

Composición e Interpretación Concurrente de Melodías Musicales en Tiempo Real Sintetizadas en Base a Emociones Humanas utilizando Técnicas de Inteligencia Artificial

Efraín Astudillo ⁽¹⁾, Pedro Lucas ⁽¹⁾, Enrique Peláez ⁽¹⁾
Facultad de Ingeniería en Electricidad y Computación ⁽¹⁾
Escuela Superior Politécnica del Litoral (ESPOL)
Campus Gustavo Galindo, Km 30.5 vía Perimetral
Apartado 09-01-5863. Guayaquil-Ecuador
{ejastudi, pepaluca, epelaez}@espol.edu.ec ⁽¹⁾

Resumen

Este trabajo explora la aplicación de la Inteligencia Artificial con respecto a melodías musicales y las técnicas que soportan la generación de ellas. Se detalla un modelo que representa el conocimiento musical, basado en la revisión del estado del arte y en el criterio de practicantes de la música experimental; además de la descripción de un sistema que soporta este modelo, el cuál procesa la entrada un teclado musical para componer e interpretar melodías de manera concurrente en tiempo real. Este sistema está guiado por la intención emocional del intérprete de tal manera que genere sentimientos en la audiencia comparable a como un artista humano lo haría. La arquitectura del modelo propuesto y el sistema prototipo fueron sometidos a una prueba de estética, prueba de Turing, prueba de percepción emocional con respecto a oyentes, y pruebas de rendimiento. Estas pruebas son analizadas para evaluar el desempeño del enfoque escogido para la solución del problema, en el que se implementaron dos componentes del modelo que fueron: Cadenas de Markov, por su uso para secuencia de elementos, y Lógica Difusa, por su utilización en situaciones inciertas, para este caso en el uso de emociones humanas.

Palabras Claves: Ingeniería del Conocimiento, Composición Musical, Patrones de Melodías, Emociones Humanas, Cadenas de Markov, Lógica Difusa, Inteligencia Artificial.

Abstract

This work explores the application of Artificial Intelligence regarding music melodies and the techniques that supports the generation of them. A model that represents the musical knowledge is explained based on a revision for the state of the art and the criteria from experimental music's practitioners; also, a description for a system that supports this model, which processes the input from a musical keyboard for composing and performing real-time concurrent melodies. This system is guided by the emotional intention from the performer such way it generates emotions on the audience compared to what a human artist would do. The architecture of this purposed model and the system prototype were subject to an aesthetic test, Turing test, emotional perception test for listeners, and computational performance test. These tests are analyzed in order to evaluate the effectivity for the selected approach, in which two components were implemented: Markov Chains, because of its use for sequence of elements, and Fuzzy Logic, due to its use in uncertainty situations, for this case, human emotions.

Keywords: Knowledge Engineering, Music Composition, Melody Patterns, Human Emotions, Markov Chains, Fuzzy Logic, Artificial Intelligence.

1. Introducción

Las investigaciones relacionadas con la reproducción musical basada en representaciones abstractas, utilizando técnicas de Inteligencia Artificial, se han enfocado en tres aspectos básicos: composición, improvisación e interpretación expresiva, [1] y [2]. De estos, el aspecto más explorado ha sido la composición musical, como un proceso de *armar* o ensamblar varias partes del sonido de un mismo instrumento para lograr música polifónica, o de varios sonidos distintos para crear obras musicales; como lo define el Diccionario de Música Harvard [3].

Establecer normas generales válidas para la creación musical, aplicadas a cualquier ámbito, es prácticamente imposible; por esta razón, existen estrategias de composición para ámbitos específicos, que dependiendo del entorno sugieren normas de composición, permitiendo la aplicación de técnicas definidas en la Ingeniería de Conocimiento para estructurar estas normas.

Desde el punto de vista de la música, existen estándares para la transmisión de información musical. En [4] se toman en cuenta dos aspectos importantes sobre el conocimiento musical, la teoría musical y el

análisis musical, cuyas interacciones a lo largo del tiempo han generado el conocimiento impartido hoy en la academia. Tomando en cuenta otras áreas, Christopher Alexander [5], quien realizó un estudio exhaustivo en diseño arquitectónico, establece la utilización de patrones y un lenguaje que permite la interacción con esos ellos para alcanzar lo que él denomina la *calidad sin nombre*, que en el ámbito musical sería una característica deseada complicada de comunicar fuera del ámbito de las obras en sí.

Las investigaciones relacionadas a composición musical [2] en su gran mayoría están enfocadas a la música occidental, con énfasis en música clásica, Jazz o Blues, pero poco en la música experimental, donde según Holmes [6] existen innovadoras formas de expresar nuevas ideas para la composición provenientes de un proceso explorativo; donde más que comenzar con una partitura, se comienza con *escuchar*. El tipo de música experimental toma en consideración no sólo criterios elementales de armonía, melodía y ritmo, sino el tipo de sonido que se desee crear.

Las estrategias utilizadas en el ámbito de composición musical algorítmica han sido de gran interés para varias investigaciones como se detalla en [2] y [7], las cuáles son basadas en técnicas de Inteligencia Artificial, donde algunas de ellas utilizan la teoría lingüística para representar reglas de composición, e identificar patrones; tales como, frases melódicas, transiciones, arreglos de armonías, entre otros [2], de esta forma se brinda variabilidad a la salida del algoritmo de composición.

El modelo para la adquisición del conocimiento, desarrollado en esta investigación, permite crear una base de conocimientos que toma en consideración la experiencia humana de dos compositores músicos entrevistados, relacionada a la composición de música experimental y las técnicas de inteligencia artificial aplicadas al proceso creativo de los artistas, bajo el mismo esquema que utilizarían cuando crean este tipo de música. Sobre este modelo se aplicaron estrategias para poder componer e interpretar en tiempo real melodías musicales en un proceso que denominamos *composición Humano-Máquina* donde el músico real interpreta una base armónica que es seguida por el agente artificial, el cual genera melodías musicales.

En la sección 2 se detalla la metodología utilizada en el desarrollo de este trabajo, la sección 3 presenta el modelo correspondiente a la arquitectura de representación del conocimiento musical y la descripción de sus componentes con las estrategias a aplicar, en la sección 4 se describe al sistema concurrente en tiempo real que implementa el modelo, la sección 5 muestra los resultados acerca de la pruebas realizadas, y finalmente, en la sección 6, se presentan las conclusiones y futuros trabajos relacionados a la temática.

2. Metodología

Para la adquisición del conocimiento humano en el proceso de composición musical se aplicó una entrevista no estructurada a dos músicos, siguiendo los lineamientos de Kendal & Creen [8]. Los músicos entrevistados son compositores con formación académica en música y practicantes de la música experimental, además fueron familiarizados con el proceso de adquisición a utilizar. Éste proceso de adquisición será detallado en profundidad en la sección 3.

Los lineamientos extraídos de la entrevista realizada, además de una revisión de literatura con respecto a la temática, permitieron adquirir el conocimiento general del proceso creativo. Siendo este primer componente el conocimiento explícito representado a través del lenguaje natural, y en base a estos lineamientos se diseñó una arquitectura tecnológica de adquisición del conocimiento para capturar un segundo componente que es el conocimiento tácito, que por su particularidad no puede ser articulado o transmitido fácilmente, el cual está implícitamente representado en las melodías musicales.

Posteriormente, se desarrollaron los componentes de la arquitectura tecnológica relacionados con Cadenas de Markov para secuencias melódicas, y Lógica Difusa para la clasificación de emociones, con el objetivo de representar este conocimiento. Este proceso dio lugar a la extracción del conocimiento tácito dado por las grabaciones de secuencias de notas hechas por los músicos para melodías, las cuales fueron manipuladas bajo el estándar MIDI.

El conocimiento extraído fue utilizado en la generación de melodías musicales basadas en emociones por parte de un sistema concurrente en tiempo real, el cual acompaña a un músico humano quien lo guía con la intención emocional que él pretende hacia su audiencia mediante la asignación de niveles de emociones como parte de la entrada de este sistema.

Las experimentaciones realizadas implicaron la realización de una prueba de estética y una de Turing que considera únicamente melodías generadas por el componente de Cadenas de Markov, y una prueba de percepción emocional con oyentes que incluye el componente de Lógica Difusa, además de pruebas de rendimiento computacional del sistema.

Finalmente, se estableció un análisis de la experimentación y sus implicaciones en el ámbito de la Inteligencia Artificial aplicada a la música.

3. Arquitectura para el Sistema de Adquisición y Representación del Conocimiento

Para generar una arquitectura capaz de adquirir y representar el conocimiento implícito que poseen los compositores de música, se ha tomado en consideración las descripciones e idiosincrasias de los entrevistados, y se han estructurado los patrones específicos que les permitan componer con su estilo. La música experimental de los entrevistados es de carácter occidental, por lo tanto, su representación estará basada en el sistema de doce notas musicales. Esta arquitectura se presenta a continuación:

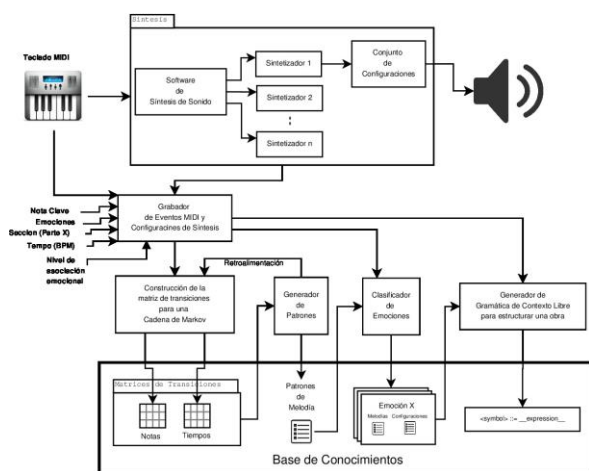


Figura 1. Diagrama de bloques para la arquitectura del sistema de adquisición del conocimiento

A continuación se describen los componentes de esta arquitectura tomando como ejemplo una obra de uno de los compositores con gran influencia en los músicos entrevistados, consideraremos *la Toccata y fuga en re menor, BWV 565 de Johann Sebastian Bach* que es muy conocida mundialmente y controversial por sus elementos únicos para la época.

3.1. Entradas

Tal como se muestra en la figura 1, los datos de entrada son:

- La información de la ejecución de la composición generada por los músicos; es decir, una secuencia de eventos MIDI que consta de tres elementos: nota musical que se está ejecutando (0 a 127), velocidad (0 a 100) y duración (BPM beats por minutos).
- La nota clave (*key*) que rige la interpretación del músico.

- La emoción que expresa el sentimiento de lo que se está interpretando en ese momento (alegría, nostalgia, ira, etc.).
- La sección o tramo de la interpretación, tal como fue asignado por el autor a su composición; es decir, la parte A, B, C, etc., de la composición.
- El tempo o la restricción de tiempo dada en BPM (beats por minuto), guiado por un metrónomo de sincronización entre la interpretación y la grabación que el sistema realiza. Y,
- El nivel de asociación emocional, que etiqueta con un valor entre 0 y 100 la intención que pretende un músico con cada emoción asignada.

Considerando la obra de Bach, podemos estructurar un conjunto hipotético de entradas que recibiría el sistema de adquisición y representación basado en el criterio de los músicos de esta manera:

Teclado MIDI: A, G, A, G, A, E, D, C#, D,..... (Cada una debe ser acompañada por su velocidad y tiempo)
Nota Clave: D menor
Tempo: 80 BPM
Emociones y Partes:

Tabla 1. Partes y emociones de la Toccata y fuga en re menor, BWV 565 de Johann Sebastian Bach

Intervalo de tiempo (mm:ss)	Parte	Emociones
[0:00, 00:24)	A	Ira
...
[02:48, 04:05)	F	Melancolía, Nostalgia, tristeza
...
[08:11, 09:10)	J	Sensación de libertad

3.2. Síntesis

En este módulo se producen los sonidos de las composiciones por medio de sintetizadores digitales de los cuáles se almacenan sus configuraciones con el fin de enlazarlas con los datos recolectados en el módulo de grabación.

En el caso de la obra de Bach es una composición barroca que generalmente es interpretada por un órgano, por lo tanto una configuración de síntesis para esta obra conllevaría a registrar la configuración de elementos de un sintetizador tales como osciladores, filtros y envolventes de tal manera que simulen el modelamiento físico del órgano.

3.3. Grabador de Eventos MIDI y Configuraciones de Síntesis

Este módulo recolecta las entradas y las procesa de acuerdo al requerimiento de los otros módulos del sistema; también, almacena la interpretación combinada con el resto de entradas, de tal manera que puedan reproducirse después. La utilización de aplicaciones de grabación MIDI reduce la complejidad de la implementación.

3.4. Representación del Conocimiento a través de una Matriz de Transiciones con Cadenas de Markov

Las Cadenas de Markov son un proceso estocástico secuencial de estados o eventos dependientes. Donde la transición de un estado al otro depende sólo del estado actual y en su probabilidad de ocurrencia. Los dos componentes principales de las cadenas de Markov son: el conjunto de estados y las transiciones entre ellos [7].

En Nierhaus [7] se propone la representación de cadenas de Markov para composición musical. En base a ello y al conocimiento adquirido de los músicos entrevistados, en este módulo se generan dos matrices, una para las transiciones entre notas musicales; y, otra para las transiciones entre tiempos y duraciones, representados por (*te*, *tp*). Esta representación nos permite a futuro generar composiciones más ricas, combinando patrones de tiempo con patrones de secuencias de notas.

Por ejemplo, las matrices de transiciones de notas y duraciones para la Toccata y Fuga de Bach presentan una particularidad en sus melodías, donde algunas de ellas no se encuentran dentro de la clave D menor, y las duraciones se pueden encontrar en su mayoría representadas por semicorcheas (0.375 segundos si BPM es 80) [9]. A continuación se presenta la matriz de transiciones de notas para esta obra, construida con el algoritmo para las cadenas de Markov.

Tabla 2. Matriz de transiciones de notas de la Toccata y Fuga en re menor, BWV 565 de Johann Sebastian Bach

	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
C	0.06	0.01	0.19	0.04	0.07	0.09	0.02	0.08	0.00	0.11	0.27	0.05
C#	0.02	0.04	0.36	0.00	0.24	0.02	0.00	0.02	0.00	0.15	0.13	0.03
D	0.11	0.08	0.21	0.05	0.12	0.08	0.02	0.10	0.01	0.15	0.06	0.01
D#	0.15	0.02	0.42	0.02	0.00	0.23	0.02	0.08	0.00	0.02	0.02	0.02
E	0.05	0.13	0.19	0.00	0.05	0.23	0.01	0.16	0.00	0.15	0.02	0.00
F	0.04	0.02	0.17	0.05	0.19	0.04	0.00	0.23	0.01	0.17	0.06	0.01
F#	0.02	0.01	0.22	0.09	0.03	0.04	0.01	0.35	0.01	0.18	0.05	0.00
G	0.08	0.01	0.09	0.03	0.14	0.17	0.04	0.05	0.01	0.23	0.11	0.04
G#	0.00	0.00	0.17	0.00	0.08	0.06	0.04	0.17	0.12	0.25	0.00	0.12
A	0.03	0.08	0.17	0.00	0.08	0.12	0.02	0.19	0.00	0.11	0.17	0.04
A#	0.11	0.10	0.10	0.02	0.05	0.02	0.01	0.20	0.01	0.34	0.03	0.01
B	0.24	0.22	0.09	0.01	0.01	0.04	0.00	0.17	0.01	0.17	0.00	0.04

3.5. Generador de Patrones

En este módulo de la arquitectura propuesta, se crean los patrones melódicos usando un corpus de pequeños segmentos de melodías musicales. Cada

segmento es una limitada secuencia de notas con sus respectivas duraciones generadas por las cadenas de Markov y deben cumplir con ciertas reglas musicales usadas en el proceso de composición. Adicionalmente, se agregan otros segmentos a partir de pequeñas melodías musicales que han sido grabadas a través del tiempo por los músicos, de igual manera deben ser previamente analizadas y procesadas para que cumplan con las reglas musicales mencionadas anteriormente. Estos patrones son representados de la siguiente manera:

Patrón de Melodía M: ($S_1, S_2, S_3, \dots, S_n$)

Donde los elementos de M son segmentos de melodías, y cada segmento S_i , contiene: (*nota_i*, *te_i*, *tp_i*)

El proceso para generar estos patrones consiste en combinar N diferentes segmentos del corpus, formando una secuencia concatenada de estos segmentos. Debido a que el corpus puede ser vasto, debemos realizar una búsqueda informada entre las diferentes soluciones. Los resultados de esta búsqueda deben ser las características fundamentales y estructurales de una composición musical que debe cumplir cada patrón y tener coherencia perceptiva en el oyente.

En el proceso de composición existe una retroalimentación al módulo de construcción de matrices para facilitar el proceso iterativo que ocurre durante la composición, y así generar más patrones que son el resultado del autoaprendizaje del sistema. Algunos de esos patrones no necesariamente son congruentes con las emociones que se desean transmitir, de manera que este modelo concibe un módulo clasificador de emociones, para relacionar estos patrones.

Por ejemplo, en la obra de Bach se generan secuencias de notas con largas duraciones y esa secuencia es similar a las notas musicales de una parte que evocaba esperanza, pero con duraciones cortas, posiblemente, el resultado emocional sería más sublime apuntando a sensaciones positivas.

3.6. Clasificador de Emociones a través de Lógica Difusa

Una de las entradas al sistema son las etiquetas de emociones que el músico coloca a sus composiciones. Este módulo permite asignar esas etiquetas a los patrones generados, dependiendo de su nivel de relación con patrones ya etiquetados; es decir, si son patrones nuevos provenientes del generador de patrones discutido en la sección anterior, y deseamos saber si pertenecen a la emoción “alegría”, basta con comparar con los patrones almacenados de alegría, y determinar su grado de cercanía a través de *lógica difusa*; esta comparación también nos permite determinar el grado de membresía que podría tener con respecto a otras emociones. Un proceso similar se

realizaría para los sonidos producidos por el módulo de síntesis cuyo elemento no serían patrones de melodías, más bien configuraciones de síntesis.

Considerando que es posible representar a los elementos involucrados como secuencias que podrían ser almacenadas en vectores, el grado de cercanía puede ser determinado por cálculo de distancias tales como la Euclidian, de Coseno, Manhattan, entre otras.

Este módulo permite la representación del “rango de emociones” que los músicos mencionaron en la entrevista y que desean plasmar en sus composiciones; emociones que en este modelo son representadas a través de funciones de membresía para los patrones.

Por ejemplo, la figura 2 muestra las funciones de membresía correspondientes a patrones melódicos con respecto al ejemplo de emociones identificadas en la tabla 1 para la obra de Bach. Nótese que en el eje X (patrones melódicos), los valores $P_i[]$ serán discretos y representados por arreglos de patrones, pues se debe considerar elementos con grados de membresía similar.

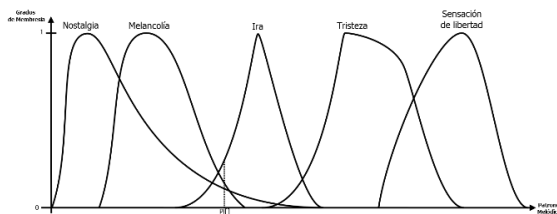


Figura 2. Funciones de Membresía para patrones melódicos con respecto a emociones

3.7. Generador de Gramática de Contexto Libre para Estructurar una Obra

A través de este módulo se logra estructurar la obra como un *todo* a partir de la definición de las secciones (partes), los patrones de melodías y las configuraciones del sintetizador. El producto de este módulo es el conjunto de reglas que representan la gramática de contexto libre, a través de la cual se pueden generar las estructuras de la obra.

Para obtener esta gramática se combinan dos estrategias: Obtener reglas de los expertos, que permitan clasificar las partes, y a través de un mecanismo de *inferencia de gramática* obtener las reglas desde las composiciones.

Por ejemplo, para una gramática de la estructura de Johann Sebastian Bach, considerando el criterio de los músicos y patrones genéricos sería:

```
<Parte A> ::= 'Patrón 1' | 'Patrón 2'....
<Parte C> ::= 'Patrón 9' | 'Patrón 3' |
'Patrón 10' |....
<Parte F> ::= <Parte A> | 'Patrón 5' |
<Parte B> <Parte H> <Parte C> |....
<Parte I> ::= <Parte D> | <Parte F>
'Patrón 6' | ....
```

.....

4. Sistema concurrente en tiempo real para la composición en interpretación de melodías musicales.

Para probar el modelo propuesto en un ambiente en tiempo real, se consideró la construcción de una base de conocimientos cuyo entrenamiento permitió generar patrones melódicos utilizando Cadenas de Markov, y la asignación de emociones para estos patrones junto a su respectiva intención emocional (ponderación entre 0 y 100) para cada una de ellas utilizando Lógica Difusa, como se describió en la sección anterior.

Esta base de conocimientos se utiliza para seleccionar los patrones melódicos adecuados mediante un proceso de defusificación, dependiendo de lo que uno o varios músicos reales estén interpretando en un teclado MIDI, el cual brinda las entradas al sistema. Este proceso comprende un nivel de concurrencia tanto en las entradas como en las salidas como se explicará posteriormente.

4.1. Componentes de Hardware

En la figura 3 se pueden apreciar los componentes físicos que participan en el sistema. En las entradas tenemos N controladores MIDI, que en este caso son teclados musicales, los cuales brindan entradas paralelas hacia el computador a través de una interfaz de audio que viene a ser una tarjeta de sonido externa con procesamiento de audio eficiente. Estas entradas son manipuladas por los algoritmos que el computador ejecuta y son reproducidas a través de los parlantes, completando el ciclo de ejecución en cada intervalo de tiempo. Si el computador soporta las entradas de los controladores directamente, se podría obviar la interfaz de audio siempre y cuando no haya retardos considerables.

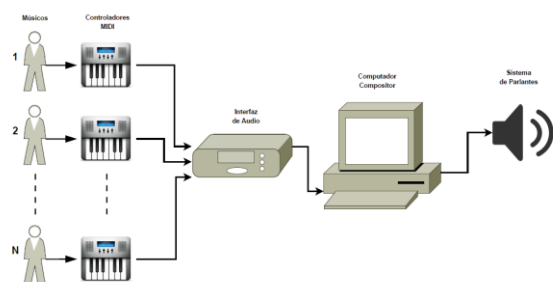


Figura 3. Infraestructura de hardware del sistema

4.2. Componentes de Software

Para el software del sistema se diseñaron componentes que soportan cada etapa del procesamiento, los cuales se detallan a continuación.

4.2.1. Adquisición de Datos. Como se muestra en la figura 4, las entradas deben ingresar al módulo de adquisición de datos, el cual censa el instante en que ingresan, y los transforma en abstracciones entendibles por los procesos consecuentes.



Figura 4. Módulo de adquisición de datos del sistema

4.2.1. Procesamiento Concurrente Temporizado. Este componente permite la generación del material musical en varios hilos de ejecución; es decir que se crearán instancias de las melodías a reproducir de manera paralela aplicando algoritmos que operan sobre la base de conocimientos, de tal manera que ese resultado llegue a un motor de síntesis de sonido como se lo muestra en la figura 5, para reproducirlo por los parlantes.

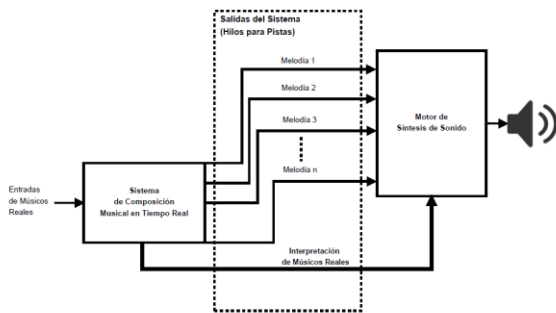


Figura 5. Hilos de ejecución de las melodías generadas por el Sistema

Este proceso es sincronizado por medio de un metrónomo, donde en cada *beat*, se ejecutan los algoritmos de composición musical cuyo tiempo de procesamiento es prácticamente instantáneo, por lo que no ocurren retardos que degradan la estética del resultado. En la figura 6 se muestra este proceso a lo largo del tiempo, el cual es marcado por el metrónomo.

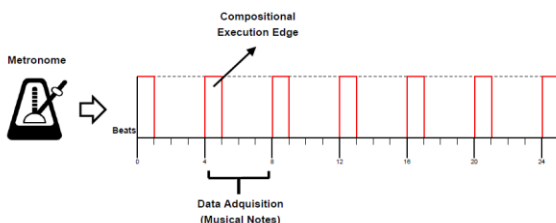


Figura 6. Ejecución del sistema restringida a un metrónomo

4.2.1. Salidas y síntesis de sonido de los resultados.

El resultado de la aplicación de los algoritmos para la generación de melodías en cada hilo de ejecución, es estructurado en un objeto que se compone de dos colecciones, una para *notas* y otra para *duraciones*, donde un elemento corresponde al otro como se ilustra en el ejemplo de la figura 7.

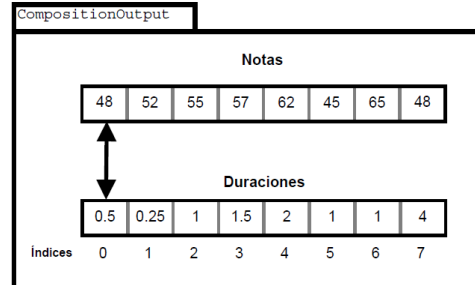


Figura 7. Ejemplo de una melodía generada por el sistema

Las notas son números MIDI y las duraciones son valores relativos al tiempo (BPM) de la obra musical que se está interpretando. La melodía está organizada en este formato para que sea entendible por el motor de síntesis de sonido utilizado, el cual transforma esta información en señales auditivas a través de parlantes completando el ciclo de procesamiento del sistema.

Un prototipo de este sistema fue implementado utilizando las técnicas de Cadenas de Markov, para la generación las secuencias melódicas, y Lógica Difusa, para la manipulación de ellas con respecto a niveles emocionales. Los resultados de las pruebas con respecto al mismo se detallan en la siguiente sección.

5. Resultados

Se pidió a cinco músicos compositores con formación académica evaluar melodías generadas por el sistema y otras producidas por los músicos entrevistados, calificándolas en un rango de 0 a 10, siendo el valor más alto la mejor calidad de la melodía según el criterio de cada músico evaluador.

El banco de melodías compuestas por los músicos experimentales fueron 30; 15 de ellas fueron usadas para el entrenamiento del sistema, y las otras 15 fueron utilizadas para evaluar el sistema. Las 15 melodías de entrenamiento tuvieron un total de 845 notas con un promedio de 56.33 notas para cada secuencia de melodía; lo mismo fue aplicado para las duraciones de cada nota, lo que en conjunto alimentó las matrices de transiciones. También, se generaron 30 melodías por el sistema, luego de ser entrenado, y a través de la estrategia de generación de secuencias para las cadenas de Markov. En total se entregaron 45 melodías a los músicos evaluadores, ordenadas al azar.

Para establecer el tipo de prueba estadística a aplicar, fue necesario verificar la normalidad de las muestras. En la tabla 3 se presentan los resultados obtenidos por el test de normalidad de Shapiro-Wilk.

Tabla 3. Resultados de la prueba de Shapiro-Wilk

Compositor	Valor p	W
Humano	0.003841	0.9479
Máquina	3.895e-07	0.924

Podemos observar que los datos no siguen una distribución normal; por lo tanto, se optó por la prueba Wilcoxon signed-rank. Wilcoxon signed-rank es una prueba no paramétrica que no asume la normalidad de los datos. Es usada para comparar dos grupos que vengan de los mismos participantes, con la finalidad de observar el cambio en los sujetos de un tiempo a otro, o cuando los individuos son sometidos a diferentes condiciones y analizar los efectos causados en los individuos. Para la prueba se asume que los dos tipos de melodías siguen el mismo estilo de los músicos que las generaron.

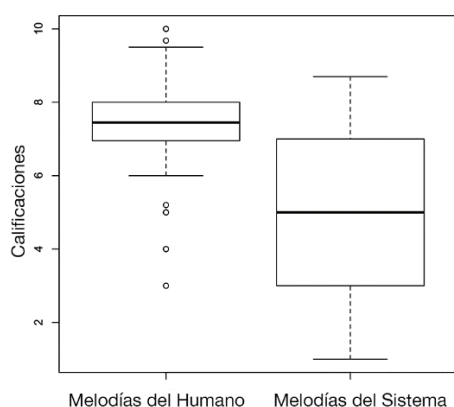


Figura 8. Diagrama de cajas para los grupos de melodías evaluados

Tabla 4. Estadística descriptiva de los dos grupos

Ente	N	Mín	Max	Media	Mediana	Desv. Stand.
Humano	75	3.0	10.0	7.34	7.45	1.45
Máquina	150	1.0	8.70	4.902	5.00	2.10

En la figura 8 y la tabla 4 se muestra un diagrama de cajas y la estadística descriptiva entre los dos grupos de melodías. Se puede observar que las medianas difieren significativamente, lo cual fue comprobado por la prueba de Wilcoxon signed-rank, como se muestra en la tabla 5, con el 95% de confiabilidad.

Tabla 5. Valores de la prueba Wilcoxon signed-rank

Valor p	W
4.733e-16	9355

La hipótesis nula en la prueba Wilcoxon signed-rank define que las diferencias de medianas entre pares de observaciones es igual a cero, por consiguiente el valor **p**, obtenido en la investigación muestra que existe una diferencia estadística significativa de las medianas de estos dos grupos de melodías.

Con respecto a la prueba de Turing realizada, se tomaron en cuenta los aciertos con respecto a la clasificación en cada grupo (humano o sistema) como lo muestra la figura 9.

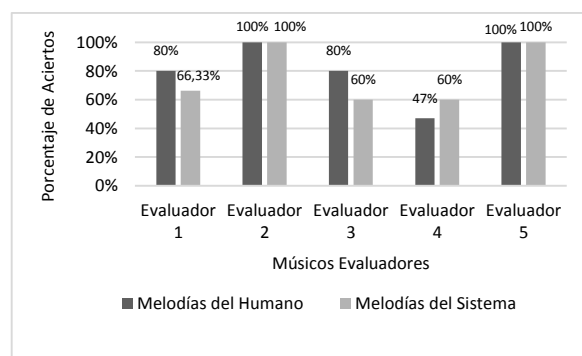


Figura 9. Porcentajes de aciertos para la prueba de Turing

Nótese que dos de los evaluadores (el 2 y el 5) lograron clasificar el material correctamente, el resto posee un porcentaje de fallo que indica que hubo melodías que no pudieron ser distinguidas para el grupo al que pertenecían. Además, los comentarios que ellos creían que eran generadas por humanos tenían estructura y coherencia de inicio a fin, mientras que las demás se sentían más aleatorias y con una estructura difusa, adicionalmente mencionaron que la calidad pudiera aumentar si existiese una base armónica que soportase cada melodía.

La siguiente prueba considera los componentes de Cadenas de Markov, y el Clasificador de Emociones basado en Lógica Difusa.

El sistema fue entrenado con 15 melodías de un promedio de 30.0 segundos por cada una. Las emociones seleccionadas por el músico que proporcionó el entrenamiento fueron: *felicidad, serenidad, tristeza, nostalgia, apasionamiento* las cuales fueron ponderadas con un nivel emocional entre 0 y 100 para cada melodía, dependiendo de la intención emocional, además la nota clave y el tempo fueron proporcionados durante el entrenamiento. El sistema generó 30 patrones de melodías.

El compositor humano interpretó la base armónica (acordes) para 15 piezas musicales de una duración de 60 segundos aproximadamente para cada una, además se proporcionó la intención emocional para estas piezas, de tal manera que el sistema interpretó melodías

consistentes con la armonía e intenciones emocionales dadas utilizando el proceso de defusificación sobre los patrones ponderados en la base de conocimientos.

Estas 15 piezas musicales fueron enviadas a personas (oyentes) familiarizadas con la música occidental con el objetivo de que ellos ponderen cada canción con su nivel de percepción emocional según lo que sintieron, 30 personas contribuyeron a la evaluación.

La figura 10 muestra el primero de los 15 diagramas de caja que muestran la intención emocional del músico y las emociones percibidas por los oyentes para una pieza musical humano-máquina.

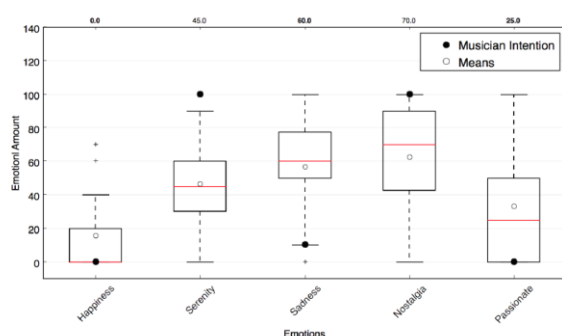


Figura 10. Intención emocional y percepciones para la pieza musical 1

Considerando la figura 10, la intención emocional del músico difiere de la tendencia de la percepción emocional del oyente; sin embargo, esta percepción se inclina de manera similar a la intención; por ejemplo, podemos notar en la ilustración más *serenidad* y *nostalgia* que *felicidad* y *apasionamiento* en las percepciones así como en las intenciones, pero la *tristeza* no se ajusta a este comportamiento. Este fenómeno es similar en el resto de diagramas donde todas las emociones se comportan de manera parecida o con un comportamiento distinto al explicado con *tristeza*.

Otras emociones que fueron sentidas por los oyentes fueron: melancolía, reminiscencia, calma, relajación, depresión, esperanza y ansiedad. Algunas de ellas son relacionadas de alguna manera con las emociones pretendidas por el músico; sin embargo, dada la subjetividad del contexto, las personas podrían tener diferentes percepciones acerca de las emociones relacionadas como *tristeza*, *nostalgia* y *melancolía*. En la tabla 6, presentamos los resultados para una prueba de Levene que es utilizada para evaluar la similitud de varianzas para una variable, calculada para dos o más grupos. En nuestro caso, la prueba es aplicada a cada canción y los grupos por canción son las emociones (*felicidad*, *serenidad*, *tristeza*, *nostalgia*, *apasionamiento*).

Tabla 6. Prueba de Levene para las varianzas de las emociones en cada pieza musical

Canciones	W	Valor p
<i>Pieza Musical 1</i>	1.731	0.146
<i>Pieza Musical 2</i>	0.377	0.824
<i>Pieza Musical 3</i>	1.833	0.125
<i>Pieza Musical 4</i>	1.353	0.253
<i>Pieza Musical 5</i>	3.289	0.0129
<i>Pieza Musical 6</i>	2.816	0.0274
<i>Pieza Musical 7</i>	1.952	0.104
<i>Pieza Musical 8</i>	0.753	0.557
<i>Pieza Musical 9</i>	1.120	0.349
<i>Pieza Musical 10</i>	0.124	0.973
<i>Pieza Musical 11</i>	1.460	0.217
<i>Pieza Musical 12</i>	0.968	0.427
<i>Pieza Musical 13</i>	0.412	0.799
<i>Pieza Musical 14</i>	0.978	0.421
<i>Pieza Musical 15</i>	1.599	0.177

Esta prueba fue aplicada con 95% de confiabilidad y nos dice que la variabilidad para cada canción, con respecto a las emociones, no difiere significativamente excepto por la *Pieza Musical 5* y *Pieza Musical 6*, lo que significa que los oyentes percibieron cada canción con un mismo grado de incertidumbre.

Finalmente, a pesar de estos resultados, los oyentes no reportaron comentario alguno que sugiriese una composición aleatoria de melodías a diferencia de los resultados presentados en el enfoque que solo utilizó Cadenas de Markov. Sin embargo, ellos sintieron algunas piezas con patrones melódicos similares; es decir, se repetían ciertas partes en distintas piezas.

Las pruebas de rendimiento del prototipo del sistema fueron ejecutadas en el contexto de la composición e interpretación de una pieza musical en tiempo real. Para ello, el sistema fue puesto en marcha en un computador con procesador Intel Core I7 de 2.20 GHz, 6 Gb de memoria RAM y sistema operativo Windows 8.1 de 64 bits. Se utilizó un teclado Yamaha PSR-E413 de cinco octavas como controlador MIDI vía USB.

Se simuló la entrada de 8 hilos, los cuales copiaron los datos de los eventos MIDI desde el controlador hacia la máquina, lo que significa que ocho instrumentos musicales son manipulados por músicos reales en el exterior para alimentar el sistema en tiempo real.

Tomando en cuenta este número de entradas, se produjeron salidas del sistema a través de motor de síntesis con varias cantidades de hilos donde cada uno representa la composición en interpretación de la máquina para un instrumento musical que acompaña a los 8 hilos de entrada. Los números de hilos escogidos para realizar las pruebas siguen una escala de Fibonacci desde 1 a 55 hilos de salida (ocho números que son: 1, 2, 3, 5, 8, 13, 21, 34, 55). Se escogió esta escala para

observar lo que sucede en las primeras ejecuciones con pocos hilos y para avanzar más rápidamente en cantidades altas.

Por consiguiente, se realizaron 8 procesos de toma de datos para cada número de la escala con una pieza musical de tiempo 80 BPM, nota clave C Mayor o A menor y ejecutando el algoritmo basado en Lógica Difusa para la composición musical de la máquina para cada hilo de salida. Por cada proceso se tomaron 100 muestras; es decir, como las estrategias de composición se ejecutan cada beat, entonces el tiempo de duración de la toma de datos para cada proceso fue de 100 beats.

En la tabla 7, se detalla la estadística descriptiva para los tiempos de procesamiento cuando un flanco de tiempo (beat) ocurre considerando el número de hilos de salida. En la figura 11 se puede observar el crecimiento del tiempo de procesamiento con respecto al promedio. Según [10], la retroalimentación hacia un usuario debe ser menor o igual a 200 ms para que pueda ser percibida de forma inmediata, y, considerando que el músico debe percibir una respuesta inmediata en la improvisación (composición e interpretación) podemos observar que con menos de 34 hilos se está llegando a ese límite, más de ello resultaría en desincronización del tiempo musical y por ende una interpretación no agradable e imposible de seguir por el músico real.

Tabla 7. Estadística descriptiva de los tiempos de ejecución en milisegundos por número de hilos

Hilos	N	Mín.	Máx.	Media	Mediana	Desv.
1	100	9.00	16.00	11.73	11.50	1.19
2	100	12.00	32.00	17.64	17.00	3.22
3	100	16.00	41.00	24.66	25.00	4.73
5	100	23.00	63.00	37.95	39.00	8.05
8	100	34.00	75.00	56.85	60.00	11.31
13	100	51.00	139.00	95.23	100.00	22.47
21	100	42.00	193.00	116.38	101.00	44.01
34	100	73.00	383.00	219.09	204.00	69.76
55	100	128.00	632.09	380.08	374.00	6137.2



Figura 21. Tiempos Promedio de Procesamiento

Cabe destacar que con 13 hilos de ejecución los sonidos producidos por el motor de síntesis comienzan

a deteriorarse, es decir que se perciben interferencias y ruidos no deseados por la carga de procesamiento que se está soportando. Por lo tanto, el sistema con menos de 13 hilos responde adecuadamente en tiempo real con respecto al procesamiento en CPU, por lo que el sistema puede soportar aproximadamente 21 instrumentos musicales ejecutándose al mismo tiempo si consideramos también las 8 entradas; sin embargo, esto no significa que se obtendrá una buena estética ya que esto depende de los algoritmos que se ejecuten.

Con respecto a la utilización de memoria, el comportamiento de crecimiento con respecto a la cantidad de hilos de salida es mostrado en la figura 12. Esta cantidad de memoria utilizada es debido a que cada hilo carga una copia de la base de conocimientos con el objetivo de evitar accesos continuos a disco y así ahorrar tiempo de procesamiento, además para evitar la complejidad de la gestión de recursos compartidos entre procesos concurrentes.

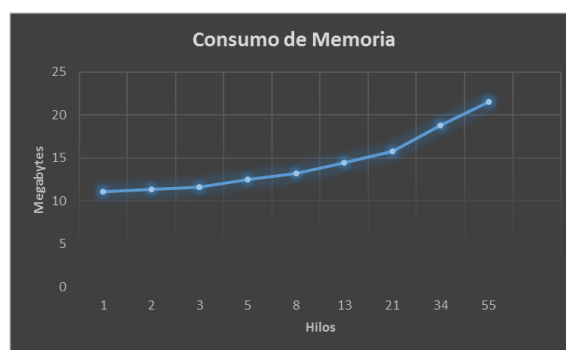


Figura 32. Consumo de memoria por hilos

6. Conclusiones

Considerando los resultados obtenidos en las pruebas realizadas y el planteamiento del tema propuesto, podemos concluir lo siguiente:

- Se levantó el estado del arte con respecto a las técnicas de inteligencia artificial relacionadas al ámbito musical, de donde se tomaron criterios de ingeniería del conocimiento aplicados a artistas practicantes de la música experimental, para guiar a la construcción de una solución tecnológica que componga e interprete melodías musicales de manera concurrente en tiempo real con carácter emocional. Por lo tanto, considerando que los resultados en las pruebas realizadas fueron satisfactorios, el enfoque escogido fue adecuado para el desarrollo de esta solución específica.
- La estrategia de generación de melodías basadas en Cadenas de Markov promete ser un enfoque adecuado cuando se provee una base armónica sobre la cual sea reproducida una melodía en

particular como se lo realizó al agregar el componente difuso. Además, considerando la versión de la prueba de Turing realizada, los resultados de la composición de este enfoque son naturales para varios músicos como si se tratara de un artista real.

- En la interpretación en tiempo real para la realización de las pruebas, se reportó que algunas veces la intención emocional del músico real cambiaba para ser consistente con lo que el sistema estaba reproduciendo; sin embargo, esta situación no afectó significativamente a la composición, es por ello que el sistema no se restringe a lo que se requiere que produzca, sino que contribuye al proceso de composición con su propio *estilo*.
- Este inesperado cambio mencionado en el punto anterior, podría haber causado que la percepción emocional de los oyentes no fuera similar de una manera significativa con la intención emocional del músico real, como lo sugieren los resultados. Además, este experimento no controló el estado emocional de cada oyente, por lo que esto pudo haber influenciado en las respuestas; sin embargo, la variabilidad con respecto a estas respuestas es similar en cada una de las 15 piezas musicales, lo que significa que existe un grado de subjetividad a ser considerado cuando las personas escuchan una canción que evita tener una tendencia marcada.
- A pesar de la conclusión del punto anterior, los oyentes reportaron que sintieron algunas de las emociones establecidas para la prueba, e incluso otras tales como melancolía, reminiscencia, calma, relajación, depresión, esperanza y ansiedad. Estos resultados demuestran que el enfoque propuesto posee influencia sentimental en las personas.
- Las estrategias de representación del conocimiento para músicos experimentales artificiales, deben tomar en cuenta el proceso creativo tanto en lo explícito como lo implícito; es decir, argumentar las representaciones en base a la identificación de sus ideologías, influencias, actividades e incluso equipos como los sintetizadores, lo cual sería lo explícito, y en lo implícito extraer ese conocimiento directamente de sus creaciones.
- La implementación de sistemas de composición e interpretación, que asistan a músicos en su proceso creativo, debe considerar los elementos de la obra que ellos quisieran mejorar estéticamente, esto implica en primera instancia un entendimiento del proceso de composición; como el realizado en este

trabajo, con el objetivo de generar recomendaciones por parte del sistema; sin embargo, existen límites impuestos por los artistas, como el no aceptar que la máquina sea un músico más, y por tanto el enfoque debe ser de *colaboración*, con el objetivo de cortar camino para llegar a componer obras intensamente emocionales y condensadas, tal como lo mencionaron los músicos entrevistados.

- El proceso ingenieril que se desarrolló en este trabajo puede ser considerado como un conjunto de directrices que aporten al grupo de conocimientos adquiridos por un estudiante de ingeniería según los objetivos del proyecto ANDAMIOS, que ejecuta el Centro de Tecnologías de Información de ESPOL, y fomentar la interacción entre el área técnica y el área artística en el ámbito de la música.

En trabajos futuros es posible utilizar el aporte de este trabajo en aplicaciones multimedia interactivas en tiempo real, como juegos de videos o experiencias que demanden diseño de sonido dinámico; además, la construcción formal de herramientas que aporten al proceso creativo de la composición musical.

7. Referencias

- [1] R. Lopez de Mantaras, "Making Music with AI: Some examples," *Proc. 2006 Conf. Rob Milne A Tribut. to a Pioneer. AI Sci. Entrep. Mt.*, pp. 90–100, 2006.
- [2] J. Fernández and F. Vico, "AI methods in algorithmic composition: a comprehensive survey," *J. Artif. Intell. Res.* 48, vol. 48, pp. 513–582, 2013.
- [3] Willi Apel, *Harvard Dictionary of Music*, 2nd ed. Cambridge, Massachusetts: The Belknap Press of Harvard University Press, 1974.
- [4] R. Lilienfeld, *An Introduction to Music*, vol. 48. 1962, p. 131.
- [5] C. Alexander, *The Timeless Way of Building*. Oxford University Press, 1979, p. 552.
- [6] T. Holmes, *Electronic and Experimental Music: Pioneers in Technology and Composition*, 3rd ed. London and New York: Routledge, 2008.
- [7] G. Nierhaus, *Algorithmic Composition: Paradigms of Automatic Music Generation*. Graz: Springer-Verlag/Wien, 2009.
- [8] S. . Kendal and M. Creen, *An Introduction to Knowledge Engineering*. Springer-Verlag New York, 2007.
- [9] Wikipedia, "Tocata and Fugue in D minor, BWV 565." 2015. Disponible en: http://en.wikipedia.org/wiki/Tocata_and_Fugue_in_D_minor,_BWV_565.
- [10] J. Schell, *The Art of Game Design: A Book of Lenses*, vol. 54. Morgan Kaufmann, 2008, p. 489.