación al estado del arte en similitud musical, sin embargo, al visualizar la construcción del grafo se pueden encontrar similitudes ocultas entre canciones que serían categorizadas como irrelevantes entre sí.

Evaluación Subjetiva: No se realizó evaluación subjetiva.

3.2.2.10. Similitud usando Arreglo Espiral

Descripción: En el trabajo de Alvarado et al. (2020), se propone un algoritmo genético para la generación de variaciones en una pieza musical de tal forma que produzca algo novedoso sin dejar de ser similar a la pieza original. Para esto proponen una función objetiva que parcialmente es influenciada por una nueva medida de similitud musical. En forma general esta medida de similitud mide la distancia euclidiana entre los vectores representativos de ambas piezas musicales. Sin embargo, el vector representativo se genera a través del modelo de arreglo espiral propuesto por Chew y Chen (2005) definiendo parámetros arbitrarios para generar un espacio de tres dimensiones en donde se puedan colocar las notas de acuerdo a su distancia tonal en una clave especificada. Dicho espacio permite una mayor especificidad y libertad de modificación en la generación y localización de distintas escalas musicales.

Características utilizadas: Parejas de nota y duración en formato MIDI agrupados por barras.

Metodología: Dados los parámetros para definir el arreglo espiral (h y r), se empieza con una nota arbitraria como la raíz del arreglo ($k = 0$). Los parámetros se definieron de acuerdo al trabajo de Chew (2002), es decir: $h = \frac{2}{15}$, $r = 1$ y D_0 como la nota raíz. Ahora cada coordenada de 3 dimensiones en el índice k se calcula de la siguiente forma:

$$P[k] = \begin{matrix} 2 \\ 4 \\ kh \end{matrix} \begin{matrix} r \sin \frac{k}{2} \\ r \cos \frac{k}{2} \\ 5 \end{matrix} \begin{matrix} 3 \\ 5 \\ 5 \end{matrix} \quad (3.26)$$

Posteriormente para obtener el "Centro de Efecto" de cada barra, se obtiene la coordenada promedio de cada nota única en la barra.

$$\delta_i \text{ } {}_b C E_i = \frac{\frac{2}{6} \frac{P_{j=0}^{n_i} P[V][0]}{n_i} \frac{3}{7}}{\frac{4}{6} \frac{P_{j=0}^{n_i} P[V][1]}{n_i} \frac{7}{5}} \quad (3.27)$$

Donde b es el número de barras y n_i el número de notas en la i -ésima barra. La representación en espiral hace que esto sea equivalente a obtener el acorde presente en cada barra. Finalmente, entre un par de piezas musicales la similitud se define como la distancia euclidiana entre cada Centro de Efecto.

$$S = 100 \prod_{i=0}^p CE_i \quad CE_i^{\theta} \quad (3.28)$$

Evaluación Objetiva: Se usó la colección de datos *12 variation on Twinkle Twinkle Little Star*, en el cual se comprobó como en cada una de las 5 secciones con variabilidad incremental, la medida de similitud fue disminuyendo, comprobando la sensibilidad en los cambios a la melodía:

Sección	Similitud
1	100
2	96.8
3	83.25
4	92.5
5	84.88

Tabla 3.17: Similitud por sección.

Evaluación Subjetiva: Las pruebas consistieron en evaluar las piezas generadas donde voluntarios le daban una puntuación a la armonía en general entre 0 y 10, en promedio sin tomar en cuenta la medida de similitud las piezas tenían un puntaje de 6.6, sin embargo, este puntaje mejora al incorporar la medida de similitud a un 6.9. Comprobando que el incorporar los elementos de similitud y diversidad aumentan la variación sin perjudicar e incluso mejorar su calidad armónica.

Capítulo 4

RNA: Una medida de similitud musical

En este capítulo se presenta *Rhythm-Note Alignment* (*RNA* por sus siglas en inglés), la medida propuesta en este trabajo para evaluar la novedad de composiciones musicales a partir de su similitud con respecto a un conjunto de piezas musicales. El nombre hace referencia a que utiliza características de ritmo y nota musical para hacer una comparación similar a la presente en Alineación Local (ver sección 3.2.2.4). Otra medida que podría considerarse similar es FANTASTIC. En el trabajo de Müllensiefen y Frierler (2006), se presenta FANTASTIC. FANTASTIC reúne características de n-gramas (secuencias nota-ritmo) y características de estadísticas con frecuencia de elementos abstractos en un corpus o dentro la misma melodía con las características reunidas hace una comparación de similitud usando distancia euclidiana para n-gramas.

El trabajo realizado con la medida propuesta fue sometido a la *International Conference on Computational Creativity (ICCC) 2021* usando el nombre en inglés con el propósito de mantener uniformes las referencias a la medida.

La descripción de la medida propuesta inicialmente incorporaba un mayor número de propiedades de las piezas musicales y está detallada en el Apéndice A. Sin embargo, con la experimentación realizada y expuesta en el Apéndice B, se obtuvieron mejores resultados con menos propiedades y una mínima modificación. Esta versión simplificada está descrita y evaluada en este capítulo.

4.1. Descripción

Con *RNA* se puede medir la similitud entre un par de melodías representadas simbólicamente en formato MIDI, como una combinación lineal entre las diferencias en ritmo y tonos presentes en la sub-secuencia de notas común de longitud máxima (SCLM) entre ambas melodías.

La intuición que originó su desarrollo es que, el usar características simples, como el

4. RNA: UNA MEDIDA DE SIMILITUD MUSICAL

tono y la duración, puede ser de gran utilidad en tareas como calcular la similitud entre melodías. Una de sus peculiaridades es la consideración de la diferencia que hay en los tiempos de inicio de las notas en las melodías, esto es generalmente ignorado en las medidas que fueron usadas y evaluadas en las primeras investigaciones. De igual forma vale la pena cuestionar la penalización estricta al no coincidir en una nota exactamente, presente en *Alineación Local*, con el propósito de evaluar su desempeño sin penalización. Para este propósito la comparación se realiza tomando en cuenta la clase de nota sin considerar la octava en la que se encuentra.

Ahora, solo considerar la cantidad de elementos atómicos que comparten las melodías carece de juicio sobre estructuras más abstractas en la música como el orden en la melodía, acordes, motivos, etc. Para minimizar el impacto de esta falta de juicio se considera el uso de un algoritmo usado en la comparación de secuencias para modificarlo e incluir las características mencionadas previamente.

El algoritmo con opciones convenientes de personalización seleccionado fue *Alineación Local* por su capacidad de considerar el orden relativo en la búsqueda de la sub-secuencia de longitud máxima (SCLM).

Las propiedades principales de *RNA* pueden observarse en la Figura 4.1.

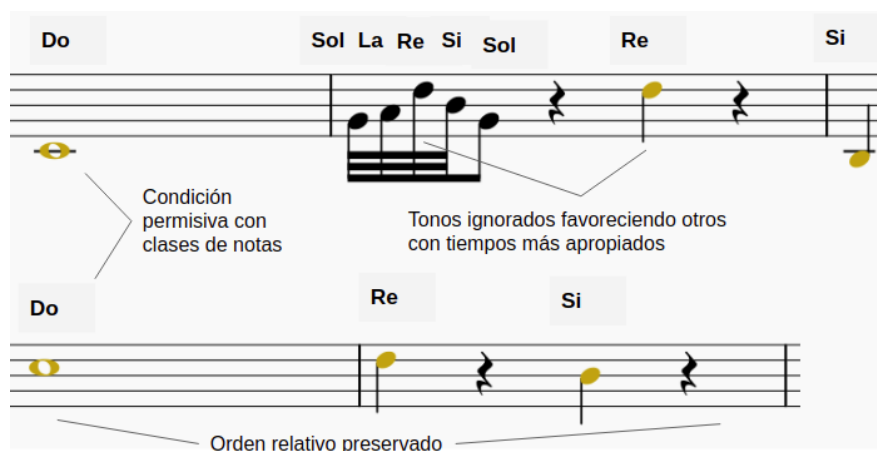


Figura 4.1: Propiedades principales de comparación y orden presentes en *RNA*.

La definición formal representa la relación que existe entre la SCLM y la melodía de longitud máxima entre **A** y **B**, se calcula de la siguiente forma:

$$RNA(\mathbf{A}; \mathbf{B}) = 1 \frac{|SCLM_{\mathbf{A}; \mathbf{B}}|}{\max(|\mathbf{A}|; |\mathbf{B}|)}$$

Para calcular la SCLM entre ambas melodías, primero requerimos eliminar la dependencia de octavas en las melodías, es decir, extraer únicamente las notas sin importar

la octava o la altura en la que estas se encuentren. Esta representación toma todos los Do como de la misma clase, aun cuando un Do₃ sea más grave que un Do₅. La razón de esto es porque en la construcción de acordes y progresiones comúnmente solo se toma en cuenta la distancia relativa entre las notas con respecto a la escala y no en su distancia absoluta en semitonos, por ejemplo, se pueden identificar patrones en la relación entre todos los Re y todos los Do para una pieza en particular, aun cuando cada pareja de Do y Re se encuentran en distintas octavas.

Una vez con la clasificación por nota podemos obtener una cadena de notas para una melodía como se muestra en (Fig. 4.2).



Figura 4.2: Extracto del himno de la alegría, en su representación simbólica (arriba) y en una representación como una cadena cromática de notas musicales (abajo).

Dadas dos melodías A y B representadas como una cadena de clase de notas la SCLM se calcula usando la siguiente función recursiva:

$$SCLM(i;j) = \begin{cases} 0; & \text{si } i = 0 \text{ o } j = 0 \\ \max(SCLM(i-1;j); & \\ SCLM(i;j-1); & \\ SCLM(i-1;j-1) + subs(i;j)); & \text{en otro caso} \end{cases} \quad (4.1)$$

Donde $subs$ es una función definida como:

$$subs(i;j) = \begin{cases} 1; & \text{si } (A[i] = B[j]) \vee (P_A[i] - P_B[j]) < (P_A[i] + P_B[j]) = C_E \\ 0; & \text{en otro caso} \end{cases} \quad (4.2)$$

Donde P_A y P_B son vectores calculados de la siguiente forma:

$$\begin{aligned} P_A[i] &= inicio(A[i]) - inicio(A[i-1]) \\ P_B[j] &= inicio(B[j]) - inicio(B[j-1]) \end{aligned} \quad (4.3)$$

4. RNA: UNA MEDIDA DE SIMILITUD MUSICAL

Donde *inicio* devuelve el tiempo inicial de una nota dada; y C_E es el coeficiente de escala, experimentalmente se obtuvo el mejor desempeño al considerar el promedio del vector \mathbf{P} en las melodías del conjunto de datos, el valor obtenido en promedio fue $C_E = 64$.

Esto fue incorporado en *RNA* con el propósito de contar con un umbral para distinguir diferencias significativas entre las notas. En este caso el coeficiente de escala puede usarse para definir qué tanta diferencia es considerada significativa por el algoritmo. No sería apropiado usar un valor fijo debido a la diversidad que puede existir en las distintas melodías, por el contrario, lo que se encuentra fijada es la escala relacionada con los tiempos de inicio con la hipótesis de que será conveniente para una mayor cantidad de conjuntos musicales.

La cardinalidad de $jSCLMf\mathbf{A};\mathbf{B}g^j$ es maximizada en la función recursiva *SCLM* llegando a una condición de paro de manera similar que en *Alineacion Local*. La llamada original a la función es con i y j inicializados en las longitudes de \mathbf{A} y \mathbf{B} respectivamente. Un ejemplo se puede observar en la Figura 4.3.

A = Do Re Do La Si Sol
B = Re Do Sol Re Do Si
SCLM = Do Re Do Si

Figura 4.3: Ejemplo de una sub-secuencia común de longitud máxima entre un par de melodías.

4.2. Evaluación de la medida propuesta

En esta sección se presentan las métricas a comparar con *RNA* y los entornos de evaluación utilizados.

Nombre	Referencia	Descripción Breve
Distancia Euclidiana (<i>ED</i>)	Steinbeck (1982)	Se determina la diferencia directa entre cada nota en las melodías.
City Block Distance (<i>CBD</i>)	Steinbeck (1982)	Se determina la distancia Manhattan entre las melodías.
Distancia de correlación (<i>CD</i>)	Scherrer y Scherrer (1971)	Se aplica la medida de correlación entre segmentos de ambas melodías.
Alineación Local (<i>LA</i>)	van Kranenburg et al. (2013)	Modificación de la distancia de edición entre melodías
<i>SIAM</i>	Meredith (2006)	Se realiza una transformación utilizando un algoritmo de inducción de estructura.
<i>BSPLINE</i>	Urbano (2012)	Se realiza una aproximación del contorno de las melodías usando interpolación con curvas Spline.
<i>RNA</i>	-	La medida propuesta descrita en la sección anterior

Tabla 4.1: Listado de métricas a comparar.

Para realizar la comparación, los resultados de las demás métricas en la Tabla 4.1 fueron normalizados tomando como referencia su valor mínimo y máximo en las evaluaciones para establecer un rango especificando el valor para 0 % y 100 % respectivamente.

Se utilizó el conjunto de datos utilizado en , *The Meertens Tune Collections* (MTC-ANN-2.0.1) La cual consiste de 355 piezas musicales en formato MIDI agrupadas por familia de melodía, además de contar con anotaciones de similitud por 3 expertos.

Se omitieron las medidas del estado del arte que no se encontraban publicadas para su

evaluación, o en su defecto que los recursos necesarios para llevar a cabo su entrenamiento y evaluación están fuera de los alcances en tiempo y objetivo de este trabajo de investigación.

4.2.1. Evaluación con porcentajes de transformación

Se consideró realizar transformaciones en el conjunto de datos. Se implementaron funciones que pueden realizar las siguientes transformaciones en alguna melodía tomando como argumento el porcentaje con el que la melodía sería afectada (0 para ningún cambio y 100 para una melodía completamente distinta).

$$\mathbf{A} = \langle (Sol;64);(Si;36);(Re;12) \rangle \quad (4.4)$$

En la ecuación 4.4 describe una melodía codificada como una secuencia ritmo-nota donde cada elemento es una dupla con la clase de nota sin incluir octava y su duración en ticks extraídos del archivo MIDI. Esta melodía se utilizará como ejemplo para mostrar el efecto de cada transformación.

1. Transformación de nota musical: consiste en cambiar alguna nota musical por otra al azar en el rango MIDI. Mientras más alto el porcentaje más notas se modificarán.

$$T_1(\mathbf{A}) = \langle (Sol;64);(Si;36);(Do;12) \rangle \quad (4.5)$$

2. Transformación temporal: se cambia la duración de la nota musical con un valor en ticks entre 1 y 4096, equivalente al rango de una barra completa con compás de 4/4. Mientras más alto el porcentaje más notas tendrán tiempos modificados.

$$T_2(\mathbf{A}) = \langle (Sol;48);(Si;36);(Re;12) \rangle \quad (4.6)$$

3. Transformación de nota musical y temporal: tanto como el tiempo de inicio como el valor de la nota musical son modificados. más alto el porcentaje más notas se modificarán tanto en su nota como en su duración.

$$T_3(\mathbf{A}) = \langle (Sol;64);(Fa;24);(Re;12) \rangle \quad (4.7)$$

4. Transformación de eliminación: Se eliminan notas al azar de la melodía. Mientras más alto el porcentaje más notas se removerán.

$$T_4(\mathbf{A}) = \langle (Sol;64);(Si;36);- \rangle \quad (4.8)$$

5. Transformación de Inserción: Se añaden notas con duraciones al azar. Mientras más alto el porcentaje más notas se añadirán.

$$T_5(\mathbf{A}) = \langle (Sol;64);(Si;36);(Re;12);(Mi;54) \rangle \quad (4.9)$$

4. RNA: UNA MEDIDA DE SIMILITUD MUSICAL

Los resultados son el promedio de las diferencias a lo largo de los 5 tipos de transformación y los 10 porcentajes de transformación.

Nombre	Diferencia Promedio
<i>RNA</i>	0.0227
Distancia Euclidiana (<i>ED</i>)	0.3093
Distancia de ciudad (<i>CBD</i>)	0.1774
Distancia de correlación (<i>CD</i>)	0.3926
Alineación Local (<i>LA</i>)	0.2149
<i>SIAM</i>	0.1424
<i>BSPLINE</i>	0.3460

Tabla 4.2: Comparación de diferencias promedio.

En la tabla 4.2 es bastante notable el desempeño que tiene la medida propuesta en comparación al resto, siendo la mejor para este entorno de pruebas. En segundo lugar, se encuentra *SIAM* con una diferencia siete veces mayor. Vale la pena mencionar que *CBD* aun siendo simple en comparación tiene el tercer mejor desempeño. Alineación local *LA*, tiene casi 10 veces más error demostrando que las modificaciones realizadas para llegar a *RNA* tuvieron un gran impacto en este entorno de pruebas.

A continuación, se incluyen las tablas y gráficas correspondientes a las diferencias promedio desglosadas por porcentajes y por tipos de transformación.

Nombre	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
RNA	0.014	0.012	0.013	0.009	0.014	0.023	0.028	0.03	0.032	0.053
ED	0.225	0.301	0.329	0.294	0.327	0.266	0.256	0.233	0.288	0.574
CBD	0.037	0.084	0.097	0.142	0.155	0.17	0.219	0.226	0.234	0.41
CD	0.464	0.451	0.414	0.297	0.317	0.266	0.329	0.33	0.399	0.659
LA	0.1	0.087	0.16	0.235	0.283	0.283	0.296	0.262	0.223	0.219
SIAM	0.1	0.167	0.272	0.243	0.211	0.178	0.096	0.049	0.04	0.067
BSPLINE	0.249	0.414	0.448	0.462	0.432	0.376	0.34	0.269	0.265	0.206

Tabla 4.3: Comparación de diferencias promedio por porcentaje de transformación.

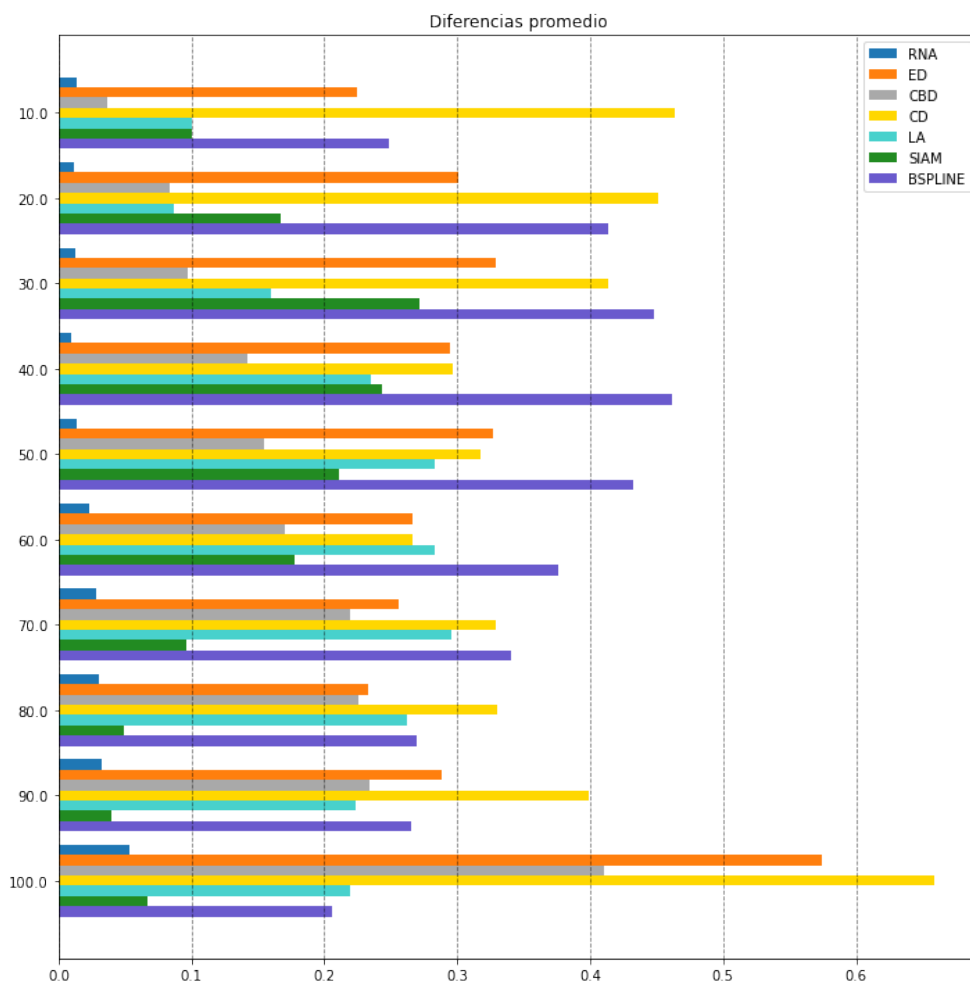


Figura 4.4: Diferencias promedio por porcentaje de transformación.

Los resultados que se muestran tanto en la tabla 4.3 y la Figura 4.4 dan una perspectiva de las distintas métricas a lo largo de los distintos porcentajes. Podemos observar lo pequeña que es la diferencia presente en *RNA* en comparación al resto de las métricas, además de notar cómo este comportamiento de incrementar el error mientras se incrementa el porcentaje no se observa tan uniformemente en las demás métricas. Por ejemplo, el segundo mejor desempeño lo obtuvo *SIAM*, muy cercano a *RNA* para los últimos porcentajes, sin embargo, tiene sus peores valores en el rango de 30-50 % por lo que podemos intuir que su fuerte se encuentra en distinguir melodías o que son muy similares o muy distintas. Un comportamiento similar se encuentra en el resto de las métricas, con la excepción de *CD* en donde pasa un caso invertido, sus peores valores se encuentran con los menores y mayores porcentajes de transformación y teniendo los mejores en el rango de 40-60 %.

4. RNA: UNA MEDIDA DE SIMILITUD MUSICAL

Nombre	T1	T2	T3	T4	T5
RNA	0.067	0.021	0.014	0.008	0.003
ED	0.193	0.55	0.092	0.282	0.43
CBD	0.055	0.55	0.06	0.158	0.064
CD	0.353	0.55	0.255	0.273	0.532
LA	0.171	0.55	0.17	0.091	0.093
SIAM	0.056	0.136	0.206	0.163	0.151
BSPLINE	0.332	0.544	0.338	0.21	0.305

Tabla 4.4: Comparación de diferencias promedio por tipo de transformación.

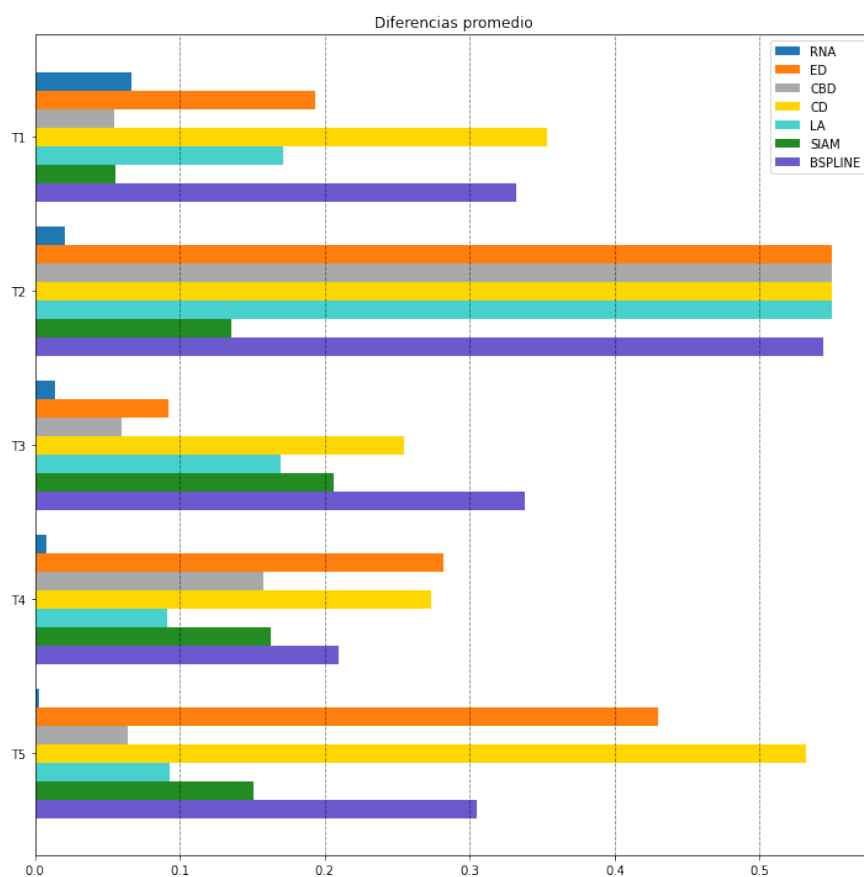


Figura 4.5: Diferencias promedio por tipo de transformación.

En la Tabla 4.4 y la Figura 4.5 podemos observar cómo *RNA* tiene una notable ventaja en general, sin embargo, este desglose por tipo de transformación deja ver los escenarios en donde resaltan las especificidades de cada métrica. Por ejemplo, en el primer tipo de transformación, donde únicamente se reemplazan notas al azar, tanto *CBD* como *SIAM* superaron a *RNA*. En el segundo tipo de transformación en donde se modifican los tiempos de las notas, los resultados fueron considerablemente peores con excepción de *RNA* y *SIAM*, puesto que son las medidas que incorporan factores de ritmo en su evaluación. *ED*, *CBD* y *CD* aun siendo medidas relativamente simples mantuvieron un desempeño con cierta similitud a medidas como *BSPLINE* y *LA*, como es el caso de la transformación 3, en donde se modifican notas y duraciones al azar, y es con resultados como en la transformación 5, en donde se eliminan notas al azar, que su desempeño en general se ve afectado.

4.2.2. Evaluación con anotadores expertos

Como parte del conjunto de datos *The Meertens Tune Collections* (MTC-ANN-2.0.1) (van Kranenburg et al., 2016), se encuentran anotaciones para 355 de las melodías, en donde 3 expertos marcaron segmentos en una melodía como idénticos, relacionados pero variados y diferentes. Para propósitos de este estudio tomamos los siguientes valores para sus tres posibles categorías: “casi idénticas”, que será referenciado con un valor numérico de 0; “relacionado pero variado”, que será referenciado con un valor numérico de 0.5; “diferente”, que será referenciado con un valor numérico de 1. Se utilizó un umbral para redondear los puntajes dados por la medida con el fin de corresponder a los dados por los anotadores. Se calcularon las diferencias promedio con cada anotador y se obtuvieron los siguientes resultados para *RNA*:

A1	A2	A3	Diferencia promedio
0.1501	0.1587	0.1475	0.1521

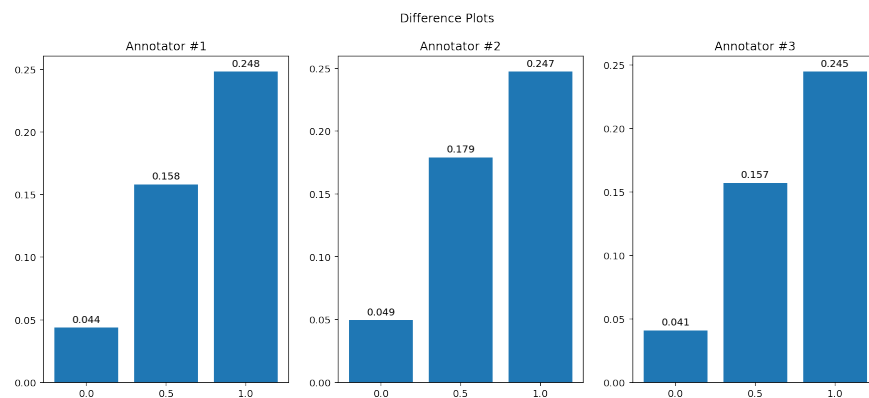


Figura 4.6: Diferencias promedio por anotador.

Para realizar una comparación con las demás métricas se extendió un entorno de evaluación utilizado en estudios anteriores. Este entorno de evaluación considera únicamente una tarea de clasificación binaria, en donde para cada par de melodías pueden ser “casi idénticas” o “diferentes”. En este trabajo se incluye el tercer tipo de anotación presente en el conjunto de datos que previamente no había sido utilizado en medidas de similitud musical. Las anotaciones de los expertos que fueron consideradas fueron donde hubo unanimidad, es decir donde la anotación de los tres expertos fue la misma. Se evaluaron las medidas previamente descritas calculando la curva ROC por cada una y obteniendo el umbral de mejor desempeño en tres tareas distintas de clasificación: entre clases 0 y 0.5, 0.5 y 1, y 0 y 1. El mejor umbral es aquel con la menor diferencia al punto ideal en la curva ROC, en donde la tasa de falsos positivos se minimiza y se maximiza la tasa de verdaderos positivos. Por cada problema de clasificación se obtuvo el valor F_1 para medir su desempeño, acompañado de los valores de precisión y exhaustividad (valores más altos se refieren a un mejor desempeño).

La exhaustividad se conoce también como la razón de verdaderos positivos. Es calculada dividiendo el número de verdaderos positivos entre el total de miembros de la misma clase, es decir, los verdaderos positivos junto con los falsos negativos.

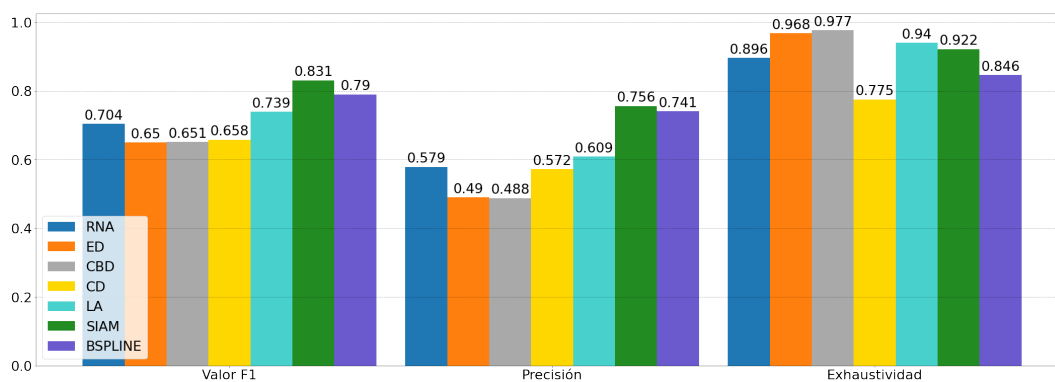
$$\text{exhaustividad} = \frac{TP}{TP + FN} \quad (4.10)$$

La precisión o el valor predictivo de positivos, se define como la división entre el número de verdaderos positivos entre el número de verdaderos positivos mas falsos positivos.

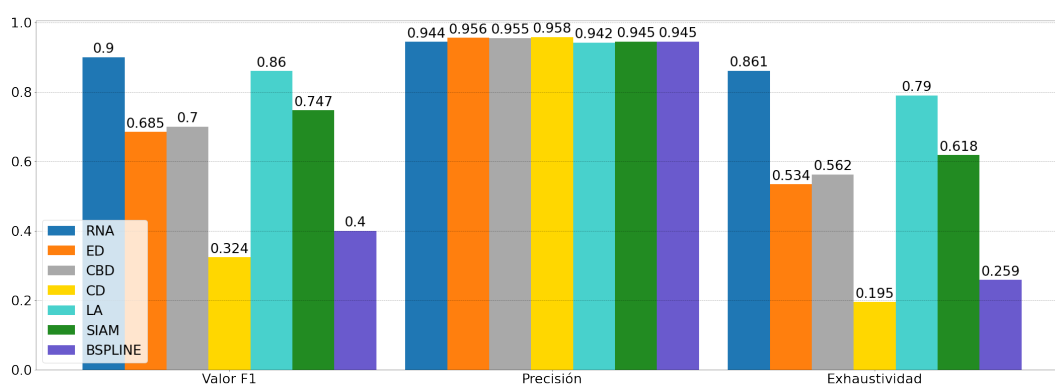
$$\text{precisión} = \frac{TP}{TP + FP} \quad (4.11)$$

El valor F_1 está definido como la media armónica entre la precisión y la exhaustividad. Este tipo de media mitiga la influencia de valores muy distantes entre sí y es apropiada para este tipo de tasas.

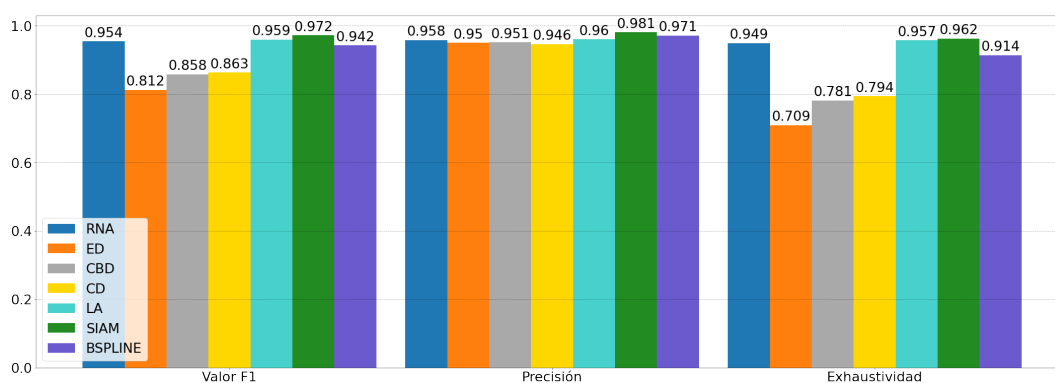
$$F_1 = 2 \frac{\text{precisión} \cdot \text{exhaustividad}}{\text{precisión} + \text{exhaustividad}} \quad (4.12)$$



(a) 0-0.5



(b) 0.5-1



(c) 0-1

Figura 4.7: Desempeño de las distintas medidas de similitud en las tareas de clasificación: a) “casi idénticas” ó “relacionadas pero variadas”, b) “relacionadas pero variadas” ó “diferente”, y c) “casi idénticas” ó “diferente”.

En la primera gráfica de la Figura 4.7 es en donde los valores obtenidos con la medida propuesta *RNA* son menores a los de las métricas. Es también la tarea de clasificación en donde el desempeño en general de todas las métricas es bajo, especialmente en el caso de la precisión, su exhaustividad no se ve tan afectada, pero tomando la relación de ambas para calcular el valor F1 este se ve en desventaja en comparación a las demás tareas de clasificación. En la segunda gráfica sucede un caso inverso en relación a la precisión y la exhaustividad, ahora los menores valores se encuentran en la exhaustividad. Sin embargo, los resultados para *RNA* tiene la menor variación entre ambas, lo que hace que tenga el mejor valor F1 en esta tarea de clasificación. La tercera gráfica es la que representa el caso más común en tareas de clasificación, en donde únicamente se busca segregar melodías casi idénticas de las diferentes. En general en esta prueba se obtuvieron los mejores resultados para todas las métricas teniendo muy poca diferencia en sus resultados, con la excepción de la exhaustividad. En esta prueba la mejor resultó ser *SIAM* y por debajo con muy poca diferencia *LA* y la medida propuesta *RNA*. En dos de las tres tareas de clasificación posibles con este conjunto de datos *SIAM* tuvo los mejores resultados, superada por *RNA* en la segunda clasificación y estando cerca en la tercera prueba. Sin embargo, los resultados mostrados en la Figura 4.7 son inconclusos en términos de establecer cuál es la mejor medida de clasificación debido a la variación de las mismas en las distintas pruebas. Para estos efectos en la siguiente sección tiene como objetivo probar algunas de estas medidas en un ambiente real de co-creatividad.

4.3. Caso de estudio con Impro-Visor

En esta sección se presenta un caso de estudio en el que un músico usó el software de improvisación Impro-Visor (Keller, 2012), para co-crear distintas melodías. La medida propuesta fue utilizada (al igual que otras medidas de similitud) para evaluar si el ambiente co-creativo es capaz de empujar al músico a otros panoramas musicales novedosos para él. Se le pidió al músico que creara cuatro melodías en su propio estilo. Una vez que termino, se le pidió que crearía una quinta melodía que intencionalmente sea diferente a las otras cuatro. Este conjunto será referido como *User_SOLO*.

Impro-Visor fue configurado para crear dos conjuntos de cinco melodías cada uno. Uno de los conjuntos fue usando la gramática de generación de Miles Davis (*MilesDavis_SOLO*), el otro conjunto fue usando la gramática de Charlie Parker (*CharlieParker_SOLO*). Las razones por las que estas gramáticas fueron elegidas por el músico son: a) Parte de su trabajo es “cercano” al de Miles Davis; y b) Charlie Parker es considerablemente “lejano” al suyo.

Posteriormente, Impro-Visor Fue configurado para que el músico fuera capaz de realizar un intercambio en una improvisación musical con el software y así crear otros dos conjuntos de 5 melodías cada uno. Los dos conjuntos se generaron utilizando la gramática de cada uno de los artistas de referencia, a estos conjuntos nos referiremos como: *MilesDavis_TRADE* y *CharlieParker_TRADE*.

A continuación se presentará una visualización de los resultados basada en la propuesta de Alison Pease Pease et al. (2001).

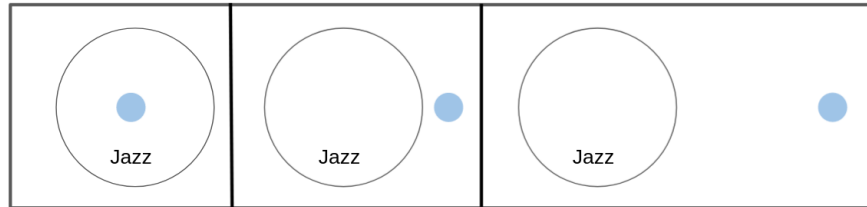


Figura 4.8: Visualización ejemplo del campo conceptual de Jazz y la posición de una melodía que no es nueva (izquierda), fundamentalmente nueva (centro) o demasiado caótica para considerarse nueva (derecha).

En la Figura 4.8, una melodía ejemplo se ubica en 3 distintas zonas dependiendo de su condición de novedad. Una formalización de esta visualización puede cuantificar dicha distancia similar a la siguiente figura.

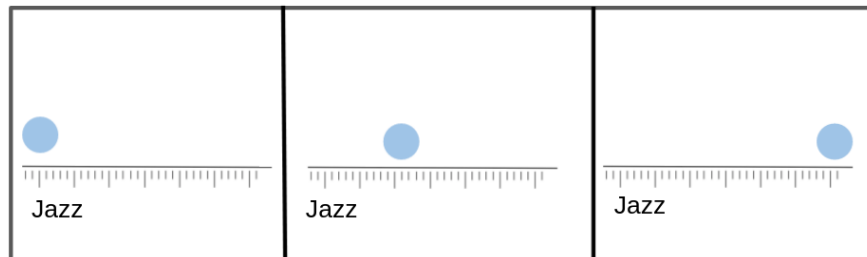


Figura 4.9: Visualización ejemplo del campo conceptual de Jazz y la cuantificación de la distancia de una melodía que no es nueva (izquierda), fundamentalmente nueva (centro) o demasiado caótica para considerarse nueva (derecha).

En la Figura 4.9, El valor donde se ubica a la melodía es el valor de la distancia promedio entre la melodía ejemplo y las melodías presentes en el conjunto de datos. Posteriormente se extiende esta idea al considerar dos ejes simultáneamente para ubicar en dos dimensiones cada melodía dependiendo de su distancia a dos conjuntos de datos.

4. RNA: UNA MEDIDA DE SIMILITUD MUSICAL

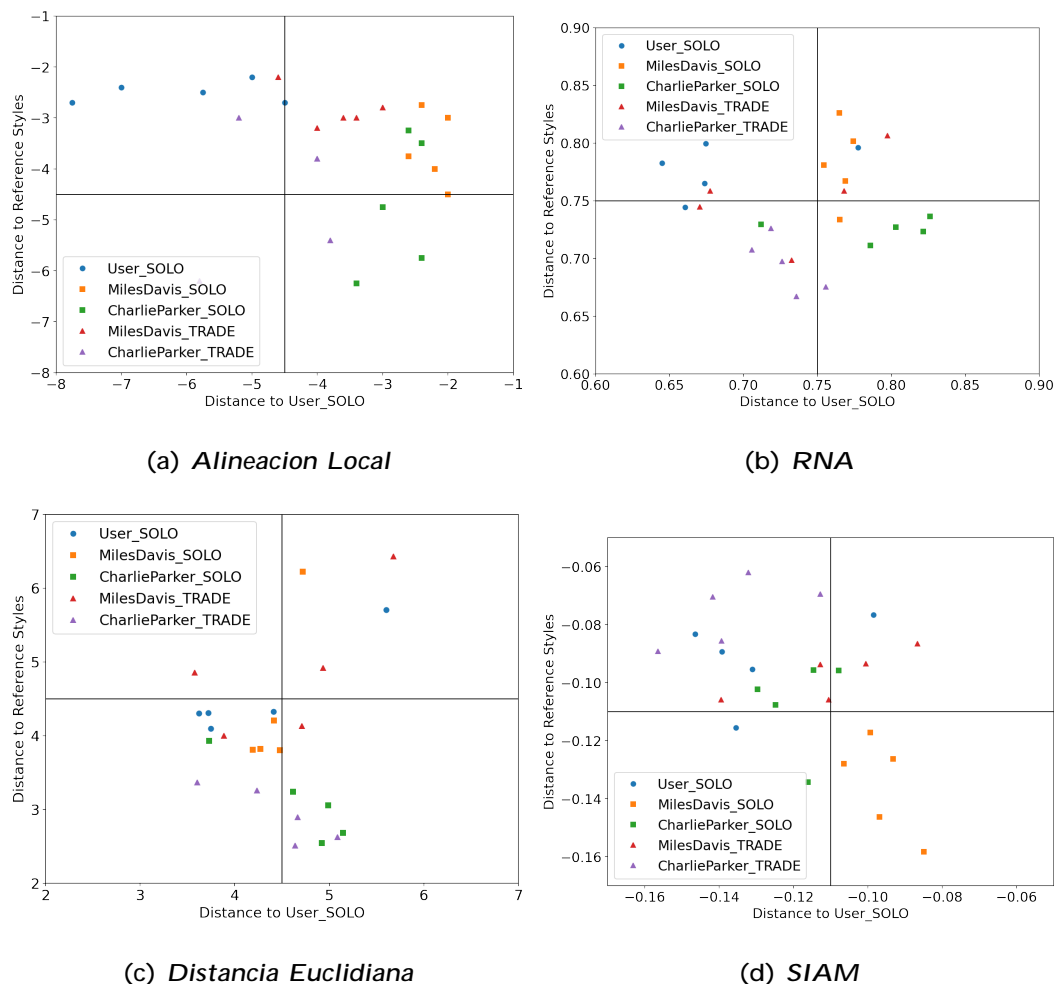


Figura 4.10: Medidas de similitud en un proceso co-creativo.

En la Figura 4.10, los resultados se muestran en una representación que define un tipo de plano de similitud, asumiendo que cada conjunto de melodías también representa un estilo musical específico. Cada sub-figura muestra el plano de similitud con la medida propuesta (*RNA*), *LA*, *ED* y *SIAM*. Estas medidas fueron escogidas puesto que: *SIAM* tuvo el segundo mejor desempeño (por debajo de *RNA*), *LA* es la medida de la cual se derivó *RNA* y *ED* es una medida básica que provee una base de referencia.

El eje horizontal de cada sub-figura muestra la distancia promedio entre cada melodía y el conjunto de melodías del músico que creó por sí solo, es decir *User_SOLO*. En términos del eje vertical: a) si las melodías son del conjunto *User_SOLO*, muestra la distancia promedio a los otros conjuntos *SOLO* de los dos artistas de referencia; b) si es de otro conjunto de *SOLO*, muestra la distancia promedio hacia las melodías del mismo conjunto; c) si es del conjunto de *TRADE*, muestra la distancia promedio

hacia su respectivo conjunto *SOLO*, por ejemplo con *MilesDavis_TRADE* se calcula su distancia a *MilesDavis_SOLO*.

Esta visualización está construida de tal forma que el espacio que ocupa un grupo representa su variación de estilo entre el mismo grupo. El ancho de un grupo puede interpretarse como la variación con respecto al estilo del músico (*User_SOLO*), mientras que su altura es la variación con respecto a los estilos de referencia (*MilesDavis_SOLO* y *CharlieParker_SOLO*).

El valor de distancia en ambos ejes implica una “distancia” entre las melodías, por lo que mientras más pequeño sea el valor, más alta será su similitud.

Observando a la Figura 4.10 podemos notar que las distintas métricas logran segregar los distintos conjuntos de melodías en distintas formas cada una. *RNA* parece tener los grupos más claramente definidos que el resto de las métricas, sugiriendo que es más apropiada para juzgar elementos de un mismo estilo. *ED* siendo una métrica simple relativamente, fue capaz de reconocer la melodía del conjunto *User_SOLO* que es intencionalmente distinta al resto, de igual forma mantiene ciertos grupos cercanos, aunque no tan consistentemente teniendo melodías dispersas en cada grupo.

Tomando en cuenta el eje horizontal en *RNA* existe una tendencia visible de separar el conjunto del usuario (círculos) de los conjuntos de los artistas de referencia (cuadrados), colocando a los primeros a la izquierda (cercaos al estilo del usuario) y a los segundos a la derecha (lejanos al estilo del usuario). Para los intercambios en improvisación (triángulos) en donde se tiene una especie de mezcla de estilos, son colocados en el medio del plano. Siendo justos una tendencia similar está presente en *LA*, aunque teniendo las carencias presentes en *ED* en donde los grupos no están definidos claramente. Tanto *ED* y *SIAM* carecen de esta segregación horizontal de los conjuntos. En el caso del eje vertical *RNA* es capaz de mostrar que el intercambio con Miles Davis (*MilesDavis_TRADE*) tiene una distancia similar al conjunto del usuario, lo que va acorde con la presunción de una variación similar dentro de ambos estilos. También, el grupo del intercambio con Charlie Parker (*CharlieParker_TRADE*) se encuentra más abajo que el conjunto del usuario (*USER_SOLO*) haciendo ver la consistencia presente en el estilo de Charlie Parker hace que acerque otros estilos a él. Esta consistencia se observa al ubicar el conjunto *CharlieParker_SOLO* en el plano, encontrándose con una mínima variación en el eje vertical, a diferencia del conjunto de Miles Davis (*MilesDavis_SOLO*).

4.4. Discusion

En esta sección se realizó la evaluación de la medida propuesta (*RNA*) en comparación al subconjunto de métricas de similitud musical en tres distintos escenarios.

En el entorno de evaluación propuesto en este trabajo por porcentajes y tipos de transformación *RNA* tuvo mejores que el resto de las métricas, teniendo únicamente a *SIAM* como la métrica con los resultados más cercanos. *RNA* tiene una mayor consistencia en a lo largo de los distintos porcentajes a diferencia de *SIAM* que tiene un mejor desempeño al diferenciar entre melodías o muy similares o muy diferentes careciendo de un mejor criterio al considerar melodías al medio. Esto es importante en sesiones creativas puesto que pueden permitir a los usuarios humanos adaptarse de una manera fluida en lugar del comportamiento por extremos de *SIAM*. Las medidas relativamente simples como *CBD* tienen buenos desempeños al establecer que dos melodías son similares en distintos tipos de transformación, aunque teniendo considerablemente malos resultados en el resto de transformaciones. Cabe mencionar que la medida de la que se basó *RNA*, Alineación Local (*LA*) tuvo peores resultados en una proporción significativa por lo que podemos concluir que las modificaciones realizadas lograron cumplir el objetivo de mejorar su desempeño consistentemente con distintos parámetros de transformación.

En el escenario donde se evaluaron la precisión, exhaustividad y valor F1 en distintas tareas de clasificación, *RNA* tiene el mismo o mejor resultado en la mayoría de las circunstancias, con la excepción de la tarea que consistía en separar las melodías “casi idénticas” (0) de las “relacionados pero variados” (0:5), en dicha tarea fue superada por *BSPLINE* y *SIAM*. En esta tarea de clasificación, incluso es superada en su exhaustividad por *CBD* y *ED*. En general en esta tarea de clasificación todas las métricas tuvieron valores menores que en las demás tareas. Podemos inferir entonces que discriminar melodías de la clase 0 y de la clase 0:5 es un problema difícil para las métricas de similitud musical que fueron analizadas en este trabajo. En contraste, en la tarea de clasificación entre las melodías “casi idénticas” (0) y “diferentes” (1) las métricas tuvieron su mejor desempeño lo que indica que es sencillo generalmente diferenciar estas clases. Ahora, para la última tarea de clasificación entre las clases 0:5 y 1, *RNA* logró superar a las demás métricas. Cabe mencionar que no hubo gran diferencia entre *LA* y *RNA* lo que sugiere que las modificaciones realizadas no tienen mucho impacto en el desempeño de estas tareas de clasificación. Las medidas relativamente simples como *CBD* y *ED* tienen comportamientos similares, lo cual se explica a que únicamente calculan una distancia entre vectores de notas.

En el tercer y último escenario evaluado, el caso de estudio con Impro-Visor, el plano mostrado por *RNA* tiene una facilidad de interpretación es parcialmente debido a que tiene un rango definido y cerrado de valores entre 0 y 1. *LA*, de la cual esta derivada *RNA*, tiene un rango abierto de valores lo cual dificulta tomar medidas como pequeñas, medianas o altas. En el caso de *SIAM* aunque trabaja con un rango cerrado de valores, en la prueba realizada no parece dar valores muy variados por lo que se dificulta en-

contrar una escala adecuada para reconocer la discriminación que realiza y así evaluar la novedad entre los grupos. Y en el caso de *ED* definitivamente se requiere un paso de normalización puesto que tiene grupos definidos en una región del plano mientras que al mismo tiempo tiene grupos esparcidos aun siendo del mismo estilo. Los resultados sugieren que *RNA* es más consistente y fácil de interpretar las cuales son características de gran utilidad en un escenario co-creativo. Su rango cerrado de valores apoya para la evaluación de novedad, y al ser continuo permite observar de manera fluida un intercambio entre estilos pudiendo juzgar la distancia entre ellos. Eso hace a *RNA* una medida apropiada para comparar la novedad obtenida al utilizar un software de improvisación como Impro-Visor. Considerando el plano mostrado por *RNA* se observa que la variación de valores entre los conjuntos del usuario por su cuenta y al realizar una improvisación con las gramáticas de los artistas de referencia se ven influenciados y apunta a que Impro-Visor guía al usuario en dirección a crear conjuntos de melodías diferentes que no hubieran sido exploradas al trabajar solo.

A lo largo de las distintas pruebas realizadas para evaluar la medida propuesta y compararla con otras métricas, se observan tareas donde tiene excelentes resultados y otros en los que tiene carencias, sin embargo, las principales fortalezas con la que fue pensada al desarrollarla han logrado su objetivo. Es capaz de dar una perspectiva granular al comparar estilos utilizando un conjunto pequeño características que no requieren de un complejo pre-procesamiento y además que están presentes de manera nativa en un formato tan popular como MIDI. La variedad de pruebas y escenarios hacen ver que en la mayoría de los casos tuvo un buen desempeño, lo cual era otro objetivo, no solo poder evaluar conjuntos de datos establecidos sino ser lo suficientemente robusta como ser utilizada en un caso de la vida real en una sesión con un agente artificial co-creativo.

Capítulo 5

Conclusiones

Establecer la novedad del producto de un proceso creativo es un tema importante de discusión en distintas áreas. En la música se puede tomar el acercamiento utilizado en este trabajo al medir la similitud de piezas representadas simbólicamente. En el capítulo 2 se introdujo al lector de manera general a los temas de creatividad computacional, compositores musicales artificiales co-creativos y a los marcos de evaluación de los sistemas co-creativos. Posteriormente, en el capítulo 3 se describieron las medidas de similitud musical más usadas, por sus distintas representaciones.

La aportación principal de este trabajo es la definición de *RNA* o *Rhythm-Note Alignment* en la sección 4, una nueva medida de similitud musical que se deriva de la existente medida *Alineación Local LA*, incluyendo elementos rítmicos y de clase de nota musical con el propósito de mejorar su desempeño de evaluación. Se puso a prueba esta hipótesis en tres entornos de evaluación descritos en la sección 4.2. Como aportación menor se propuso una evaluación consistente de 5 transformaciones aplicadas continuamente a la colección de melodías utilizada. En este primer entorno de evaluación *RNA* superó significativamente al resto de las medidas al calcular la diferencia de la distancia calculada por las medidas y la distancia esperada entre la versión modificada y la versión original de la melodía. Aunque inconclusa el resultado del segundo entorno de evaluación, que consistió en tareas de clasificación binarias anotadas por expertos, creemos que son lo suficientemente favorecedores como para pasar a un tercer entorno de evaluación de un caso de estudio de una interacción real entre un músico y un agente artificial co-creativo. En este tercer entorno de evaluación las características, como su rango cerrado de valores, le dieron la ventaja a *RNA* en la representación de los planos de similitud, puesto que facilitaron su interpretación y estudio. De manera adicional pudo definir grupos de melodías pertenecientes a un estilo en particular y ubicando intuitivamente a los grupos de melodías que fueron productos colaborativos entre el músico y el sistema co-creativo, en este caso Impro-Visor.

5. CONCLUSIONES

Como trabajo a futuro se pueden explorar diversos cuestionamientos que surgen con las propuestas realizadas en este trabajo, como lo son experimentar con distintas configuraciones del sistema co-creativo, aplicar los entornos de evaluación en distintos conjuntos de datos con melodías en otros estilos así como extender la evaluación por planos de evaluación en otros sistemas co-creativos para trabajar en la viabilidad de incorporar este tipo de evaluación en el flujo de colaboración en sesiones creativas entre agentes humanos y artificiales.

Bibliograf a

- Alvarado, F. H., Lee, W.-H., Huang, Y.-H., y Chen, Y.-S. (2020). Melody similarity and tempo diversity as evolutionary factors for music variations by genetic algorithms. In *ICCC*. [74](#)
- Bogdanov, D., Serra, J., Wack, N., Herrera, P., y Serra, X. (2011). Unifying Low-Level and High-Level Music Similarity Measures. *IEEE Transactions on Multimedia*, 13(4):687–701. Conference Name: IEEE Transactions on Multimedia. [49](#), [51](#), [52](#), [53](#), [55](#)
- Bretan, M. y Heck, L. (2019). LEARNING SEMANTIC SIMILARITY IN MUSIC VIA SELF-SUPERVISION. *ISMIR 2019*, page 8. [72](#)
- Brown, A. R. (2018). Creative improvisation with a reflexive musical bot. *Digital Creativity*, 29(1):5–18. [26](#)
- Chakrabarty, S., Chatterjee, S., Islam, M. R., y Kumar Deva Sarma, H. (2019). Music Similarity Mapping through Fundamental Frequencies by Coefficient of Concurrent Deviation. In *2019 International Conference on Computing, Power and Communication Technologies (GUCON)*, pages 910–915. [58](#)
- Chew, E. (2002). The spiral array: An algorithm for determining key boundaries. In *Proceedings of the Second International Conference, ICMAI 2002*, pages 18–31. Springer. [74](#)
- Chew, E. y Chen, Y.-C. (2005). Real-time pitch spelling using the spiral array. *Computer Music Journal*, 29(2):61–76. Publisher: MIT Press. [74](#)
- Cleveland, J., Cheng, D., Zhou, M., Joachims, T., y Turnbull, D. (2020). Content-based Music Similarity with Triplet Networks. *arXiv:2008.04938 [cs, eess]*. arXiv: 2008.04938. [56](#)
- Dannenberg, R. y Raphael, C. (2006). Music score alignment and computer accompaniment. *Commun. ACM*, 49:38–43. [67](#)
- Davis, N., Hsiao, C.-P., Popova, Y., y Magerko, B. (2015). An Enactive Model of

- Creativity for Computational Collaboration and Co-creation. In Zagalo, N. y Branco, P., editors, *Creativity in the Digital Age*, Springer Series on Cultural Computing, pages 109–133. Springer, London. [6](#)
- Donahue, C., Mao, H. H., Li, Y. E., Cottrell, G. W., y McAuley, J. (2019). LakhNES: Improving multi-instrumental music generation with cross-domain pre-training. *arXiv:1907.04868 [cs, eess, stat]*. arXiv: 1907.04868. [33](#)
- Dong, H.-W., Hsiao, W.-Y., Yang, L.-C., y Yang, Y.-H. (2017). MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment. *arXiv:1709.06298 [cs, eess]*. arXiv: 1709.06298. [13](#)
- Duggan, B. y O’Shea, B. (2011). Tunepal: Searching a digital library of traditional music scores. *OCLC Systems & Services*. [67](#)
- Fan, Y. y Chen, N. (2019). Music similarity model based on CRP fusion and Multi-Kernel Integration. *Multimedia Tools and Applications*, 78(12):16245–16260. [64](#)
- Gillick, J., Roberts, A., Engel, J., Eck, D., y Bamman, D. (2019). Learning to Groove with Inverse Sequence Transformations. *ICML*. [34](#)
- Guzdial, M. y Riedl, M. (2019). An Interaction Framework for Studying Co-Creative AI. *arXiv:1903.09709 [cs]*. arXiv: 1903.09709. [10](#)
- Hadjeres, G. y Nielsen, F. (2020). Anticipation-RNN: enforcing unary constraints in sequence generation, with application to interactive music generation. *Neural Computing and Applications*, 32(4):995–1005. [18](#)
- Hadjeres, G., Pachet, F., y Nielsen, F. (2017). DeepBach: a Steerable Model for Bach Chorales Generation. *arXiv:1612.01010 [cs]*. arXiv: 1612.01010. [30](#)
- Hamanaka, M., Hirata, K., y Tojo, S. (2015). Applying Melody-Morphing Method to Composition. *ICMC*, page 16. [23](#)
- Hao, T. H. (2019). ChordAL: A Chord-Based Approach for Music Generation using Bi-LSTMs. *In ICCG*, page 2. [15](#)
- Herremans, D. y Chew, E. (2019). MorpheuS: generating structured music with constrained patterns and tension. *IEEE Transactions on A ective Computing*, 10(4):510–523. arXiv: 1812.04832. [37](#)
- Hsu, W. y Sosnick, M. (2009). *Evaluating Interactive Music Systems: An HCI Approach*. NIME. [9](#)
- Huang, C.-Z. A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., y Howcroft, J. (2019). The Bach Doodle: Approachable music composition with machine learning at scale. *arXiv:1907.06637 [cs, eess, stat]*. arXiv: 1907.06637. [21](#)
- Janssen, B., van Kranenburg, P., y Volk, A. (2015). A comparison of symbolic similarity

-
- measures for finding occurrences of melodic segments. In *16th International Society for Music Information Retrieval Conference, 2015*. 70
- Janssen, B., van Kranenburg, P., y Volk, A. (2017). Finding Occurrences of Melodic Segments in Folk Songs Employing Symbolic Similarity Measures. *Journal of New Music Research*, 46(2):118–134. 65, 66, 67, 68, 71
- Johnson, D. D., Keller, R. M., y Weintraut, N. (2017). Learning to Create Jazz Melodies Using a Product of Experts. In *ICCC.*, page 8. 27
- Kantosalo, A. y Riihiaho, S. (2019). Experience evaluations for human–computer co-creative processes – planning and conducting an evaluation in practice. *Connection Science*, 31(1):60–81. 11
- Kantosalo, A., Toivanen, J. M., y Toivonen, H. (2015). Interaction Evaluation for Human-Computer Co-creativity: A Case Study. *ICCC*, page 8. 9
- Karimi, P., Grace, K., Maher, M. L., y Davis, N. (2018). Evaluating Creativity in Computational Co-Creative Systems. *arXiv:1807.09886 [cs]*. arXiv: 1807.09886. 6, 8, 10
- Keller, R. M. (2012). Continuous improvisation and trading with impro-visor. *Proceedings of the Third International Conference on Computational Creativity*. 90
- Kitahara, T., Giraldo, S., y Ramírez, R. (2018). JamSketch: Improvisation Support System with GA-Based Melody Creation from User’s Drawing. In Aramaki, M., Davies, M. E. P., Kronland-Martinet, R., y Ystad, S., editors, *Music Technology with Swing*, Lecture Notes in Computer Science, pages 509–521, Cham. Springer International Publishing. 28
- Lim, H., Rhyu, S., y Lee, K. (2017). Chord Generation from Symbolic Melody Using BLSTM Networks. *ISMIR*. 25
- Liu, H.-M. y Yang, Y.-H. (2018). Lead Sheet Generation and Arrangement by Conditional Generative Adversarial Network. *arXiv:1807.11161 [cs, eess]*. arXiv: 1807.11161. 16
- Lubart, T. (2005). How can computers be partners in the creative process: Classification and commentary on the Special Issue. *International Journal of Human-Computer Studies*, 63(4):365–369. 5, 11
- Meredith, D. (2006). Point-set algorithms for pattern discovery and pattern matching in music. In Crawford, T. y Veltkamp, R., editors, *Content-based Retrieval (Dagstuhl Seminar No. 06171)*, number 06171 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany. 68, 82
- Meredith, D. (2016). Computational music analysis. *Springer, Cham*. 68
-

- Mongeau, M. y Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities*, 24(3):161–175. [67](#)
- Müllensiefen, D. y Frieler, K. (2006). Evaluating different approaches to measuring the similarity of melodies. In Batagelj, V., Bock, H.-H., Ferligoj, A., y Žiberna, A., editors, *Data Science and Classification*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 299–306. Springer. [77](#)
- Panteli, M. y Dixon, S. (2016). On the evaluation of rhythmic and melodic descriptors for music similarity. *New York City*, page 7. [59](#)
- Pati, A., Lerch, A., y Hadjeres, G. (2019). Learning to Traverse Latent Spaces for Musical Score Inpainting. *arXiv:1907.01164 [cs, eess, stat]*. arXiv: 1907.01164. [20](#)
- Pease, A., Winterstein, D., y Colton, S. (2001). Evaluating Machine Creativity. *ICCBR Workshop on Creative Systems*, page 9. [41](#), [91](#)
- Ravikumar, P. T., McGee, K., y Wyse, L. (2018). Back to the experiences: Empirically grounding the development of musical co-creative partners in co-experiences. *ICCC*, page 7. [11](#)
- Ritchie, G. (2007). Some Empirical Criteria for Attributing Creativity to a Computer Program. *Minds and Machines*, 17(1):67–99. [45](#)
- Roberts, A., Engel, J., Raffel, C., Hawthorne, C., y Eck, D. (2019). A Hierarchical Latent Vector Model for Learning Long-Term Structure in Music. *arXiv:1803.05428 [cs, eess, stat]*. arXiv: 1803.05428. [22](#)
- Scherrer, D. K. y Scherrer, P. H. (1971). An Experiment in the Computer Measurement of Melodic Variation in Folksong. *The Journal of American Folklore*, 84(332):230–241. Publisher: University of Illinois Press. [65](#), [82](#)
- Silva, D. F. y Yeh, C.-C. M. (2016). Simple: assesing music similarity using subsequences joins. *New York City*, page 7. [63](#)
- Simonetta, F., Carnovalini, F., Orio, N., y Rodà, A. (2018). Symbolic Music Similarity through a Graph-Based Representation. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion*, AM’18, pages 1–7, New York, NY, USA. Association for Computing Machinery. [73](#)
- Steinbeck, W. (1982). *Struktur und Ähnlichkeit : Methoden automatisierter Melodieanalyse* /. Bärenreiter,, Kassel :. [66](#), [67](#), [82](#)
- Stowell, D., Robertson, A., Bryan-Kinns, N., y Plumbley, M. D. (2009). Evaluation of live human–computer music-making: Quantitative and qualitative approaches. *International Journal of Human-Computer Studies*, 67(11):960–975. [8](#)
- Sturm, B. L., Santos, J. F., Ben-Tal, O., y Korshunova, I. (2016). Music transcription modelling and composition using deep learning. *arXiv:1604.08723 [cs]*. arXiv: 1604.08723. [35](#)

- Taylor, C. W. (1988). *Various Approaches to and Definitions of Creativity*. Cambridge University Press, Cambridge. [1](#)
- Tran, S. N., Ngo, S., y Garcez, A. d. (2020). Probabilistic approaches for music similarity using restricted Boltzmann machines. *Neural Computing and Applications*, 32(8):3999–4008. [62](#)
- Urbano, J. (2012). MIREX 2012 symbolic melodic similarity: Hybrid sequence alignment with geometric representations. *MIREX*, page 4. [70](#), [82](#)
- van Kranenburg, P., Janssen, B., y Volk, A. (2016). The meertens tune collections: The annotated corpus (mtc-ann) versions 1.1 and 2.0.1. *Meertens Online Reports, No. 2016-1*. [87](#)
- van Kranenburg, P., Volk, A., y Wiering, F. (2013). A Comparison between Global and Local Features for Computational Classification of Folk Song Melodies. *Journal of New Music Research*, 42(1):1–18. [67](#), [82](#)
- Velarde, G., Weyde, T., y Meredith, D. (2013). An approach to melodic segmentation and classification based on filtering with the haar-wavelet. *Journal of New Music Research*, 42(4):325–345. [71](#)
- Yang, L.-C., Chou, S.-Y., y Yang, Y.-H. (2017). MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation. *arXiv:1703.10847 [cs]*. arXiv: 1703.10847. [31](#)
- Yannakakis, G. N., Liapis, A., y Alexopoulos, C. (2014). Mixed-initiative co-creativity. *FDG*, page 8. [7](#)