



Vol. 3 N.º 9 / Guayaquil
mayo 2024
ISSN 2953-657x

Estudio comparativo de enfoques probabilísticos para la generación de progresiones armónicas

Juan Camilo Núñez Jaramillo
Estudiante en Facultad de Ingenierías,
Universidad San Buenaventura
Medellín, Colombia
juan.nunez191@tau.usbmed.edu.co

Cristian Bajaña Díaz
Estudiante en Facultad de Ingenierías,
Universidad San Buenaventura
Medellín, Colombia
cristian.bajana191@tau.usbmed.edu.co

Mateo Yepes Díaz
Docente en Facultad de Ingenierías,
Universidad San Buenaventura
Medellín, Colombia
mateo.yepes@usbmed.edu.co

Daniel Gómez Marín
Music Technology Group, Universitat Pompeu Fabra, Barcelona, España
Facultad de Ingeniería, Universidad Icesi
Cali, Colombia
daniel.gomez@upf.edu
dgomez@icesi.edu.co

Resumen

En este artículo se presenta la comparativa de dos métodos para la armonización de melodías a través de modelos generativos de progresiones de acordes. La motivación de esta investigación surge de los avances logrados en algoritmos anteriores, específicamente los basados en modelos ocultos de Markov (HMM, por sus siglas en inglés), que han demostrado éxito en este campo y que han obtenido resultados prometedores al combinarse con modelos estocásticos de duración variable, o modelos bayesianos basados en funciones sintácticas tonales.

Recientemente, se han realizado avances en algoritmos para extraer datos cuantitativos de la estructura tonal en progresiones armónicas mediante espacios de tensión tonal y sus perfiles correspondientes. En este artículo se emplean los perfiles de tensión tonal para entrenar un HMM, independiente de las características melódicas, y, además, se entrena un segundo modelo utilizando las características tonales de la melodía. La validación de estos métodos se realiza a través de métricas objetivas y subjetivas, basadas en la comprensión del dominio musical.

El objetivo de este estudio es ampliar la exploración en el campo del aprendizaje automático musical, introduciendo nuevos métodos que permitan a músicos y productores interactuar con herramientas digitales y crear de manera más diversa composiciones musicales.

Palabras clave: Melodía, progresión de acordes, perfiles de tensión tonal, modelo oculto de markov

Abstract

In this paper, the comparison of two methods for harmonizing melodies through generative models of chord progressions is presented. The motivation for this research originates from the advances achieved in previous algorithms, specifically those based on Hidden Markov Models (HMM), which have demonstrated success in this field and have shown promising results when combined with stochastic models of variable duration or Bayesian models based on tonal syntactic functions.

Recently, there have been advances in algorithms for extracting quantitative data from the tonal structure in harmonic progressions through tonal tension spaces and their corresponding profiles. In this article, tonal tension profiles are used to train an HMM (Hidden Markov Model), independent of melodic features, and, in addition, a second model is trained using tonal characteristics of the melody at specific time points. The validation of these methods is carried out through objective and subjective metrics, based on the understanding of the musical domain.

The aim of this study is to broaden exploration in the field of musical machine learning, introducing new methods that enable musicians and producers to interact with digital tools and create musical compositions in a more diverse manner.

Keywords: Melody, chord progression, tonal tension profiles, hidden markov model

1. Introducción

Tradicionalmente, los procesos de aprendizaje y producción musical suelen ser lentos y requieren una gran inversión de recursos tanto para novatos como para expertos en música. Por ello, con el aprendizaje de máquina y sus aplicaciones en procesos de creación¹ se han desarrollado herramientas para la creación musical asistida.

La investigación en esta área ha permitido el desarrollo de diferentes sistemas de generación de elementos musicales, entre los que se encuentra la generación automática de acompañamientos melódicos y armónicos, mezcla y masterización. Estos sistemas son entrenados a partir de ejemplos musicales que los sistemas procesan, aprendiendo de ellos y conformando modelos estadísticos que son utilizados en diversos contextos y aplicaciones. Por ejemplo, el proyecto Flow Machines ha derivado en una herramienta de composición asistida denominada Flow Composer, con la cual se han asistido composiciones musicales y se han compuesto de álbumes como *Hello World*, por SKYGG.² También, investigaciones en ritmo han derivado en sistemas interactivos para el análisis y la generación de patrones de batería.³ Así mismo, técnicas derivadas del aprendizaje profundo o *deep learning*, como LSTM o *transformers*, han sido utilizadas en la creación sistemas generativos

de melodías musicales⁴, o incluso piezas completas.⁵

Las redes neuronales también han sido utilizadas para la armonización automática. Estas difieren de los modelos estadísticos clásicos en su capacidad de realizar predicciones más precisas, pero tienen la desventaja de que necesitan muchos más datos para lograr un buen rendimiento.

Un ejemplo de ello es el modelo bidireccional de redes de memoria a corto y largo plazo (LSTM por sus siglas en inglés) implementado por Lim, Rhyu y Lee⁶. Este modelo demostró un desempeño superior a un modelo oculto de Markov (HMM por sus siglas en inglés) que generaba progresiones armónicas a través de las notas de una melodía. Este desempeño se midió a partir de una métrica objetiva, que evaluaba la capacidad de predecir progresiones iguales al corpus de entrenamiento y una evaluación subjetiva con sujetos de prueba, que evaluaron la concordancia entre una melodía de una canción y la armonía generada de estos dos modelos.

Aun así, según Tsushima, Nakamura y Yoshii⁷, las redes neuronales tienen dificultades para representar la composición musical de manera similar a un humano debido a la dificultad de análisis e interpretabilidad de la estructura interna de las redes neuronales. En su investigación, se proponen métodos estadísticos de armonización que incorporan funciones sintácticas, estructuras de árboles y estructuras métricas de los acordes para generar progresiones.

Los HMM se han utilizado en áreas como el reconocimiento de habla,⁸ en fi-

1 Eduardo Reck Miranda, «Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity», en *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*, edición de Eduardo Reck Miranda (Springer International Publishing, 2021), <https://doi.org/10.1007/978-3-030-72116-9>

2 Shulei Ji, Jing Luo y Xinyu Yang, «A Comprehensive Survey on Deep Music Generation: Multi-level Representations, Algorithms, Evaluations, and Future Directions», *ArXiv* (2020), <http://arxiv.org/abs/2011.06801>

3 François Pachet, Pierre Roy y Benoit Carré, «Assisted Music Creation with Flow Machines: Towards New Categories of New», *ArXiv* (2020), <http://arxiv.org/abs/2006.09232>

4 Daniel Gómez-Marín, Sergi Jordà y Perfecto Herrera, «Drum Rhythm Spaces: From Polyphonic Similarity to Generative Maps», *J New Music Res* vol. 49, n.º 5 (2020): 438-456.

4 Kejun Zhang et al., «WuYun: Exploring Hierarchical Skeleton-guided Melody Generation Using Knowledge-enhanced Deep Learning», *ArXiv* (2023).

5 Andrea Agostinelli et al., «MusicLM: Generating Music From Text», *ArXiv* (2023), <http://arxiv.org/abs/2301.11325>

6 Hyungui Lim, Seungyeon Rhyu y Kyogu Lee, «Chord Generation From Symbolic Melody Using BLSTM Networks», *ArXiv* (2017).

7 Hiroaki Tsushima, Eita Nakamura y Kazuyoshi Yoshii, «Bayesian Melody Harmonization Based on a Tree-Structured Generative Model of Chord Sequences and Melodies», *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2020).

8 Lawrence R. Rabiner, «A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition», *Proceedings on the IEEE*, vol. 77, n.º 2 (1989).

nanzas y economía, para la predicción del precio de las acciones⁹ y también en el reconocimiento de patrones, que permiten identificar a diferentes hablantes basado en el timbre de sus voces.¹⁰ La generación automática de progresiones de acordes también ha sido objetivo de investigación para los HMM, utilizando métodos estadísticos para generar secuencias armónicas. Los HMM modelan procesos de Markov con dos variables de estado: los estados ocultos y los estados observables. Con estas variables un HMM utiliza la probabilidad para determinar la transición de un estado oculto hacia otro a través de un estado observable.¹¹

Aplicados al campo de la generación de acordes musicales, los HMM utilizan las notas de una melodía como estados observables y los acordes que las acompañan como estados ocultos, logrando generar una secuencia de acordes para una melodía introducida al modelo. Basados en esta metodología, se han desarrollado aplicaciones como MySong, la cual permite la armonización de una melodía vocal reconocida a través de detección de tono.¹² Otros HMM más complejos han sido diseñados utilizando modelos de grafos, a través de los cuales es posible capturar la estructura de los acordes y la interacción de estos basados en una melodía.¹³

Apesar de haber sido usados con éxito, una limitación de los HMM es que asumen que la probabilidad de cada estado oculto solo depende del estado anterior. Esto

restringe la capacidad de modelamiento de las distribuciones de probabilidad de los estados ocultos. En la generación musical esto hace que se asuma que un acorde solo está relacionado con una nota en una pieza musical sin tener en cuenta otras notas que puedan acompañar a ese acorde. Para abordar este desafío, se han implementado modelos semiocultos de Markov, los cuales tienen en cuenta las duraciones variables de los acordes. Estos modelos permiten modelar las características temporales de las secuencias entre notas y acordes.¹⁴

Los HMM mencionados anteriormente generan progresiones tomando variables musicales (por ejemplo, la armonía o la melodía de una pieza musical) como estados ocultos y como observaciones variables la secuencia de acordes, la melodía y/o la duración de los acordes. Aunque las dos últimas son las más relevantes para entrenar dichos modelos,¹⁵ es posible buscar otras variables que sirvan como estados para realizar armonización automática. Un ejemplo de esto¹⁶ combina modelos probabilísticos de diversas variables musicales utilizando métodos de interpolación logarítmica-lineal, a través de submodelos que representan las progresiones de acordes, la relación tonal y la relación melódica. También, Tsushima et al.¹⁷ utilizaron una combinación de diferentes modelos de Markov y gramática libre de contexto probabilístico (PCFG por sus siglas en inglés) para describir el ritmo, funciones armónicas y la estructura jerárquica de los acordes para generar progresiones. En esta investiga-

9 Xin An, «Prediction of Stock Price by Hidden Markov Model», *BCP Business & Management*, 38 (2023): 309–316.

10 Faisal Farhan, Ziaul Hoque y Provat Raihan, «Study of Pattern recognition system with Hidden Markov Model through the development of a raspberry pi based speaker recogniser» (tesis de grado, Ahsanullah University of Science & Technology, 2020), <https://www.researchgate.net/publication/348150123>

11 Christopher M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2006).

12 Ian Simon, Dan Morris y Sumit Basu, «MySong: Automatic Accompaniment Generation for Vocal Melodies», *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* Association for Computing Machinery (2008): 725–734.

13 Jean-François Paiement, Douglas Eck y Samy Bengio, «Probabilistic Melodic Harmonization», en *Advances in Artificial Intelligence. Canadian AI 2006. Lecture Notes in Computer Science*, edición de Lamontagne, L., Marchand, M., 218–229 (Berlín: Springer, 2006).

14 Ryan Groves, «Automatic Harmonization Using a Hidden Semi-Markov Model», *AIIDE Workshop Technical Report WS-13-22 (Musical Metacreation)*, vol. 9, n.º 5 (2013): 48–54.

15 Dan Morris, Ian Simon y Sumit Basu, «Exposing Parameters of a Trained Dynamic Model for Interactive Music Creation», *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*. (2008).

16 Stanisław Raczyński, Satoru Fukayama, Emmanuel Vincent, «Melody Harmonisation with Interpolated Probabilistic Models» (reporte de investigación, RR-8110, INRIA, 2012), <https://hal.inria.fr/hal-00742957>

17 Hiroaki Tsushima, Eita Nakamura, Katsutoshi Itoya-ma, Kazuyoshi Yoshii, «Function-And Rhythm-Aware Melody Harmonization Based On Tree-Structured Parsing And Split-Merge Sampling Of Chord Sequences» (18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017).

ción, se propone otra alternativa diferente basada en espacios tonales, en particular los perfiles de tensión tonal. Estos perfiles serán utilizados como los estados observables de un modelo oculto de Markov, es decir que este HMM genera progresiones armónicas basado en la información proporcionada por estos perfiles.

Los espacios tonales se han estudiado para describir de manera matemática y geométrica características que influyen en la armonía. Una de ellas es la tensión tonal, introducida primeramente por Lerdahl.¹⁸ A partir de esta introducción, se deriva la investigación de Randall y Khan¹⁹ que utilizaron esta teoría para formular una comparación con tres modelos empíricos para así comprobar si dicho modelo teórico se aproxima a los datos que modelan. La teoría desarrollada por Lerdahl también ha servido como base para implementar otros modelos, como el Espacio de Intervalo Tonal (TIS por sus siglas en inglés).²⁰ Este último modelo permite relacionar cada acorde en una progresión armónica específica con un valor numérico que describe un perfil de tensión tonal (PTT), permitiendo crear curvas de tensión tonal con dichos perfiles.

Los avances de Navarro y compañía en los TIS y PTT permiten cuantificar de manera sistemática y numérica aspectos relevantes de una pieza musical. Por un lado, los TIS dan la capacidad de encontrar la semejanza entre diferentes acordes y caracterizarlos, por medio de su respectiva posición dentro del espacio de intervalo tonal. Por el otro, los PTT ofrecen la facilidad matemática de cuantificar de una manera eficiente la tensión entre los acordes que

18 Fred Lerdahl, «Tonal Pitch Space», *Music Perception: An Interdisciplinary Journal*, vol. 5, n.º 3 (1988): 315-350, <https://www.researchgate.net/publication/257925384>

19 Richard R. Randall y Bilal Khan, «Lerdahl's Tonal Pitch Space Model and Associated Metric Spaces», *Journal of Mathematics and Music*, vol. 4, n.º 3 (2010): 121-131, <https://doi.org/10.1080/17459737.2010.529654>

20 Marcelo Caetano, María Navarro-Cáceres y Gilberto Bernardes, «Objective Evaluation of Tonal Fitness for Chord Progressions Using the Tonal Interval Space», *Artificial Intelligence in Music, Sound, Art and Design* (2020): 150-164.

componen una progresión alineada con la percepción del oyente.²¹

En la investigación presentada en este artículo se busca implementar un HMM a través de perfiles de tensión tonal y comparar dicho método con un HMM que crea progresiones a través de las notas de una melodía por medio de estudios objetivos y subjetivos que informen de su desempeño. Así se tendrá información de las ventajas y desventajas de ambos métodos a la hora de realizar la misma tarea. Esto con el fin de sentar las bases para un sistema de asistencia para la composición musical.

Para el entrenamiento de los modelos se hace uso de una base de datos de música electrónica EDM. Este género musical se elige por su gran popularidad en las últimas décadas y su influencia en personas de edades variadas, dado que el mismo ha sido tendencia desde los años 90. El estilo es denominado Electropop Fusion, y contiene diversas canciones de artistas que han estado en *rankings* de más escuchados de la última década.

El artículo está organizado de la siguiente manera: en la sección II se explica el preprocesamiento y entrenamiento de los modelos; en la sección III se explican los métodos de evaluación subjetiva y objetiva y se discuten los resultados obtenidos; luego, en la sección IV, se dan las conclusiones y el trabajo a futuro.

2. Metodología

Los algoritmos de los HMM diseñados en este trabajo buscan generar progresiones de acordes por medio de dos estados observables diferentes y un mismo estado oculto para los dos modelos. Uno de los modelos toma como estado observable los perfiles de tensión tonal (PTT) que describen las curvas de tensión encontradas en el corpus

21 María Navarro-Cáceres et al., «A Computational Model of Tonal Tension Profile of Chord Progressions in the Tonal Interval Space», *Entropy*, vol. 22, n.º 11 (2020): 1-30, <https://doi.org/10.3390/e2211291>

de la base de datos. Este modelo se denomina como PTT. El segundo modelo propone estados observables basados en las notas de la melodía de las canciones presentes en el corpus, y se le da el nombre de Melodía. Este modelo tomará una secuencia de notas y encontrará la secuencia de acordes más probable según cada nota.

La metodología propuesta para desarrollar este algoritmo se puede dividir en tres partes. Primero, se realiza un preprocesamiento de la base de datos para extraer información armónica y melódica (progresiones de acordes, notas, tonalidad y duraciones de nota y acordes). Segundo, se generan las matrices de transición, que representan la probabilidad de transición de un estado oculto a otro. Las matrices de emisión, que representan la probabilidad de tener un estado oculto dado un estado observable. Y de estado inicial, que representan la probabilidad inicial de un estado oculto. La figura 1 representa el diagrama de flujo para el entrenamiento de los dos modelos anteriormente mencionados. Finalmente, se procesan los datos del corpus musical con el propósito de entrenar cada modelo para generar progresiones armónicas y luego pasar a su debida evaluación subjetiva y objetiva.

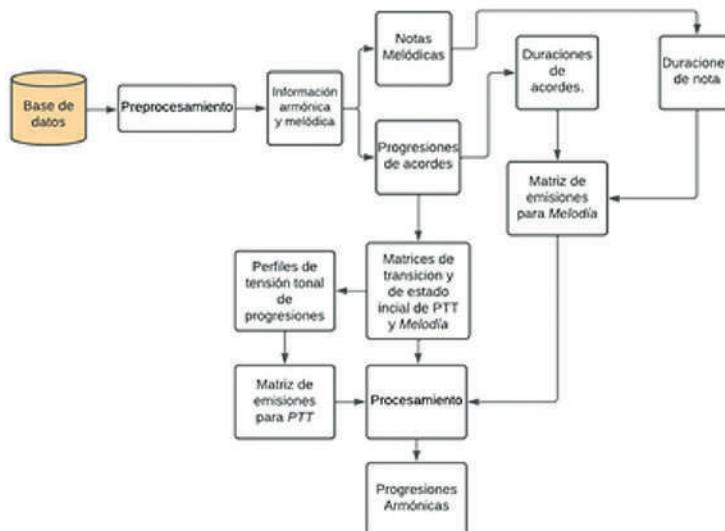


Figura 1: diagrama de flujo para el entrenamiento de los modelos.

2.1. Preprocesamiento

El corpus utilizado está compuesto por 83 fragmentos de cuatro compases extraídos de 25 canciones publicadas por artistas del género musical denominado EDM (entre ellos se encuentran artistas como Avicii, Calvin Harris, David Guetta, Dua Lipa, Marshmello). Estos fragmentos son elegidos porque cuentan con información tanto melódica como armónica en pistas MIDI independientes y por su reconocimiento en la última década. Los archivos MIDI fueron descargados de una base de datos pública disponible en línea.²²

De cada fragmento se extraen las siguientes características simbólicas musicales que representan la información extraída de armonía y melodía:

Para armonía son:

1. Progresión de acordes: representa la secuencia de acordes encontrada en cada fragmento.
2. Duración de acorde: indica el tiempo de duración de un acorde por medio de semicorcheas.
3. Perfil de tensión tonal: representan los cambios de tensión entre una progresión de acordes.
4. Tonalidad: indica la tónica o nota principal de cada fragmento del corpus musical.
5. BPM: indica la cantidad de pulsaciones por minuto de cada fragmento del corpus musical

Para la melodía son:

1. Tiempo inicial: indica el inicio de una nota melódica en un fragmento en semicorcheas.

22 Unison, <https://unison.audio/free-packs/>

2. Duración de cada nota: es la duración total de cada nota de un fragmento y está dada en semicorcheas.
3. MIDI pitch: indica el valor MIDI de cada nota de forma numérica a través de un número entero del 0 al 127.
4. Conjunto de tono de clase: es una representación de cada nota por medio de números enteros del 0 al 11 (cada uno representa una nota musical desde C a B).

Cinco algoritmos fueron utilizados para extraer la información armónica y melódica. Primero, para extraer la información de las progresiones de acordes de las canciones del corpus se implementó un algoritmo para reconocer la progresión de cada fragmento basado en la teoría propuesta por Pardo y Birmingham²³ y complementada por Scholz y Ramalho²⁴. El algoritmo divide cada fragmento en puntos de partición que marcan el inicio o final de una nota. Posteriormente, compara cada segmento con todas las posibles combinaciones de notas fundamentales y formas de acordes para identificar los acordes, las formas que compara el algoritmo son presentadas en la tabla I. Por último, utiliza un Grafo Acíclico Dirigido (DAG) para seleccionar los puntos de inicio y fin de los acordes. Se elige dicho algoritmo porque en una investigación actual que se encuentra en proceso este algoritmo obtuvo una eficiencia de reconocimiento del 79.77 %.

Tabla I
Forma y plantilla de acorde

Forma de acorde	Plantilla
M	[0, 4, 7, 0]
m	[0, 3, 7, 0]
o	[0, 3, 6, 0]
Aug	[0, 4, 8, 0]
Aug 7	[0, 4, 8, 10]
Aug Maj7	[0, 4, 8, 11]
7	[0, 4, 7, 10]
m7	[0, 3, 7, 10]
Maj 7	[0, 4, 7, 11]
m Maj7	[0, 3, 7, 11]
m7b5	[0, 3, 6, 10]
o7	[0, 3, 6, 9]
sus4	[0, 5, 7, 0]

Para extraer la información de la tonalidad de cada canción se utilizó un algoritmo basado en la investigación hecha por Madsen y Widmer²⁵. Se hace una normalización hacia C mayor o menor según corresponda el modo en el que se encuentra la canción, esto con el fin de normalizar todas las progresiones obtenidas y también las notas melódicas. Seguido a esto se le asigna a cada acorde su grado correspondiente en la tonalidad, por ejemplo, el acorde C M se clasifica como un primer grado I M.

Las progresiones normalizadas son utilizadas para obtener los perfiles de tensión tonal de los acordes. Los PTT representan los cambios de tensión que un oyente experimenta en una progresión de acordes, y se obtienen a través del cálculo de la tensión tonal instantánea en un intervalo de espacio tonal (TIS por sus siglas en inglés) para cada acorde, esta representa una combinación lineal de indicadores tonales proporcionados por Caetano, Navarro-Cáceres y Bernardes²⁶ descritos por la ecuación (1):

$$M(T_i, P) = \omega_1 \delta(T_i, P) + \omega_2 c(T_i) + \omega_3 M(T_i, P) + \omega_4 h(T_i, P) \quad (1)$$

²³ Bryan Pardo y William P. Birmingham, «Algorithms for Chordal Analysis», *Computer Music Journal*, vol. 26, n.º 2 (2002).

²⁴ Geber Ramalho y Ricardo Scholz, «COCHONUT: Recognizing Complex Chords from MIDI Guitar Sequences» (Conference: International Conferences on Music Information Retrieval, 2008), <https://www.researchgate.net/publication/269038716>

²⁵ Søren Tjagvad Madsen y Gerhard Widmer, «Key-Finding With Interval Profiles» (Proceedings of the International Computer Music Conference ICMC, Copenhague, Dinamarca, 2007).

²⁶ Caetano, Navarro-Cáceres y Bernardes, «Objective Evaluation of Tonal...».

Donde T_i representa el acorde i -ésimo de una progresión P , δ representa la distancia tonal, c disonancia, m conducción de voces, h tensión jerárquica y ω_j representa el peso de cada indicador tonal.

δ Representa una combinación lineal de tres descriptores de distancias tonales obtenidos a partir de un TIS que toma vectores croma de doce dimensiones y los convierte en un vector de intervalos tonales (TIV por sus siglas en inglés). Cada TIV puede representar un acorde, una tonalidad (C mayor o menor) y las tres funciones armónicas categóricas (tónica, subdominante y dominante). Con estos se calculan los indicadores de tensión de distancia entre acordes, distancia de cada acorde hacia la tonalidad y distancia de cada acorde hacia las funciones tonales.

Cada indicador de tensión tonal representa un valor que se le asignará a un acorde en una progresión dada y de esta forma se obtienen las curvas de tensión necesarias para entrenar el modelo PTT.

Para extraer la información de duraciones de los acordes reconocidos y de las notas melódicas de cada canción se hace uso de la librería de Python Mido²⁷, donde se consigue información del BPM de las canciones por medio de los archivos MIDI. Para conocer la duración de cada acorde, se hace uso de los puntos de partición y el DAG obtenidos con el algoritmo de reconocimiento de acordes de Pardo y Birmingham²⁸ porque estos establecen el inicio y el final de cada acorde aproximado a la semicorchea más cercana.

Para la obtención de la duración de cada nota melódica, es necesario realizar la consecución del momento de inicio y final de esta. Sigue un procedimiento consistente para ambos datos: se selecciona la pista correspondiente a la melodía y, posterior-

mente, se recorre dicha pista en búsqueda de los mensajes Note-On y Note-Off del archivo MIDI. Estos mensajes representan los puntos temporales en los que inicia y termina una nota musical en el archivo. La librería Mido proporciona el tiempo inicial de cada nota con relación al último mensaje MIDI encontrado, por lo tanto, para calcular el tiempo inicial acumulativo de cada nota, se realiza una suma de los valores de tiempo de los mensajes Note-On y Note-Off encontrados por separado. Es importante tener en cuenta que los tiempos obtenidos inicialmente están dados en ticks MIDI, por lo que se debe realizar la respectiva conversión a semicorcheas. Esto se logra dividiendo el número de ticks por el equivalente en ticks a un dieciseisavo. Dada la elección de la unidad temporal elegida en la investigación, con los datos de inicio y final obtenidos, se calcula la diferencia entre el tiempo final y el tiempo inicial para determinar la duración de cada nota melódica.

Con la librería anteriormente descrita también se obtiene el descriptor de MIDI *pitch* de cada nota, el cual se pasa al descriptor de conjunto de tono de clase (PCS, por sus siglas en inglés) que representa la nota de la escala musical mediante un número de 0 a 11 donde 0 representa la nota C y 11 la nota B.

2.2. Modelos Ocultos de Markov

Diferentes autores²⁹ definen los HMM a través de las siguientes asunciones:

1. Una cadena de Markov finita de tiempo discreto X , que representa los estados ocultos.
2. Un proceso observable Y cuyas transiciones dependen de los estados X .

27 Mido, <https://mido.readthedocs.io/en/stable/>
 28 Pardo y Birmingham, «Algorithms for Chordal Analysis».

29 Jonh van der Hoek y Robert J. Elliot, *Introduction to Hidden Semi-Markov Models* (Cambridge University Press, 2018).
 Lawrence R. Rabiner, «A Tutorial on Hidden...».

3. Una distribución de probabilidad de estados de transición, que representa la matriz de transición, $A = \{a_{ij}\}$ donde:

$$a_{ij} = P(X_{n+1} = e_j | X_n = e_i) \quad (2)$$

4. Una distribución de probabilidad inicial de estados que representa la matriz de estados iniciales $\pi = \{\pi_i\}$ donde:

$$\pi_i = P(X_0 = e_i) \quad (3)$$

5. Una distribución de probabilidad de observaciones Y_n en un estado X_n que representa la matriz de emisiones C_{ki}

$$C_{ki} = P(Y_n = f_k | X_n = e_i) \quad (4)$$

Con los descriptores de armonía y melodía ya obtenidos, se procede a generar las matrices de estado inicial, de transición y de emisión para los modelos a entrenar, el primer modelo es un HMM que utiliza como estado observable las curvas de tensión tonal generadas por los PTT de todas las progresiones de acordes existentes en la base de datos y el segundo es un modelo oculto convencional que utiliza como estado oculto las notas de una melodía.

Las curvas de PTT se obtienen a partir de cada progresión presente en el corpus y se toman como observaciones para el modelo PTT; el esquema de la figura 2 muestra la forma gráfica del HMM basado en PTT que es agnóstico a la melodía. Por otro lado, en el modelo Melodía las observaciones se obtienen a partir de las notas melódicas que se asocian con un acorde en un fragmento musical; el esquema de la figura 3 muestra la forma gráfica del HMM basado en Melodía.

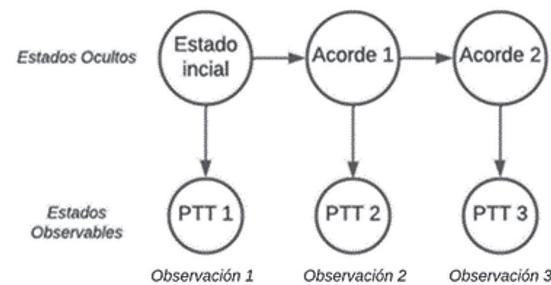


Figura 2: representación gráfica HMM a partir de perfiles de tensión tonal.

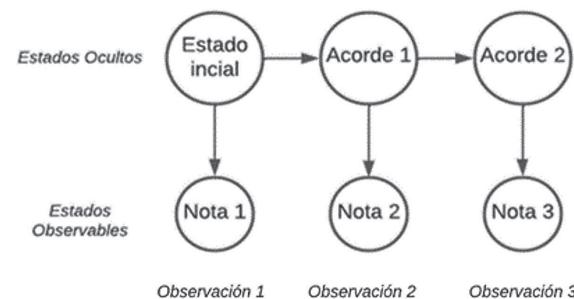


Figura 3: representación gráfica HMM a partir de notas de una melodía.

2.3. Procesamiento de datos del corpus musical

Una vez definidos los dos modelos ocultos de Markov, se procesa la información musical de la base de datos para obtener cada matriz descrita en la sección anterior.

Las progresiones obtenidas primero se dividen en tonalidad mayor y menor, ya que cada una cuenta con grados diferentes que pueden aparecer en una progresión; estos son representados por la tabla II. Luego de estar organizadas según su tonalidad se procede a obtener la matriz de estado inicial, con los grados iniciales en una progresión. De esta forma se obtiene la probabilidad de que una progresión inicie con un determinado grado inicial.

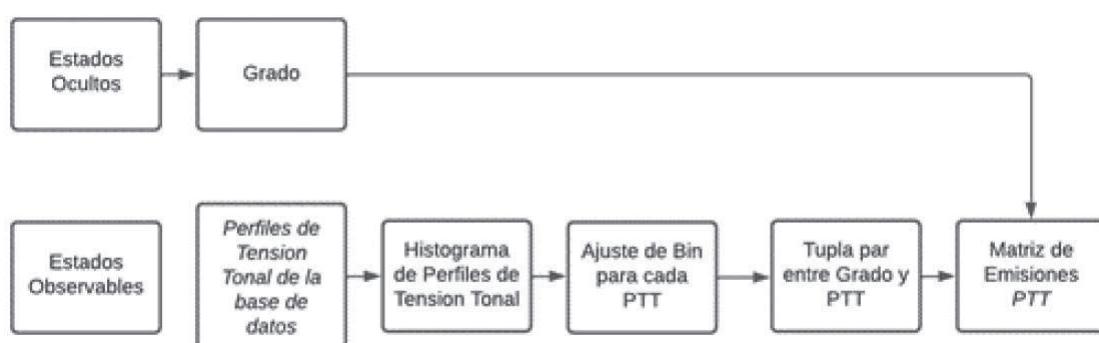
Tabla II
Grados según tonalidad

Tonalidad Mayor	Tonalidad Menor
I	I
I#/IIb	I#/IIb
II	II
II#/IIIb	III
III	III#/IVb
IV	IV
IV#/Vb	IV#/Vb
V	V
V#/VIb	VI
VI	VI#/VIIb
VI#/VIIb	VII
VII	VII#

Para generar la matriz de transición de acordes que representa la probabilidad de transición de un grado X_{i+1} dado otro grado X_i se toman todos los grados existentes en una progresión emparejándose de a dos, generando una tupla par que contiene un grado y su sucesor. Esto permite generar una cadena de Markov de orden uno y representar los estados ocultos X . Para determinar la probabilidad de transición se analiza un estado en el espacio, buscando en las tuplas que tengan en primera posición dicho estado y se realiza un conteo de

tonalidad de C menor). Dicha agrupación se realiza de manera empírica y corresponde a un ajuste del ancho de cada *bin* perteneciente al histograma generado. El objetivo de este ajuste es lograr que la cantidad de *bins* en la distribución sea proporcional a la cantidad de grados existentes en el corpus. De esta manera, se reduce la cantidad de estados ocultos asociados a un grado específico, lo que las hace más significativas para el modelo.

Seguido a la normalización, para generar la matriz de emisiones se analizan los posibles PTT generados por un grado en particular, generando ahora una tupla par de grado y perfil de tensión tonal. Como se indicó anteriormente un grado puede tener diferentes perfiles de tensión tonal por ello la probabilidad de cada observación está dada por $P(y_m|x_m)$ esta es obtenida contando la cantidad de veces que un grado x_m se asocia con un PTT y_m y luego se realiza una normalización sumando el total de perfiles asociados a un estado X . La figura 4 representa el diagrama de flujo de este proceso.



cuántas veces ese acorde pasa a otro en toda la base de datos, luego de dicho conteo se normaliza sumando la cantidad de posibles cambios para ese estado.

Antes de realizar la matriz de emisiones para el primer modelo, es relevante realizar una agrupación en intervalos de datos (también llamados *bin*), basados en los valores de PTT encontrados en el corpus (166 para tonalidad de C mayor y de 172 para

Figura 4: diagrama de flujo de obtención de la matriz de emisiones de PTT.

En contraste con las observaciones del modelo PTT, las de Melodía se derivan directamente de la melodía a la que se desea aplicar la armonización correspondiente. Es importante asegurar que la cantidad de acordes generados sea adecuada, por esto, tal como en el primer modelo, la longi-

tud del conjunto de observaciones debe ser igual a la longitud del conjunto de duraciones de acordes. Es común que una melodía contenga considerablemente más notas que acordes que la acompañen. Por lo tanto, al definir el conjunto de observaciones de notas, es necesario recortar las notas reconocidas originalmente del archivo MIDI ingresado. Para realizar este proceso se utilizan los descriptores de melodía mencionados en la etapa de preprocesamiento. Usando los puntos temporales de inicio y duración de las notas de la melodía, se establece una correspondencia temporal con la cantidad de duraciones de los acordes y los valores respectivos de esas duraciones. De esta manera, cada observación de la melodía se asocia con la primera nota reconocida que se encuentre dentro de la duración y el tiempo inicial de las duraciones de los acordes.

Con dichas notas se realiza un proceso similar al primer modelo para generar la matriz de emisiones del segundo modelo, como se observa en la figura 5; ahora la tupla par simboliza un grado y la nota con la que tiene correspondencia. Por ende, la probabilidad de cada observación se obtiene contando la cantidad de veces que un grado x_m se asocia con una nota y_m . Para encontrar la secuencia de estados más probable dado una cadena de observaciones de cada modelo, se utiliza el algoritmo de Viterbi.³⁰

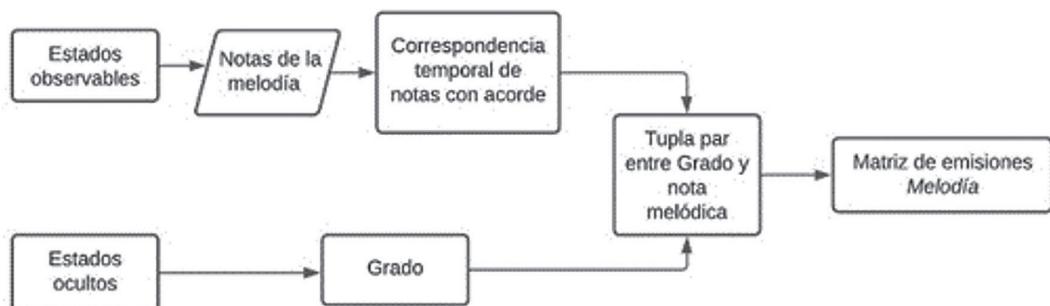


Figura 5: diagrama de flujo de obtención de la matriz de emisiones de Melodía.

Finalmente, para evaluar la eficiencia de los modelos creados se divide el corpus y se reentrenan los modelos, quedando así un 75 % de datos para el entrenamiento y un 25 % restante para la evaluación. Esta división es representada por el diagrama de flujo de la figura 6.

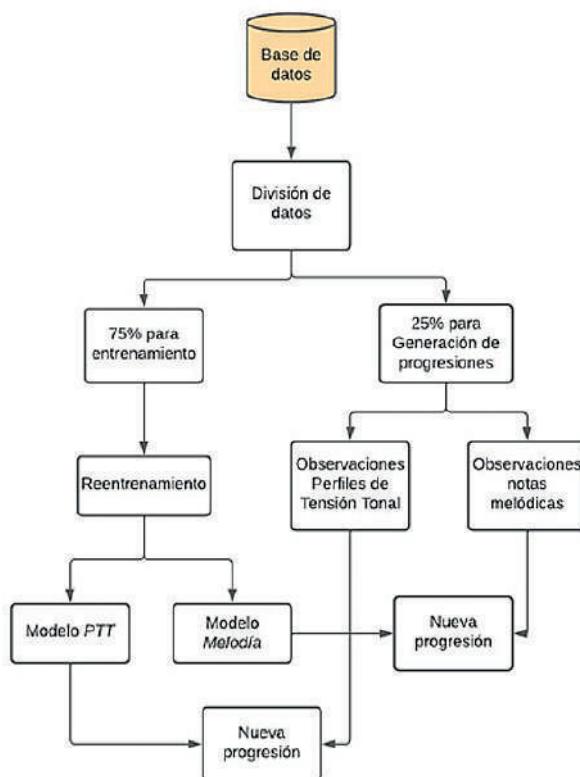


Figura 6: diagrama de flujo reentrenamiento de modelo PTT y Melodía.

30 G. David Forney, «The Viterbi Algorithm», *Formal Structure of Electromagnetics*. Amsterdam vol. 61, n.o 3 (1973).

3. Resultados

Para la evaluación de la eficiencia de los modelos PTT y Melodía se realizan tres evaluaciones objetivas y una evaluación subjetiva con un grupo de personas. Para estas pruebas se toma aleatoriamente una progresión que pertenezca al 25 % de datos que sobran luego del reentrenamiento de los modelos que se indicó en la sección anterior. Seguido a esto se genera una progresión nueva con PTT, tomando la curva de tensión de la progresión elegida. Luego se crea una progresión nueva con Melodía, tomando las notas que acompañan la progresión elegida. Y por último se crea una progresión aleatoria con acordes pertenecientes al estilo musical, es decir que solo se toman en cuenta acordes que se reconocieron en el corpus musical. Este proceso se hace hasta obtener 21 progresiones para cada uno de los grupos que serán denominados como «Original» (grupo de verdad), «Modelo PTT», «Modelo Melodía» y «Aleatorio».

3.1. Experimento objetivo 1: evaluación de función de densidad de probabilidad

Con este primer experimento se busca comparar las generaciones obtenidas con las progresiones originales de la base de datos. Esto a través de una Estimación de Densidad de Kernel (EDK)³¹ y la Divergencia de Kullback-Leibler (DKL)³² y el Área de Superposición (AS). Una EDK permite estimar la función de densidad de probabilidad (FDP) de un conjunto de datos, de esta forma se obtiene una representación numérica de cómo se distribuyen los acordes de las

³¹ Murray Rosenblatt, «Remarks On Some Nonparametric Estimates Of A Density Function», *Annals of Mathematical Statistics* (1956).

³² I. Csiszar, « \mathcal{I} -Divergence Geometry Of Probability Distributions And Minimization Problems», *The Annals of Probability*, vol. 3, n.º 1 (1975): 146–158.

progresiones de cada grupo. Esto permite que se pueda comparar la FDP de un grupo respecto a la FDP del grupo «Original» que representa el grupo de verdad a través del cálculo de DKL y AS.

3.1.1. Procedimiento

Se crean entonces nuevas curvas de tensión con las progresiones generadas por los dos modelos y un tercer método denominado «Aleatorio». Para generar el EDK de cada grupo se calcula la distancia de Frechet³³ entre las curvas del grupo de verdad frente a cualquier otro grupo, es decir, se comparan las distancias del grupo «Original» con las del grupo «Modelo PTT».

Siguiendo la metodología de Yang y Lerch³⁴, se genera una EDK para obtener una FDP con cada grupo. De esta forma para hallar la similitud entre una FDP y la FDP del grupo de verdad se realiza la medida de DKL y AS que permite determinar qué tanto diverge cada distribución frente al grupo «Original».

3.1.2. Resultados y discusión

La figura 7 muestra las distribuciones obtenidas de cada grupo y en ella se puede observar qué tan similares son las progresiones generadas con respecto al grupo de verdad, en términos de sus perfiles de tensión tonal. Se puede ver que existe una mayor semejanza para los grupos «Modelo PTT» y «Aleatorio» con la FDP del grupo «Original» que representa el estilo musical del corpus. Algo que detallar en la similitud obtenida con el grupo anterior es que esta

³³ Helmut Alt y Michael Godau, «Computing The Fréchet Distance Between Two Polygonal Curves», *International Journal of Computational Geometry & Applications*, vol. 5, n.º 1 y 2 (1995): 75–91. <https://doi.org/10.1142/s0218195995000064>

³⁴ Li-Chia Yang y Alexander Lerch, «On the Evaluation of Generative Models in Music», *Neural Computing and Applications*, vol. 32, n.º 9 (2020): 4773–4784. <https://doi.org/10.1007/s00521-018-3849-7>

se da porque las progresiones contenidas en él se generan teniendo en cuenta los acordes que representan a toda la base de datos. Por otro lado, el modelo Melodía es el que menor influencia tiene del grupo de verdad en los PTT generados por este.

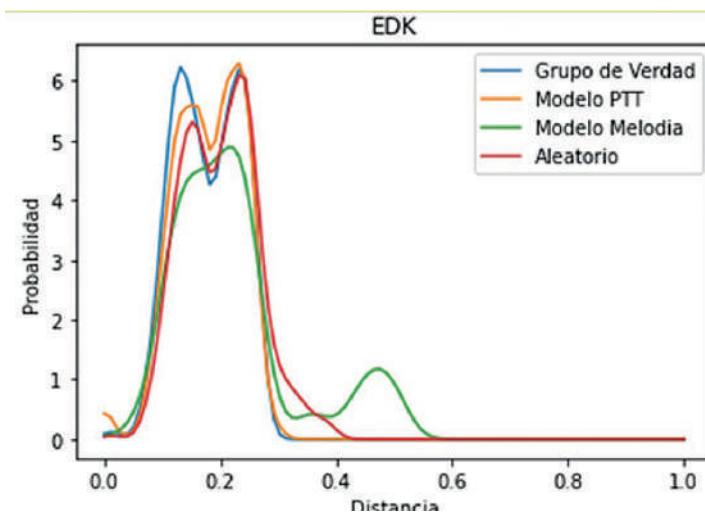


Figura 7: estimación de Densidad de Kernel (EDK) de cada grupo evaluado.

Tabla III
Divergencia de Kullback-Leibler (DLK)
y área de superposición

	DLK	AO
«Original» y «Modelo PTT»	1.578	0.95
«Original» y «Modelo Melodía»	19.698	0.80
«Original» y «Aleatorio»	9.903	0.89

Por otro lado, en Melodía se observa que este presenta la mayor divergencia y una menor área de superposición, indicando así que el modelo no es capaz de recrear de manera adecuada los PTT de las progresiones originales, incluso las progresiones generadas aleatoriamente lo superan. Un factor que afecta este resultado son las observaciones de este modelo porque se asume que un acorde solo puede estar ligado a una nota en una duración determinada, sin tener en cuenta que otras notas pueden acompañarlo.

3.2. Experimento objetivo 2: evaluación de descriptor de transición de acordes

Con el objetivo de evaluar la similitud en las transiciones de acordes generadas por cada grupo experimental en comparación con el grupo de referencia, se utiliza una métrica de evaluación que genera una matriz de transición entre acordes para cada uno de ellos. De esta forma se puede determinar cómo cada uno de los modelos define los patrones armónicos del estilo musical a través de las progresiones evaluadas.

3.2.1. Procedimiento

En primer lugar, se hace un proceso similar al señalado en la sección II donde se describe la obtención de la matriz de transición que representa una cadena de Markov de primer orden. Esto se aplica para esta prueba dividiendo las canciones en sus respectivas tonalidades (mayor y menor), teniendo en cuenta todos los acordes existentes en cada una de estas; estos acordes fueron descritos previamente en la tabla II. Luego se agrupan los acordes en tuplas pares para cada grupo y se analizan las transiciones de acordes existentes en cada progresión y finalmente se normalizan las matrices. Después de obtenerlas, se procede a realizar un cálculo de distancia de cada matriz respecto al grupo de verdad dependiendo la tonalidad, a partir de una norma matricial descrita por la ecuación (5).

$$d_2(A, B) = \sum_{i=1}^n \sum_{j=1}^n (a_{ij} - b_{ij})^2 \quad (5)$$

3.2.2. Resultados y discusión

La tabla IV representa las distancias descriptas por la ecuación 5 obtenidas para las progresiones con tonalidad mayor y menor respectivamente.

Tabla IV
Norma matricial d2

	Mayor	Menor
«Original» y «Modelo PTT»	0.636	0.0
«Original» y «Modelo Melodía»	2.741	3.647
«Original» y «Aleatorio»	3.449	4.573

En los resultados de la tabla IV se observa que el modelo PTT tiene un grado alto de similitud con los patrones armónicos que se encuentran en el estilo musical del grupo «Original». Además de esto, se nota que, para las progresiones evaluadas en tonalidad menor, este modelo genera patrones armónicos idénticos, porque la distancia obtenida es 0, y para la mayor genera patrones muy cercanos. Los resultados del segundo modelo Melodía indican que este tiene una menor capacidad para imitar los patrones armónicos, en concreto los patrones de la tonalidad menor con respecto a la mayor. Por último, las distancias obtenidas por el grupo «Aleatorio» son las más largas; esto es de esperar debido a que sus progresiones contienen patrones armónicos elegidos al azar.

3.3. Experimento objetivo 3: evaluación de similitud

Se establece una métrica de similitud de progresiones a partir de la similitud entre acordes³⁵ para evaluar los acordes de cada

³⁵ C. Wongsaroj, N. Prompoon y A. Surarerks, *Digital Information and Communication Technology and its Applications (DICTAP)*, 2014 Fourth International Conference (2014).

modelo con las progresiones evaluadas y determinar qué tan similar estas están frente al estilo musical elegido.

3.3.1. Procedimiento

La primera métrica es calculada el índice de Jaccard a través de la ecuación (6) por medio de cromas, en este caso entre dos cromas que representan los acordes a comparar. La similitud entre acordes representa una relación entre las notas compartidas entre ellos dividido por las notas de cada uno de ellos, en otras palabras, la intersección de sus tonos de clase frente a la unión de sus tonos de clase.

$$C_sim(C_1, C_2) = \frac{|P(C_1) \cap P(C_2)|}{|P(C_1) \cup P(C_2)|} \quad (6)$$

La segunda métrica deriva de la primera y se encarga de comparar la similitud entre acordes de dos progresiones diferentes representadas por vectores $a[i]$ y $b[j]$ cada uno con una longitud determinada. En este caso las longitudes de acordes se tomarán como iguales, lo que deriva a computar la suma de similitud entre cada acorde $a[j]$ con un acorde $b[j]$ y normalizar por la cantidad de acordes en cada progresión, la ecuación (7) describe este proceso.

$$C_sim(C_1, C_2) = \frac{\sum_{j=1}^n C_sim(a[j + i - 1], b[j])}{m} \quad (7)$$

3.3.2. Resultados y discusión

La figura 8 representa los resultados obtenidos de similitud de acordes de los grupos estudiados frente al grupo «Original» y la tabla V representa los valores obtenidos de promedio y desviación estándar.

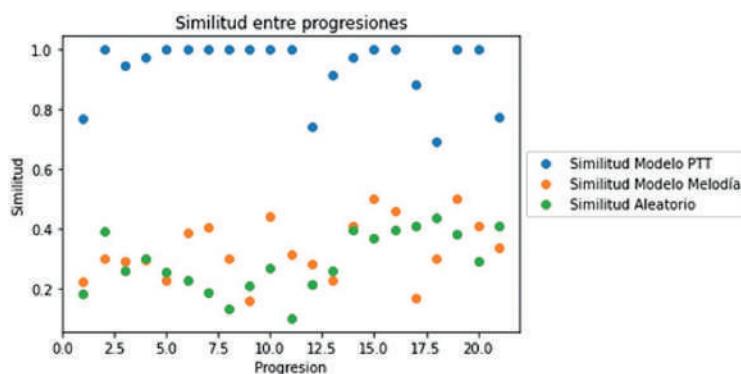


Figura 8: gráfica de similitud entre progresiones frente al grupo de verdad.

Tabla V
Resultados estadísticos

	Media	Desviación estándar
«Original» y «Modelo PTT»	93.6%	0.100
«Original» y «Modelo Melodía»	33%	0.099
«Original» y «Aleatorio»	28.9%	0.0976

Los resultados indican que el modelo PTT genera progresiones muy similares al corpus; esto significa que las secuencias armónicas generadas por este método tienen la capacidad de reproducir el estilo musical que representa el corpus. Por otro lado, el segundo modelo Melodía tiende a generar progresiones diferentes a las originales, lo que indica que su capacidad de imitación es inferior al primer modelo y que sus observaciones no permiten predecir el estilo musical subyacente en el corpus.

En general, el grupo «Modelo PTT» obtiene un promedio de similitud frente al grupo original del 93.6 %, superando así el promedio de «Modelo Melodía» de 33 % y el de «Aleatorio» de 28.9 %. Esto refuerza lo anteriormente dicho e indica que en promedio el modelo PTT captura en mejor medida las características y patrones de las progresiones de acordes de la base de datos original.

3.4. Experimento 4: prueba subjetiva

El experimento de prueba subjetiva se realiza con la finalidad de evaluar y comparar el rendimiento de los dos modelos HMM implementados, que mide qué tan bien se relacionan las armonías generadas con la melodía a la cual se le está aplicando la armonización. Además, se busca explorar la capacidad de los modelos desde otra dimensión teniendo en cuenta la subjetividad, factor importante que se debe tener presente en un arte como la música.

3.4.1. Preprocesamiento

Para realizar una evaluación del desempeño de los métodos implementados se eligen diez progresiones armónicas de las veintiuna generadas que se menciona al inicio de esta misma sección, obteniendo así cuatro muestras por canción correspondientes a los grupos de evaluación mencionados al inicio de esta sección (Original, PTT, Melodía y Aleatorio).

Se generan los archivos MIDI de cada progresión y se procede a formar los diez grupos de muestras, generando para cada canción escogida cuatro archivos de audio denominados «Fragmentos musicales» con melodía y armonía. Se usa un tempo de 120 BPM y todas las progresiones se establecen en la octava número cuatro del piano.

3.4.2. Procedimiento

En primera instancia se introduce el experimento y se explican las dos métricas correspondientes a la evaluación. Posterior a ello se realiza una sección de entrenamiento para que el sujeto comprenda el contexto

to y se familiarice con la prueba. Se busca poner dos fragmentos musicales que sean completamente opuestos en percepción de desempeño, es decir, uno correcto y otro que no lo es, teniendo en cuenta progresiones armónicas de la música popular occidental relacionadas al corpus utilizado en la investigación.³⁶

En cada iteración del experimento se dan al usuario cuatro fragmentos musicales para escuchar, todos ellos con la misma melodía y diferente armonía, cada armonía varía según los grupos de evaluación Original, PTT, Melodía y Aleatorio en la subsección 1 de este experimento. Después de escuchar cada «Fragmento musical», el sujeto procede a calificarlos en una escala de Likert de 0 a 5. Allí se encuentra con dos métricas³⁷ que, con sus respectivas escalas, son las siguientes:

- Fluidez de la armonización, en términos de transición de acordes: donde 0 es «la armonización presenta transiciones bruscas y abruptas entre acordes, dificultando la cohesión musical» y 5 corresponde a «la armonización muestra transiciones suaves y fluidas entre acordes, creando una progresión fluida y agradable».
- Enriquecimiento y transmisión emocional de la armonía a la melodía: donde 0 es «la armonía no realza ni complementa la expresión emocional de la melodía» y 5 corresponde a «la armonía enriquece y amplifica la expresión emocional de la melodía de manera notable».

La escala de calificación elegida tiene como objetivo evitar sesgos en el puntaje final

³⁶ Cuestionario de armonización de melodías: <https://form.jotform.com/231597775210056>

³⁷ Métricas que se basan en la investigación de Yin-Cheng Yeh et al., «Automatic Melody Harmonization with Triad Chords: A Comparative Study», *ArXiv* (2021).

de las métricas sin posibilidad de un punto medio en la calificación, permitiendo cuantificar el desempeño de forma tanto positiva como negativa. Se considera que un puntaje igual o mayor a 3 representa un desempeño positivo, mientras que aquellos inferiores a 3 se consideran negativos. Así, se plantean las siguientes hipótesis para el experimento:

- H1: se plantea que la métrica «Fluidez de la armonización» tenga una calificación promedio mayor o igual a 3 para el modelo PTT, lo que indica un desempeño positivo para dicho modelo.
- H2: la métrica «Enriquecimiento y transmisión emocional» tiene una calificación promedio mayor o igual a 3 para el modelo PTT, presentando un desempeño positivo para el modelo en ambas métricas evaluadas en el experimento.
- H3: la diferencia de medias entre el modelo Original y el modelo PTT es más baja que la diferencia de medias entre Original-Melodía y Original-Aleatorio.
- H4: con el objetivo de refutar la similitud observada en la figura 7, del modelo Aleatorio con el modelo Original, se plantea que la diferencia de medias sea la más alta para las dos métricas propuestas en la prueba.

3.4.3. Participantes

Para la consecución de los sujetos de prueba, se recopila para un total de 37 sujetos con experiencia académica específica en armonía. Divida en grupos de 1 a 3 años, 3 a 5 años y más de 5 años de experiencia.

3.4.4. Resultados y discusión

Se presentan en la tabla VI los estadísticos de media y desviación estándar para las

dos métricas evaluadas en el experimento. Además, se muestra de manera detallada la distribución de los datos en un diagrama de caja en las figuras 9 y 10 para las dos métricas evaluadas, respectivamente.

Tabla VI
Resultados estadísticos de los modelos

Métrica	Modelo	Media	Desviación estándar
Fluidez de la armonización	Original	3.72	1.12
	Melodía	2.92	1.28
	PTT	3.42	1.33
	Aleatorio	2.29	1.35
Enriquecimiento y transmisión emocional	Original	3.52	1.23
	Melodía	2.61	1.41
	PTT	3.18	1.42
	Aleatorio	1.89	1.36

Según se muestra en la tabla VI, las medias del puntaje del modelo PTT para las métricas «Fluidez de la armonización» y «Enriquecimiento y transmisión emocional» son 3.42 y 3.18, respectivamente. Estos resultados confirman las hipótesis H1 y H2, lo que indica un desempeño positivo del modelo PTT en cuanto a la generación de progresiones, tanto en la fluidez de la armonización como en el enriquecimiento y la transmisión emocional.

Es observable, además, e importante destacar que, aunque la media de puntaje del método Melodía no es mayor o igual a 3, se acerca por 0.08 al 3, lo que indica un desempeño que se acerca a un funcionamiento positivo, por lo que puede funcionar en casos donde se requiera acompañamientos armónicos menos técnicos. Esto sugiere que pueda ser valioso para la generación de progresiones armónicas menos predecibles o ceñidas a lo típico.

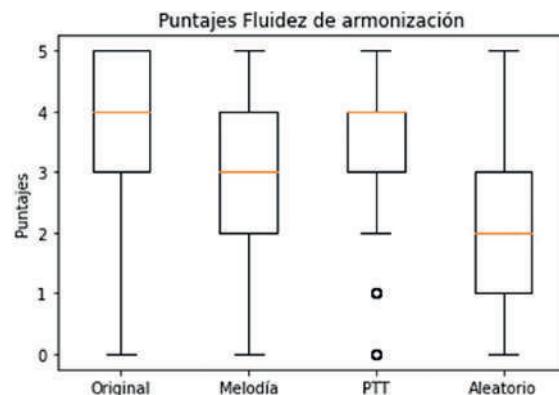


Figura 9: distribución de datos métrica «Fluidez de armonización».

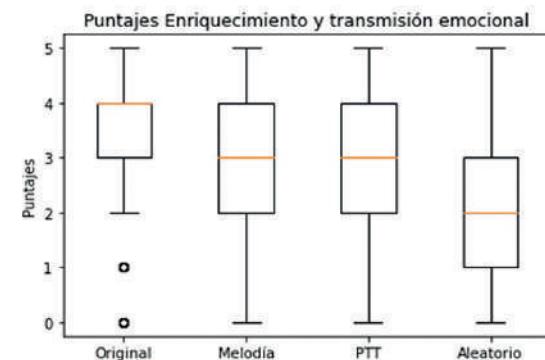


Figura 10: distribución de datos métrica «Enriquecimiento y transmisión emocional».

Se observa en la figura 9 que, aunque hay datos atípicos en el modelo PTT, en la métrica «Fluidez de armonización» comparte un intervalo en la distribución de la mayoría de los datos con el modelo Original siendo entre 3 y 5 para este, y entre 3 y 4 para el modelo PTT.

Tabla VII
Resultados de prueba anova métrica "fluidez de armonización"

Métrica	Valor P	Valor F	GL Num	GL Den
Fluidez de la armonización	1.14x10 ⁻⁵²	88.51	3	1476
Enriquecimiento y transmisión emocional	1.67x10 ⁻⁵⁹	101.06	3	1476

Se realiza un análisis de varianzas (ANOVA) para así determinar si hay diferencias significativas entre los grupos. Se observan los resultados del ANOVA en la tabla VII, con un valor de significancia $p < 0.05$ para ambas métricas, lo que indica que hay por lo menos un grupo con diferencias significativas a los demás en cuestiones de media esta-

dística. Con dicha información se procede a realizar una prueba de *post-hoc* de Tukey, que proporciona información más detallada acerca de las diferencias de medias entre grupos (resultados mostrados en la tabla VIII para la métrica «Fluidez de transición» y tabla IX para la métrica «Enriquecimiento y transmisión»).

Tabla VIII
Resultados de prueba *post-hoc* tukey
métrica «fluidez de armonización»

Grupo 1	Grupo 2	Diferencia de medias	Valor P
Aleatorio	Melodía	-0.3054	0.0066
Aleatorio	Original	-1.4324	0.0
Aleatorio	PTT	-0.8027	0.0
Melodía	Original	-1.127	0.0
Melodía	PTT	-0.4973	0.0
Original	PTT	0.6297	0.0

Tabla IX
Resultados de prueba *post-hoc* tukey para la
métrica «enriquecimiento y transmisión emocional»

Grupo 1	Grupo 2	Diferencia de medias	Valor P
Aleatorio	Melodía	-0.3378	0.0044
Aleatorio	Original	-1.6297	0.0
Aleatorio	PTT	-0.9081	0.0
Melodía	Original	-1.2919	0.0
Melodía	PTT	-0.5703	0.0
Original	PTT	0.7216	0.0

Se confirma la hipótesis H3 al observar los resultados en las tablas VIII y IX para ambas métricas evaluadas. La diferencia media entre el modelo Original y el modelo PTT es de 0.6297 en la primera métrica, un valor más bajo en comparación con los valores de 1.127 y 1.4324 obtenidos en las comparaciones entre el modelo Original y los modelos Melodía y Aleatorio, respectivamente. En la segunda métrica, se repite este patrón, con una diferencia media de 0.7216 entre el modelo Original y el modelo PTT, valor inferior a los 1.2919 y 1.6297 correspondientes a las comparaciones de medias entre el modelo Original y los otros dos modelos. Se observa, además, un valor *p* menor a 0.05 de las parejas de grupos, lo que significa una

diferencia estadística significativa entre ellos, lo cual no indica que haya un posible buen imitador del método Original con otro grupo, pero deriva en un análisis de la diferencia de medias para así determinar mayor o menor cercanía al grupo Original. Es importante destacar que los valores negativos de la diferencia medias observados en las tablas VIII y IX indican que la media del grupo 2 es mayor, mientras que un valor positivo indica que la media del grupo 1 es mayor. Sin embargo, la magnitud de este parámetro indica la diferencia de medias de igual manera.

Finalmente, se analiza las diferencias de medias del modelo Aleatorio con el Original y se observa que efectivamente muestra los valores más altos para ambas métricas. Estos valores son 1.4324 para la métrica «Fluidez de armonización» y 1.6297 para la métrica de «Enriquecimiento y transmisión emocional» por lo que se confirma la hipótesis H4 y se refuta una similitud del modelo Aleatorio con el modelo Original.

4. Conclusiones

En esta investigación se han desarrollado y comparado dos algoritmos de armonización de melodías basados en HMM. Uno de ellos utilizando notas presentes en la melodía y el otro utilizando perfiles de tensión tonal. A través de las pruebas objetivas y subjetivas con los grupos de evaluación «Original», «Aleatorio», «PTT» y «Melodía», se ha demostrado que el modelo PTT supera al modelo Melodía en la generación de patrones armónicos y progresiones similares al corpus con el que fueron entrenados.

Este descubrimiento representa un aporte en el campo de la generación automática de armonías musicales a través de HMM, especialmente al utilizar progresiones basadas en perfiles de tensión tonal

para un estilo musical particular.

Finalmente, se plantea la posibilidad de mejorar el modelo en el futuro mediante la eliminación de la dependencia de los puntos de inicio de los acordes y la reducción de los datos ingresados por el usuario.

5. Referencias

- Agostinelli, Andrea, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, Christian Frank. «MusicLM: Generating Music From Text». *ArXiv* (2023). <http://arxiv.org/abs/2301.11325>
- Alt, Helmut y Michael Godau. «Computing The Fréchet Distance Between Two Polygonal Curves». *International Journal of Computational Geometry & Applications*, vol. 5, n.º 1 y 2 (1995): 75–91. <https://doi.org/10.1142/s0218195995000064>
- An, Xin. «Prediction of Stock Price by Hidden Markov Model». *BCP Business & Management*, 38 (2023): 309–316.
- Bishop, Christopher M. *Pattern Recognition and Machine Learning*. Springer, 2006.
- Caetano, Marcelo, María Navarro-Cáceres y Gilberto Bernardes. «Objective Evaluation of Tonal Fitness for Chord Progressions Using the Tonal Interval Space». *Artificial Intelligence in Music, Sound, Art and Design* (2020): 150–164.
- Csiszar, I. «I-Divergence Geometry Of Probability Distributions And Minimization Problems». *The Annals of Probability*, vol. 3, n.º 1 (1975): 146–158.
- Farhan, Faisal, Ziaul Hoque y Provat Raihan. «Study of Pattern recognition system with Hidden Markov Model through the development of a raspberry pi based speaker recogniser». Tesis de grado, Ahsanullah University of Science & Technology, 2020. <https://www.researchgate.net/publication/348150123>
- Forney, G. David. «The Viterbi Algorithm».
- Formal Structure of Electromagnetics. Amsterdam vol. 61, n.º 3 (1973).
- Gómez-Marín, Daniel, Sergi Jordà y Perfecto Herrera. «Drum Rhythm Spaces: From Polyphonic Similarity to Generative Maps». *J New Music Res* vol. 49, n.º 5 (2020): 438–456.
- Groves, Ryan. «Automatic Harmonization Using a Hidden Semi-Markov Model». *AI-IDE Workshop Technical Report WS-13-22 (Musical Metacreation)*, vol. 9, n.º 5 (2013): 48–54.
- Ji, Shulei, Jing Luo y Xinyu Yang. «A Comprehensive Survey on Deep Music Generation: Multi-level Representations, Algorithms, Evaluations, and Future Directions». *ArXiv* (2020). <http://arxiv.org/abs/2011.06801>
- Lerdahl, Fred. «Tonal Pitch Space». *Music Perception: An Interdisciplinary Journal*, vol. 5, n.º 3 (1988): 315–350. <https://www.researchgate.net/publication/257925384>
- Lim, Hyungui, Seungyeon Rhyu y Kyogu Lee. «Chord Generation From Symbolic Melody Using BLSTM Networks». *ArXiv* (2017).
- Madsen, Søren Tjagvad y Gerhard Widmer. «Key-Finding With Interval Profiles». *Proceedings of the International Computer Music Conference (ICMC)*, Copenhague, Dinamarca, 2007.
- Miranda, Eduardo Reck. «Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity». En *Handbook of Artificial Intelligence for Music: Foundations, Advanced Approaches, and Developments for Creativity*. Edición de Eduardo Reck Miranda. Springer International Publishing, 2021. <https://doi.org/10.1007/978-3-030-72116-9>
- Morris, Dan, Ian Simon y Sumit Basu. «Exposing Parameters of a Trained Dynamic Model for Interactive Music Creation». *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence* (2008).
- Navarro-Cáceres, María, Marcelo Caetano, Gilberto Bernardes, Mercedes Sánchez-Barba y Javier Merchán Sánchez-Jara. «A Computational Model of Tonal Tension Profile of Chord Progressions in the Tonal Interval Space». *Entropy*, vol. 22, n.º 11 (2020): 1–30. <https://doi.org/10.3390/e22111291>

- Pachet, François, Pierre Roy y Benoit Carré. «Assisted Music Creation with Flow Machines: Towards New Categories of New». *ArXiv* (2020). <http://arxiv.org/abs/2006.09232>
- Paiement, Jean-François, Douglas Eck y Samy Bengio. «Probabilistic Melodic Harmonization». En *Advances in Artificial Intelligence. Canadian AI 2006. Lecture Notes in Computer Science*. Edición de Lamontagne, L., Marchand, M., 218–229. Berlín: Springer, 2006.
- Pardo, Bryan y William P. Birmingham. «Algorithms for Chordal Analysis». *Computer Music Journal*, vol. 26, n.º 2 (2002).
- Rabiner, Lawrence R. «A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition». *Proceedings on the IEEE*, vol. 77, n.º 2 (1989).
- Raczynski, Stanislaw, Satoru Fukayama, Emmanuel Vincent. «Melody Harmonisation with Interpolated Probabilistic Models». Reporte de investigación, RR-8110, INRIA, 2012. <https://hal.inria.fr/hal-00742957>
- Ramalho, Geber y Ricardo Scholz. «COCHONUT: Recognizing Complex Chords from MIDI Guitar Sequences». Conference: International Conferences on Music Information Retrieval, 2008. <https://www.researchgate.net/publication/269038716>
- Randall, Richard R. y Bilal Khan. «Lerdahl's Tonal Pitch Space Model and Associated Metric Spaces». *Journal of Mathematics and Music*, vol. 4, n.º 3 (2010): 121–131. <https://doi.org/10.1080/17459737.2010.529654>
- Rosenblatt, Murray. «Remarks On Some Nonparametric Estimates Of A Density Function». *Annals of Mathematical Statistics* (1956).
- Simon, Ian, Dan Morris y Sumit Basu. «My-Song: Automatic Accompaniment Generation for Vocal Melodies». *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery (2008): 725–734.
- Tsushima, Hiroaki, Eita Nakamura, Katsutoshi Itoyama, Kazuyoshi Yoshii. «Function-And Rhythm-Aware Melody Harmonization Based On Tree-Structured Parsing And Split-Merge Sampling Of Chord Sequences». 18th International Society for Music Information Retrieval Conference, Suzhou, China, 2017.
- Tsushima, Hiroaki, Eita Nakamura y Kazuyoshi Yoshii. «Bayesian Melody Harmonization Based on a Tree-Structured Generative Model of Chord Sequences and Melodies». *IEEE/ACM Transactions on Audio, Speech, and Language Processing* (2020).
- Van der Hoek, Jonh y Robert J. Elliot. *Introduction to Hidden Semi-Markov Models*. Cambridge University Press, 2018.
- Williams, J. K. «An Interactive, Multimedia Environment for Exploring Tonal Pitch Space». Conference ICL2007, Septiembre 26 -28, 2007.
- Wongsaroj, C., N. Prompoon y A. Surarerk. *Digital Information and Communication Technology and it's Applications (DICTAP)*, 2014 Fourth International Conference. 2014.
- Yang, Li-Chia y Alexander Lerch. «On the Evaluation of Generative Models in Music». *Neural Computing and Applications*, vol. 32, n.º 9 (2020): 4773–4784. <https://doi.org/10.1007/s00521-018-3849-7>
- Yeh, Yin-Cheng, Wen-Yi Hsiao, Satoru Fukayama, Tetsuro Kitahara, Benjamin Genczel, Hao-Min Liu, Hao-Wen Dong, Yian Chen, Terence Leong, Yi-Hsuan Yang. «Automatic Melody Harmonization with Triad Chords: A Comparative Study». *ArXiv* (2021).
- Zhang, Kejun, Xinda Wu, Tieyao Zhang, Zhi-jie Huang, Xu Tan, Qihao Liang, Songruo-yao Wu, Lingyun Sun. «WuYun: Exploring Hierarchical Skeleton-guided Melody Generation Using Knowledge-enhanced Deep Learning». *ArXiv* (2023).