



Centro de Investigación  
en Computación  
Instituto Politécnico Nacional

**INSTITUTO POLITÉCNICO NACIONAL  
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN**

**TESIS**

**Atribución de autoría en la música  
mediante el uso de aprendizaje automático**

**PARA OPTAR POR EL GRADO DE:  
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:**

Lic. Daniel Alejandro Pérez Alvarez

**DIRECTOR DE TESIS**

Dr. Alexander Gelbukh

Ciudad de México

junio 2019



**INSTITUTO POLITÉCNICO NACIONAL  
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

**ACTA DE REVISIÓN DE TESIS**

En la Ciudad de México siendo las 12:00 horas del día 03 del mes de junio de 2019 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

**Centro de Investigación en Computación**

para examinar la tesis titulada:

**"Atribución de autoría en la música mediante el uso de aprendizaje automático"**

Presentada por el alumno:

**PÉREZ**

Apellido paterno

**ÁLVAREZ**

Apellido materno

**DANIEL ALEJANDRO**

Nombre(s)

Con registro:

<b>A</b>	<b>1</b>	<b>7</b>	<b>0</b>	<b>6</b>	<b>2</b>	<b>6</b>
----------	----------	----------	----------	----------	----------	----------

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

**LA COMISIÓN REVISORA**

Director de Tesis

Dr. Alexander Gelbukh

Dr. Sergio Suárez Guerra

Dr. Jigar Batyrshin

Dr. Grigori Sidorov

Dra. Olga Kolesnikova

Dr. Luis Manuel Vilches Blázquez



**PRESIDENTE DEL COLEGIO DE PROFESORES**

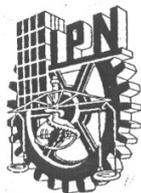
**INSTITUTO POLITÉCNICO NACIONAL**

**CENTRO DE INVESTIGACIÓN**

**EN COMPUTACIÓN**

**DIRECCIÓN**

Dr. Marco Antonio Ramírez Salinas



**INSTITUTO POLITÉCNICO NACIONAL**  
**SECRETARÍA DE INVESTIGACIÓN Y POSGRADO**

*CARTA CESIÓN DE DERECHOS*

En la Ciudad de México el día 12 del mes junio del año 2019, el (la) que suscribe Daniel Alejandro Pérez Alvarez alumno (a) del Programa de Maestría en Ciencias de la Computación con número de registro A170626, adscrito a Centro de Investigaciones en Computación CIC, manifiesta que es autor (a) intelectual del presente trabajo de Tesis bajo la dirección de Dr. Alexander Gelbukh y cede los derechos del trabajo intitulado Atribución de autoría en la música mediante el uso de aprendizaje automático, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección daperezalvarez@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

Daniel Alejandro Pérez Alvarez

Nombre y firma

## RESUMEN

En el presente estudio se aborda la atribución de autoría en la música como un problema supervisado de clasificación. Se comparan dos novedosas representaciones para piezas musicales en formato simbólico. La primera representación se basa en n-gramas y toma en cuenta las relaciones melódicas entre instrumentos en piezas polifónicas. La segunda representación está basada en vectores densos y es capaz de atrapar relaciones melódicas y armónicas entre notas musicales. Los resultados obtenidos indican que los modelos basados en la representación densa superan ampliamente a los modelos basados en n-gramas. Nuestro modelo de red neuronal convolucional basado en vectores densos de notas logra superar el estado del arte para un conjunto de datos compuesto por cuartetos de cuerda de Mozart y Haydn.

## **ABSTRACT**

In the present study the authorship attribution in music is approached as a supervised problem of classification. Two novel representations for musical pieces in symbolic format are compared. The first representation is based on n-grams and takes into account the melodic relationships between instruments in polyphonic pieces. The second representation is based on dense vectors and is able to catch melodic and harmonic relationships between musical notes. The results obtained indicate that models based on dense representation far outperform n-gram based models. Our convolutional neural network model based on dense vectors of musical notes manages to overcome the state of the art for a dataset composed of string quartets by Mozart and Haydn.

## **AGRADECIMIENTOS**

A mi esposa Ariadna, sin la cual este trabajo no hubiera sido posible.

A mi padre y mi madre.

A toda mi familia.

A mi asesor Dr. Alexander Gelbukh.

A los miembros de mi comité tutorial Dr. Grigori Sidorov, Dr. Sergio Suárez, Dr. Ildar Batyrshin, Dra. Olga Kolesnikova, Dr. Luis Manuel Vilches.

A profesores y amigos de esta excelente Universidad. A Conacyt por el apoyo económico para poder realizar mis estudios en este maravilloso país.

## ÍNDICE

RESUMEN	4
ABSTRACT	5
AGRADECIMIENTOS	6
<b>Capítulo 1. Introducción</b>	
1.1 Planteamiento del problema	12
1.2 Objetivos	14
1.2.1 Objetivo General	14
1.2.2 Objetivos Específicos	14
1.3 Justificación	14
1.4 Aportaciones	15
1.5 Estructura del documento	16
<b>Capítulo 2. Marco Teórico</b>	
2.1 Elementos de la teoría musical. Definiciones	18
2.1.1 Nota	18
2.1.2 Semitono	19
2.1.3 Alteraciones	20
2.1.4 Escala	20
2.1.5 Intervalo	23
2.1.6 Acorde	24
2.1.7 Melodía	25
2.1.8 Armonía	25
2.1.9 Ritmo	26
2.1.10 Transposición	26
2.1.11 Tonalidad	27
2.1.12 Círculo de quintas	29
2.2 Métodos de clasificación	31
2.2.1 Aprendizaje automático supervisado	31
2.2.1.1 Regresión Logística	31
2.2.1.2 Máquinas de soporte vectorial	31
2.2.2 Aprendizaje profundo	32
2.2.2.1 Redes neuronales convolucionales	33
2.3 Audio vs Representación Simbólica	34
2.4 Piano roll	35
2.5 Matriz términos-frecuencia. N-gramas	35
2.6 Representación densa	37
<b>Capítulo 3. Estado del Arte</b>	
3.1 Antecedentes de estudios sobre Atribución de autoría	38
3.2 Antecedentes de estudios sobre N-gramas	41
3.3 Antecedentes de estudios sobre Representación Densa	42
<b>Capítulo 4 Método</b>	
4.1 Fase de Pre-procesamiento	46
4.2 Fase de Representación	47
4.2.1 Representación basada en n-gramas	47
4.2.2 Representación basada en vectores densos	48
4.2.2.1 Pre-entrenamiento de vectores	48

4.3 Fase de Clasificación	50
4.3.1 Modelos basados en representación de n-gramas	50
4.3.2 Modelos basados en representación densa	50
<b>Capítulo 5 Resultados y Discusión</b>	
5.1 Modelos basados en n-gramas	53
5.2 Visualización de vectores densos	55
5.3 Uso de vectores densos mediante aprendizaje profundo	59
<b>Conclusiones y Recomendaciones</b>	61
<b>Referencias Bibliográficas</b>	63

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Notas musicales	19
<b>Figura 2.</b> Tempo de cada nota y su silencio correspondiente	19
<b>Figura 3.</b> Ejemplo de tonos y semitonos	19
<b>Figura 4.</b> Escala de Do mayor representada en forma de escalera	21
<b>Figura 5.</b> Escala diatónica donde se muestran tonos y semitonos	21
<b>Figura 6.</b> Representación de una escala diatónica en el piano	22
<b>Figura 7.</b> Escala cromática donde aparecen notas naturales y alteradas	22
<b>Figura 8.</b> Representación de una escala cromática	23
<b>Figura 9.</b> Ejemplo de intervalo de 5ta justa	23
<b>Figura 10.</b> Ejemplo de acordes de tres notas	24
<b>Figura 11.</b> Ejemplo de melodía	25
<b>Figura 12.</b> Ejemplo de armonía a 4 voces. Relación horizontal y vertical de la armonía	25
<b>Figura 13.</b> Distribución del tiempo según las figuras	26
<b>Figura 14.</b> Grados en la escala en do mayor y en sol mayor	27
<b>Figura 15.</b> Ejemplo de transposición	27
<b>Figura 16.</b> Tonalidad de do mayor y su relativa la menor	28
<b>Figura 17.</b> Círculo de quintas	30
<b>Figura 18.</b> Visualización de vectores densos basados en acordes	43
<b>Figura 19.</b> Vista general de procedimientos	46
<b>Figura 20.</b> Modelo de red convolucional utilizado en este trabajo	52
<b>Figura 21.</b> Visualización de vectores densos. Alteraciones	55
<b>Figura 22.</b> Visualización de vectores densos. Relación de octavas	56
<b>Figura 23.</b> Visualización de vectores densos. Relación de terceras y sextas	57
<b>Figura 24.</b> Visualización de vectores densos en dirección melódica (I)	58
<b>Figura 25.</b> Visualización de vectores densos en dirección melódica (II)	58

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Nomenclaturas musicales (Latina y Anglosajona)	18
<b>Tabla 2.</b> Descripción de las alteraciones musicales	20
<b>Tabla 3.</b> Intervalos y semitonos	24
<b>Tabla 4.</b> Tonalidades mayores y sus relativas menores	29
<b>Tabla 5.</b> Comparación de clasificadores basados en n-gramas	53
<b>Tabla 6.</b> Comparación de modelo basado en n-gramas con estado del arte	54
<b>Tabla 7.</b> Comparación de modelo basado en representación densa con el estado del arte	60

## **Capítulo 1. Introducción**

Históricamente la ciencia de la computación ha buscado comprender y traducir a su propio lenguaje expresiones artísticas, entre ellas: la música. El uso de las nuevas tecnologías para la automatización de procedimientos relacionados con la música ha cobrado gran auge desde hace algunos años. Ejemplos comunes son el uso de técnicas de Inteligencia Artificial para la generación automática de melodías y de Procesamiento de Audio para la traducción de archivos de grabaciones musicales al lenguaje de símbolos de un pentagrama (transcripción). El reconocimiento de estilos y géneros musicales, o de estados anímicos generados por la música ha sido abordado por muchos investigadores. Con el arribo al mercado de aplicaciones como Spotify los estudios sobre recomendación de música gozan también de gran importancia.

Todo lo anterior forma parte de la Recuperación de Información Musical, del inglés MIR (Music Information Retrieval), la cual combina disciplinas como ciencias de la computación, psicología, musicología y procesamiento de señales para crear aplicaciones del mundo real capaces de extraer información relevante de la música. MIR se enfoca en resolver tareas tales como recomendación de música, transcripción musical automática y separación de pistas (Velardo, et al., 2016).

La identificación de autoría es uno de los mayores retos que se ha planteado la ciencia computacional. La propia complejidad del lenguaje musical, exige de los investigadores en este campo, adquirir primero un dominio de categorías musicales, comprender el funcionamiento e interrelación entre ellas, para posteriormente establecer análisis y realizar propuestas desde su ciencia.

El musicólogo, es el especialista encargado de analizar obras musicales, estudiar sus estilos y en también investigar, en caso de no conocerse, la autoría de piezas cuyo compositor no se ha definido. Sin embargo, esta última tarea puede resultar de gran complejidad para este profesional, al tener que comparar y evaluar gran cantidad de obras y estilos, para poder llegar a conclusiones certeras. Recientes investigaciones han demostrado que las máquinas pueden también “incursionar” en la tarea de la identificación de autoría de obras musicales.

Nuestra investigación desarrolla un sistema computacional para el problema supervisado de Atribución de Composiciones, basado en la adaptación al campo de la música de representaciones propias del dominio de lenguaje natural.

### **1.1 Planteamiento del Problema**

La Atribución automatizada de Autoría es el problema de identificar al autor de un texto anónimo o cuya autoría está en duda (Kešelj, et al., 2003). Aunque es un concepto normalmente ligado a textos escritos también puede ampliarse a la música y otros campos como análisis forense y comercio electrónico (Koppel, et al., 2009). La AA se basa en establecer el perfil de un autor mediante aquellas características que lo hagan único. Es por ello que para AA son más usados los atributos de estilo del autor que los de contenido de la obra. Entre las técnicas aplicadas en AA están las estadísticas y las de aprendizaje automático.

Desde la llegada de internet el número de archivos relacionados con música a disposición de todos ha aumentado considerablemente. Estas condiciones, sumadas a la presencia de aplicaciones cada vez más perfeccionadas para la creación de música conllevan a que una persona con conocimientos básicos de la materia pueda

mostrar sus obras a un gran número de personas. Esto hace que la cantidad de obras musicales registradas en los últimos años crezca y que resulte imposible para los expertos analizar cada una de ellas para determinar su originalidad.

Con la introducción de las computadoras, los psicólogos, los científicos de la computación y los ingenieros han perseguido el objetivo de diseñar sistemas automáticos para tres propósitos principales: análisis de la música, síntesis y comprensión. En los últimos años, la creciente dimensión de los archivos digitales y la distribución global de información en la web están llevando a la comunidad de investigación hacia un nuevo objetivo: la clasificación de la música. El objetivo es diseñar sistemas capaces de imitar la capacidad humana para hacer categorías a partir de supuestos estilísticos, observaciones perceptivas o experiencia personal. Este tipo de sistemas podrían adoptarse para etiquetar los datos musicales por género u otras características relacionadas; lo cual debería mejorar el diseño de las bases de datos multimedia indexadas por contenido y la efectividad de los sistemas de recuperación de información.

Sucede también que al existir mayor acceso a la información y con la digitalización de la sociedad los investigadores pueden indagar con más facilidad en archivos de relativa antigüedad y consultar fuentes diversas, no sólo de su localidad o país de origen. De esta forma se encuentran muchas veces valiosas partituras u otros documentos con contenido musical de los cuales se desconoce su autor.

Las técnicas computacionales son herramientas valiosas para, en casos como los anteriores, ayudar a musicólogos, críticos y demás expertos en el área, realizar una correcta toma de decisiones durante la investigación que desarrollen.

## **1.2 Objetivos**

### 1.2.1 Objetivo General

Desarrollar un sistema computacional para el problema supervisado de atribución de composiciones basado en la adaptación al campo de la música de representaciones propias del dominio de lenguaje natural.

### 1.2.2 Objetivos Específicos

1. Desarrollar una representación para piezas musicales basada en n-gramas que combinen la información hallada en las diferentes líneas instrumentales de una pieza musical.
2. Evaluar la representación del punto 1 para el problema supervisado de atribución de composiciones mediante el uso de clasificadores de aprendizaje automático.
3. Desarrollar una representación vectorial densa de notas basada en Word2vec que modele las relaciones melódicas y armónicas entre estas.
4. Evaluar la representación del punto 3 para el problema supervisado de atribución de composiciones mediante el uso de clasificadores de aprendizaje profundo.

## **1.3 Justificación**

En la actualidad existe una creciente utilización de los archivos digitales así como de su distribución global. Los investigadores que incursionan en el campo de computación unida a la música se han planteado en los últimos años un nuevo objetivo: la clasificación de la música.

Este propósito demanda diseñar sistemas capaces de imitar la capacidad humana para hacer categorías a partir de supuestos estilísticos, observaciones perceptivas o experiencia personal.

También el campo de la Musicología, requiere y demanda en la actualidad, de nuevas alternativas tecnológicas que faciliten los estudios de clasificación de autoría, géneros musicales, entre otros elementos, de ahí que se justifique la realización de nuestro estudio.

#### **1.4 Aportaciones**

La aportación principal se refleja en la propuesta de dos representaciones para piezas musicales basadas en la adaptación de técnicas propias del dominio de lenguaje natural: n-gramas y vectores densos.

La propuesta, validación y evaluación de estas representaciones ofrecerán nuevos procedimientos y modos de actuación a futuros investigadores que quieran continuar incursionando en este tema.

Dado el enfoque transdisciplinar, se explora teórica y metodológicamente un tema que puede resultar enriquecedor para la ciencias de la computación y también para la musicología.

Varias aplicaciones prácticas se derivan de este trabajo, por ejemplo, advertir a jóvenes compositores de semejanzas entre el trabajo que vienen creando y el realizado previamente por compositores clásicos, señalar a críticos de arte las influencias que posee un compositor emergente. Además pudiera ayudar a resolver cuestiones de plagio entre dos autores.

El campo de la musicología se ve beneficiado, al facilitarle una de las tareas más complejas que este campo científico se plantea: identificar un compositor a partir de una partitura musical. También contribuye a determinar a qué compositor pertenece con mayor probabilidad una pieza musical desconocida que se encuentre en alguna investigación o de forma casual.

Los modelos aquí desarrollados pueden aplicarse también a problemas similares del campo de la música como son: clasificación de música por regiones, por emociones que transmite (alegría, tristeza, fuerza, etc.), entre otras.

Dentro de las aportaciones potenciales se contribuiría a mejorar el diseño de las bases de datos multimedia indexadas por contenido y la efectividad de los sistemas de recuperación de información.

## **1.5 Estructura del documento**

En el trabajo inicialmente se ofrece una Introducción, donde se plantean los Objetivos Generales y Específicos del estudio, así como la Justificación y Aportaciones. Seguidamente se presenta un Marco Teórico sobre elementos generales de la teoría musical (Nota, Acorde, Melodía, Armonía, Ritmo, Transposición). También se explican los métodos de clasificación usados en nuestra investigación basados en aprendizaje automático y su sub-categoría: aprendizaje profundo. En este mismo acápite se abordan las Representaciones digitales, puntualizando en la dicotomía: Audio vs. Representación simbólica. Además se presenta la representación piano-roll, una de las representaciones tradicionales más usadas en los estudios computacionales sobre música. A continuación se definen los términos Matriz término-frecuencias y Representación densa. Luego de presentar el

Estado del Arte relacionado con Atribución de autoría, N-gramas y Representación densa, se ofrece la propuesta de Modelos (modelos basados en n-gramas y modelos basados en aprendizaje profundo) y se establece una comparación entre estos modelos. Por último se brindan Resultados, Conclusiones y Recomendaciones y Referencias bibliográficas.

## Capítulo 2. Marco Teórico

Es necesario explicar determinados conceptos relacionados con nuestro tema de investigación para poder comprender de una mejor manera el problema que nos ocupa.

### 2.1 Elementos de Teoría musical. Definiciones

#### 2.1.1 Nota

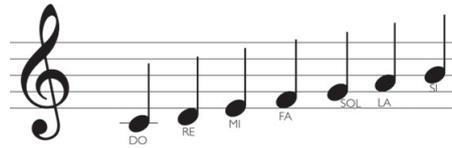
La nota es el elemento básico de la composición musical. Está compuesta por duración y altura (o tonalidad). Una nota se define como una unidad con cierto desplazamiento de tiempo desde el comienzo de una pieza musical que posee una duración y tono. Además del sistema latino de notación musical: do, re, mi, fa, sol, la, si; existe otro que es muy utilizado igualmente en la actualidad: el anglosajón. A continuación se presenta una tabla con ambos tipos de nomenclaturas.

Nomenclatura Latina	Nomenclatura Anglosajona
Do	C
Re	D
Mi	E
Fa	F
Sol	G
La	A
Si	B

**Tabla 1. Nomenclaturas musicales (Latina y Anglosajona)**

El *silencio* puede sustituir el lugar de la nota en el pentagrama. Este se define como una unidad con cierto desplazamiento de tiempo desde el comienzo de una pieza musical que posee duración y sin tono. Los descansos indican el silencio del

instrumento en particular, mientras que las notas indican el sonido (Herlands, Der, Greenberg, & Levin, 2014).



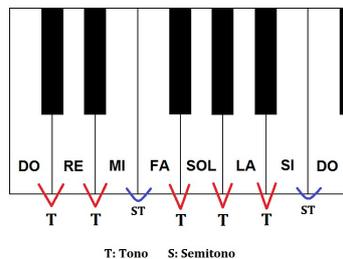
**Figura 1. Notas musicales**



**Figura 2. Tempo de cada nota y su silencio correspondiente**

### 2.1.2 Semitono

Un semitono es cada una de las dos partes, iguales o desiguales, en que se divide el intervalo de un tono, y es el menor de los intervalos que se pueden producir entre notas consecutivas de una escala diatónica. Equivale a la doceava parte de una octava.



**Figura 3. Ejemplo de tonos y semitonos**

### 2.1.3 Alteraciones

Las alteraciones musicales son modificaciones de altura de sonidos o notas musicales, por elevación o disminución. Se representan en forma de símbolos de acuerdo a la alteración. Las alteraciones en el pentagrama<sup>1</sup> se escriben al lado izquierdo de la nota musical. Ellas son: sostenido, bemol, doble sostenido, doble bemol y becuadro, como se presenta en la siguiente tabla:

Símbolo	Nombre de la Alteración	Descripción
#	Sostenido	Elevan la nota musical un semitono
×	Doble sostenido	Elevan la nota musical un tono
b	Bemol	Disminuyen la nota musical un semitono
bb	Doble bemol	Disminuyen la nota musical un tono
□	Becuadro	Anula el efecto de cualquier alteración anterior (sostenido o bemol)

**Tabla 2. Descripción de las alteraciones musicales**

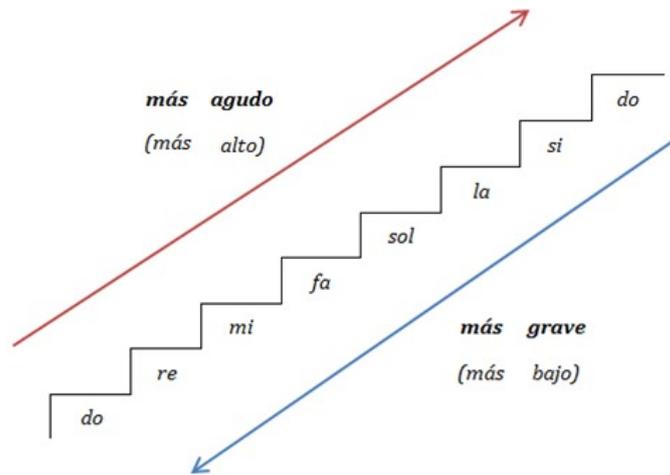
### 2.1.4 Escala

Como mismo una escala, de forma general, consiste en una secuencia de elementos, las escalas musicales son secuencias de sonidos en serie que se encuentran ordenados según su altura; cada sonido nuevo que aparece es más agudo que el anterior a medida que la escala asciende, y más grave a medida que la

---

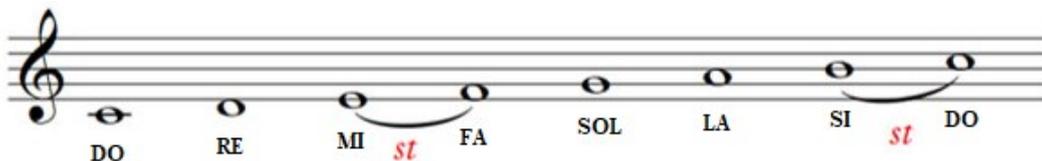
<sup>1</sup> Conjunto de 5 líneas y 4 espacios donde se escribe la música.

escala descendente. Si lo comparáramos con una escalera sería de la siguiente manera:

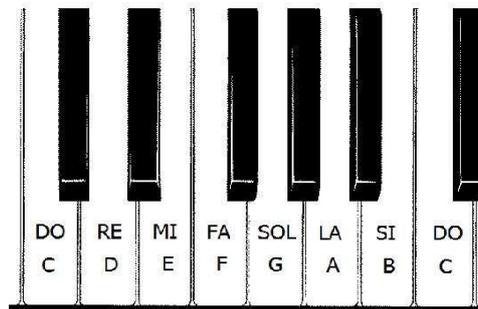


**Figura 4. Escala de Do mayor representada en forma de escalera**

Existen varios tipos de escalas. Las escalas diatónicas están formadas por ocho sonidos, siendo el último la repetición del primero, pero una octava por encima de este. Los intervalos entre las notas tienen un orden específico: dos tonos (uno entre do y re) (un tono entre re y mi) -un semitono (st) (entre mi y fa) - 3 tonos (un tono entre fa y sol) (un tono entre sol y la) (un tono entre si y do) - un semitono (st) (entre si y do) como se aprecia en la Figura 5. Esta escala (y su representación de tonos y semitonos) puede apreciarse en el piano en la Figura 6.

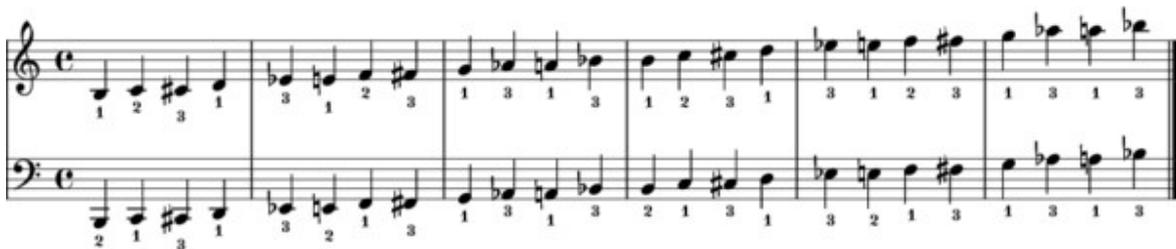


**Figura 5. Escala diatónica donde se muestran tonos y semitonos**



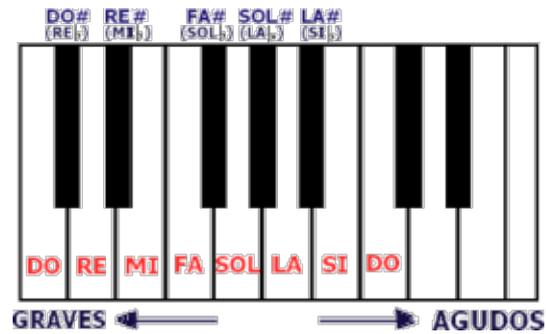
**Figura 6. Representación de una escala diatónica en el piano**

También se encuentran las escalas cromáticas, que se caracterizan por tener entre nota y nota solo un intervalo de un semitono. Estas escalas ascienden o descienden tomando en cuenta las notas naturales y también las notas alteradas (ya sea con sostenidos # o bemoles *b*) como se aprecia en la figura siguiente:



**Figura 7. Escala cromática donde aparecen notas naturales y alteradas**

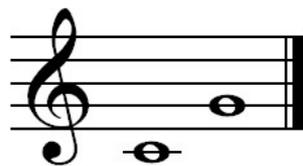
En el piano, un ejemplo de la representación de una escala cromática puede verse en la Figura 8. De forma ascendente se leen los sostenidos (#) y de forma descendente se leen los bemoles (*b*). Sostenidos y bemoles serán explicados en el siguiente acápite.



**Figura 8. Representación de una escala cromática**

### 2.1.5 Intervalo

Intervalo es la diferencia de altura (frecuencia) entre dos notas musicales, medida cualitativamente en grados o notas naturales y cuantitativamente en tonos y semitonos. Puede definirse también como la distancia que existe entre dos nota musicales. Los intervalos pueden ser de 2da, 3ra, 4ta, 5ta, 6ta, 7ma, 8va... etc. (Ver ejemplo en la Figura 12). Algunos intervalos, según la disposición en que se encuentren las notas y la cantidad de semitonos que existen entre ellas, pueden ser considerados como menor o mayor; y también aumentados, disminuidos o justos. Por ejemplo de do a mi existe un intervalo de 3ra mayor, en caso del intervalo de do a la bemol (b), se clasifica como 6ta menor, etc. Ver tabla 2.



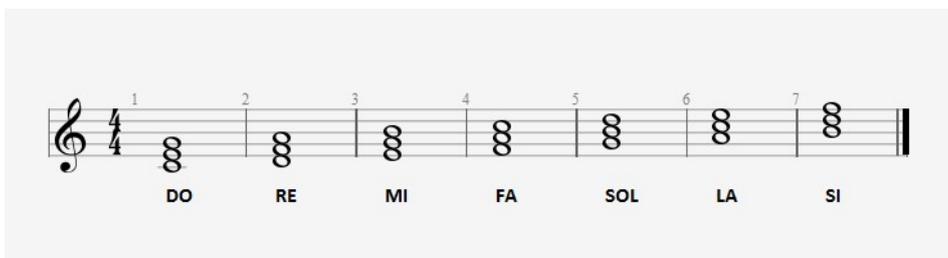
**Figura 9. Ejemplo de intervalo de 5ta justa**

Nombre del intervalo y número de semitonos que implica
Unísono = para cuando tienen el mismo sonido
Segunda menor = 1 semitono
De segunda mayor = 2 semitonos
Tercera disminuida = 2 semitonos
De tercera menor = 3 semitonos
Segunda aumentada = 3 semitonos
Tercera mayor = 4 semitonos
Cuarta disminuida = 4 semitonos
De cuarta justa = 5 semitonos
Tercera aumentada = 5 semitonos
Tritono (corresponde a la cuarta aumentada o quinta disminuida) = 6 semitonos
Quinta justa = 7 semitonos
Sexta disminuida = 7 semitonos
De sexta menor = 8 semitonos
Quinta aumentada = 8 semitonos
Sexta mayor = 9 semitonos
Séptima disminuida = 9 semitonos
Séptima menor = 10 semitonos
Séptima mayor = 11 semitonos
Octava justa = 12 semitonos

**Tabla 3. Intervalos y semitonos**

### 2.1.6 Acorde

Combinación de notas simultáneas, normalmente definida como una combinación no menor de tres sonidos. Es decir, los acordes están formados por tres notas o más (siempre que sea posible para el músico poder ejecutarlas simultáneamente). En este trabajo se asume que forman parte del acorde todas las notas que se escuchan en un mismo instante de tiempo.

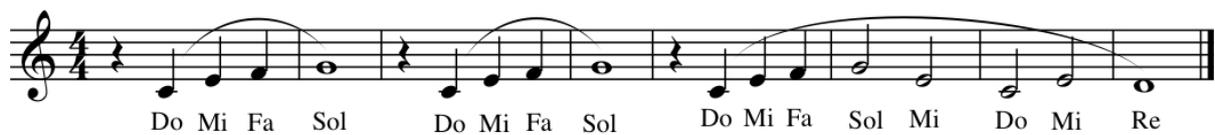


**Figura 10. Ejemplo de acordes de tres notas**

### 2.1.7 Melodía

Es la sonoridad consecutiva de notas diferentes y con significado musical sobre una misma voz o instrumento. Actúa sobre el eje tiempo.

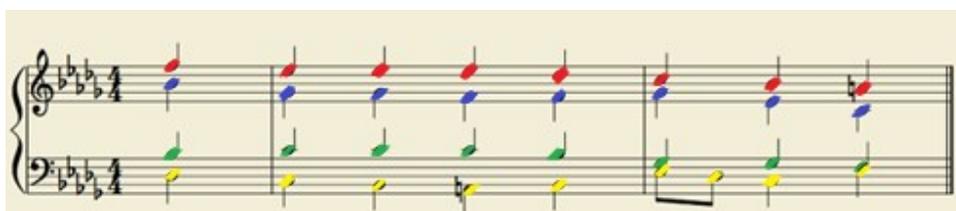
Melodía:



**Figura 11. Ejemplo de melodía**

### 2.1.8 Armonía

Es la sonoridad simultánea de notas diferentes y con significado musical. Actúa sobre el eje espacio. La armonía de una obra se basa en la sonoridad de los acordes y en su disposición tanto horizontal como vertical. La disposición horizontal contempla cada una de las voces de forma independiente. A continuación se brinda un ejemplo donde se le ha dedicado un mismo color a cada voz melódica. La disposición vertical contempla las cuatro voces (4 colores) a la vez que estas suenan simultáneamente.



**Figura 12. Ejemplo de armonía a 4 voces. Relación horizontal y vertical de la armonía**

### 2.1.9 Ritmo

Es una fuerza o un movimiento formado por una cierta sucesión de sonidos. Puede decirse que el ritmo de la música se compone de **ciclos que se reiteran en intervalos temporales**. Las notas y los silencios se presentan a lo largo de una melodía y definen el ritmo de la misma. En el ritmo entra en juego la repetición, en determinados intervalos, de sonidos breves, largos, débiles y fuertes.



**Figura 13. Distribución del tiempo según las figuras**

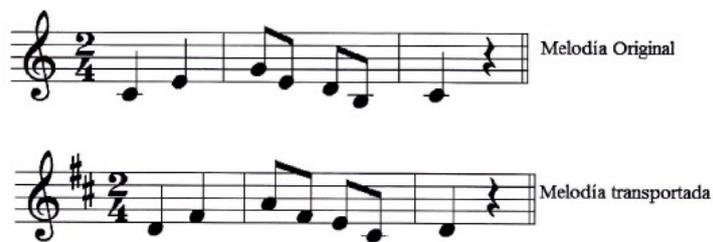
### 2.1.10 Transposición

La transposición o el transporte significa que un pasaje musical se traslada de una tonalidad (o grupo tonal) a otra. Técnicamente transponer o transportar una melodía o un grupo de acordes consiste en llevar todas las notas o acordes que la componen hacia arriba o hacia abajo en la escala musical, manteniendo en todas las notas el mismo intervalo entre nota de origen y nota de destino. Es conveniente tener presentes todos los semitonos que componen una escala cromática. Normalmente esto se hace para acomodar el acompañamiento a un registro de voz dado,

facilitando la tarea del músico o cantante. Para realizar una transposición es necesario tener claro cada uno de los grados de la escala de origen y la escala de destino (Ver Figura 14). Posteriormente se determinan los intervalos que se va dando en cada una de las líneas melódicas y se replica este intervalo en la nueva escala sin perder de vista los grados análogos en cada una de las escalas. (Ver Figura 15).

DO	RE	MI	FA	SOL	LA	SI	DO
I	II	III	IV	V	VI	VII	I
SOL	LA	SI	DO	RE	MI	#FA	SOL
I	II	III	IV	V	VI	VII	I

**Figura 14. Grados en la escala en do mayor y en sol mayor**



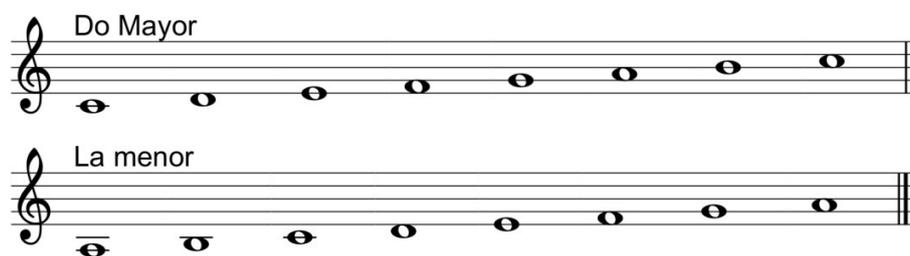
**Figura 15. Ejemplo de transposición**

### 2.1.11 Tonalidad

La tonalidad hace alusión al tono de una obra. El tono es un conjunto de sonidos que se sostienen en una misma nota principal llamada tónica. La sucesión de esas notas componen las escalas.

La tonalidad designa cuál es la nota tónica, la nota dominante y la subdominante, en las que se apoyará la estructura musical. La tonalidad define cuales son las reglas que se siguen para interpretar una obra musical. Es importante señalar que unido a la tonalidad están las categorías: relativa mayor y relativa menor.

A cada tonalidad mayor le corresponde una relativa mayor y viceversa. Un ejemplo sería la tonalidad de do mayor y su relativa la menor. Ver Figura 16.



**Figura 16. Tonalidad de do mayor y su relativa la menor**

A continuación se presenta la Tabla 4, que resume todas las tonalidades mayores que existen así como sus relativas mayores y el número de alteraciones que presenta cada una de ellas. Igualmente se ofrecen las abreviaturas que usualmente se ofrecen en la literatura.

Tonalidad mayor		Relativa menor		Alteraciones
Do bemol mayor	<b>Cb</b>	La bemol menor	<b>Abm</b>	7
Do mayor	<b>C</b>	La menor	<b>Am</b>	0
Do sostenido mayor	<b>C#</b>	La sostenido menor	<b>A#m</b>	7
Re bemol mayor	<b>Db</b>	Si bemol menor	<b>Bbm</b>	5
Re mayor	<b>D</b>	Si menor	<b>Bm</b>	2
Mi bemol mayor	<b>Eb</b>	Do menor	<b>Cm</b>	3
Mi mayor	<b>E</b>	Do sostenido menor	<b>C#m</b>	4
Fa mayor	<b>F</b>	Re menor	<b>Dm</b>	1
Fa sostenido mayor	<b>F#</b>	Re sostenido menor	<b>D#m</b>	6
Sol bemol mayor	<b>Gb</b>	Mi bemol menor	<b>Ebm</b>	6
Sol mayor	<b>G</b>	Mi menor	<b>Em</b>	1
La bemol mayor	<b>Ab</b>	Fa menor	<b>Fm</b>	4
La mayor	<b>A</b>	Fa sostenido menor	<b>F#m</b>	3
Si bemol mayor	<b>Bb</b>	Sol menor	<b>Gm</b>	2
Si mayor	<b>B</b>	Sol sostenido menor	<b>G#m</b>	5

**Tabla 4. Tonalidades mayores y sus relativas menores**

#### 2.1.12 Círculo de quintas.

En teoría musical, el círculo de quintas representa las relaciones entre los doce tonos de la escala cromática, sus respectivas armaduras de clave y las tonalidades relativas mayores y menores. Se trata de una representación geométrica de las relaciones entre los 12 tonos de la escala cromática en el espacio entre tonos. El diseño del círculo resulta útil a la hora de componer y armonizar melodías, construir acordes y desplazarse a diferentes tonalidades dentro de una composición.

En la parte superior se ubica la nota do. Moviéndonos hacia la derecha, la siguiente nota que vamos a poner es la quinta de do, o sea, sol. La siguiente, la quinta de sol, o sea, re; si seguimos así, aumentando en un intervalo de quinta la nota anterior, nos saldrán estas notas:

do-sol-re-la-mi-si-fa#

Si siguiéramos escribiendo notas con una quinta de diferencia, la siguiente nota que seguiría a la secuencia de arriba sería do#, pero a partir de esa nota vamos a poner, en lugar de los sostenidos, sus enarmónicos bemoles. La nota enarmónica de do# es reb. Ahora vamos a escribir todas las notas hasta completar el círculo:

do-sol-re-la-mi-si-fa#-reb(do#)-lab(sol#)-mib(re#)-sib(la#)-fa

La siguiente nota a fa volvería a ser do otra vez. Hemos completado el círculo, y han quedado representados los 12 sonidos de la escala cromática.

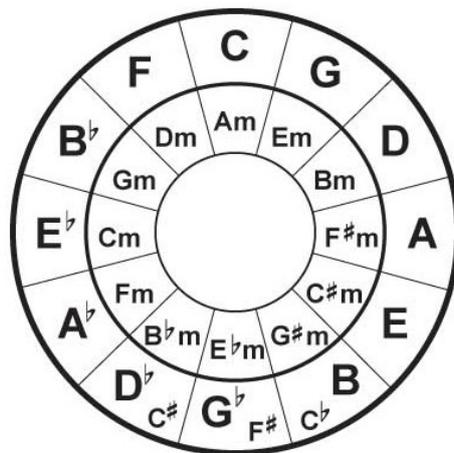


Figura 17. Círculo de quintas

## **2.2 Métodos de clasificación**

### 2.2.1 Aprendizaje automático supervisado

El aprendizaje automatizado supervisado es la búsqueda de algoritmos que razonan desde instancias suministradas externamente para producir hipótesis generales que luego hacen predicciones sobre instancias futuras. En otras palabras, el objetivo del aprendizaje supervisado es construir un modelo conciso de la distribución de las etiquetas de las clases en términos de características predictoras (Kotsiantis, et al., 2007).

#### 2.2.1.1 Regresión logística

En estadística, la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. La Regresión Logística es una de las técnicas estadístico-inferenciales más empleadas en la producción científica contemporánea. En nuestra investigación la regresión logística será utilizada como un clasificador.

#### 2.2.1.2 Máquinas de Soporte Vectorial

Entre las técnicas de Aprendizaje Automático que con más frecuencia se presentan en la literatura están las Máquinas de Soporte Vectorial (SVM por sus siglas en inglés) (Boser, et al., 1992; Boser, et al., 1992; Cortes & Vapnik, 1995). Para poner en práctica este método en primer lugar se realiza una extracción de características de los objetos a analizar que pueden ser textos, partituras musicales etc. El

resultado de este pre-procesamiento es una serie de vectores n-dimensionales ordenados de características que servirá como entrada al método de clasificación. A continuación se representan los puntos de muestra (vectores) en un espacio n-dimensional y se definen dos clases diferentes de manera tal que cada punto de muestra pertenezca solamente a una de las clases.

La idea básica de las SVMs es construir un hiperplano óptimo entre todos los posibles hiperplanos que divida las clases con el mayor margen de separación entre ellas; el margen es la suma de las distancias desde el hiperplano hasta los puntos más cercanos de cada una de las dos clases. Muchas veces los datos no pueden ser separados mediante una función lineal y deben utilizarse funciones kernel como la polinomial o la RBF (Función de Base Radial). En el caso simple cuando los datos sí pueden ser separados linealmente decimos que se usa un kernel lineal.

Las SVMs fueron diseñadas originalmente para clasificación binaria (Hsu & Lin, 2002), luego fueron extendidas para clasificaciones con múltiples clases. Se han usado con éxito en una serie de aplicaciones donde se incluyen la detección de rostros, detección de objetos, reconocimiento de dígitos y caracteres manuscritos, categorización de texto, verificación de voz, predicción, etc. (Byun & Lee, 2002).

### 2.2.2 Aprendizaje Profundo

El aprendizaje profundo es un tema que cada vez adquiere mayor relevancia en el campo de la inteligencia artificial. Es considerado una subcategoría del aprendizaje automático. El aprendizaje profundo trata del uso de redes neuronales para mejorar elementos como: el reconocimiento de voz, la visión por ordenador y el

procesamiento del lenguaje natural. Rápidamente se está convirtiendo en uno de los campos más solicitados en informática. En los últimos años, el aprendizaje profundo ha ayudado a lograr avances en áreas tan diversas como la percepción de objetos, la traducción automática y el reconocimiento de voz (todas ellas áreas especialmente complejas para los investigadores en inteligencia artificial).

#### 2.2.2.1 Redes Neuronales Convolucionales

Una red neuronal convolucional es un tipo de red artificial profunda donde las neuronas corresponden a campos receptivos de una manera muy similar a las neuronas en la corteza visual primaria de un cerebro biológico. Debido a que su aplicación es realizada en matrices bidimensionales, son muy efectivas para tareas de visión artificial, como en la clasificación y segmentación de imágenes, entre otras aplicaciones.

Las redes neuronales convolucionales trabajan modelando de forma consecutiva pequeñas piezas de información, y luego combinando esta información en las capas más profundas de la red. Una manera de entenderlas es que la primera capa intentará detectar los bordes y establecer patrones de detección de bordes. Luego, las capas posteriores tratarán de combinarlos en formas más simples y, finalmente, en patrones de las diferentes posiciones de los objetos, iluminación, escalas, etc. Las capas finales intentarán hacer coincidir una imagen de entrada con todos los patrones y arribar a una predicción final como una suma ponderada de todos ellos. De esta forma las redes neuronales convolucionales son capaces de modelar complejas variaciones y comportamientos dando predicciones bastantes precisas. En nuestro estudio serán utilizadas para clasificación.

## 2.3 Audio vs Representación Simbólica

Existen dos enfoques distintos para la representación musical: simbólico y mediante audio (Briot, et al., 2017; Khoo, et al., 2012; Wołkowicz, et al., 2008). Es importante diferenciar entre la representación de datos en modelos de clasificación que funcionan con datos de audio (por ejemplo, archivos *WAV*) y datos simbólicos (por ejemplo, archivos *MIDI*). Los tipos de características que se pueden extraer de un conjunto de datos y se utilizan para construir modelos son muy diferentes para ambas categorías (Herremans, et al., 2016).

Una de las razones por las cuáles se escoge la representación simbólica en este trabajo es por su facilidad de manejo y su parecido a la obra original que el autor concibió. Además, cada intérprete a la hora de tocar música realiza su interpretación personal de la obra que ejecuta, que muchas veces puede diferir bastante de la obra original. Así podemos encontrar disímiles versiones de una misma obra muy diferentes entre sí.

Entre las notaciones de representación simbólica de música se cuentan la notación *Finale*, la *Sibelius*, la *ABC*, la *MusicXML* y el protocolo *MIDI* (Wołkowicz, et al., 2008).

*MIDI* (Interfaz digital de instrumentos musicales) es un estándar industrial que permite la comunicación entre equipos electrónicos, instrumentos musicales y ordenadores (Naccache, et al., 2008). Los archivos *MIDI* contienen información estructurada de muy alto nivel sobre música. Describen la forma específica de realizar una pieza y contienen información como el inicio, la duración, la velocidad y el instrumento (entre otros) de cada nota (Herremans, et al., 2016).

Otra forma de codificación de música son los archivos *.kern*, estos proporcionan una descripción basada en texto de partituras musicales, y su propósito principal es el análisis musical computacional con el uso de la herramienta Humdrum Toolkit.

#### 2.4 Piano roll

Una forma muy popular de realizar la codificación en la música es utilizar *piano roll* (Lyu, et al., 2015; Velarde, et al., 2016). Esta representación es el equivalente en el dominio de la música a la representación *one hot encoding* muy usada en el lenguaje natural. A cada fragmento de la pieza se asigna un vector binario con valor 1 en las posiciones que corresponden a las notas presentes en el fragmento y valor 0 en las demás. El tamaño de este vector corresponde a la cantidad de notas distintas que se consideren, normalmente 128, ya que es la cantidad de notas distintas que soporta el protocolo MIDI. Para mayor información sobre este tema puede consultarse la sección 4.11 de (Briot, et al., 2017).

#### 2.5 Matriz términos-frecuencia. N-gramas

En el Procesamiento de Lenguaje Natural un n-grama es una subsecuencia de  $n$  letras o palabras de una secuencia dada. El n-grama de tamaño 1 se conoce por unigrama, los de tamaño 2 por bigramas, de tamaño 3 por trigramas etc. Este concepto puede llevarse a la escritura musical si establecemos una similitud entre notas o tamaños de nota con letras o palabras. Así una subsecuencia de tres notas do-re-mi resultaría ser un trigramo, una corchea seguida de una semicorchea derivaría en un bigrama etc.

## 2.6 Representación densa

El término “*Word embedding vector*” se ha puesto de moda recientemente en el área de Lenguaje Natural. A diferencia de la representación de textos tradicional conocida por *bolsa de palabras*, en la cual se representa cada texto mediante un vector que representa la frecuencia de aparición de las palabras en dicho texto, en la representación *Word embedding* se asigna a cada palabra un vector continuo y multidimensional de tal forma que las palabras con significado similar o que ocurren con frecuencia en los mismo contextos se encuentran cercanas entre sí y las palabras que no son semánticamente parecidas se encuentran lejanas en el espacio (Mikolov, et al., 2013). Dicha cercanía o lejanía entre vectores se computa usualmente mediante similitud coseno. A menudo a esta representación se le conoce como *densa*, en oposición a representaciones *díspersas* como, por ejemplo, *one hot encoding*.

El estudio realizado en (Mikolov, et al., 2013) representó el auge de la representación *Word embedding*. Los autores presentaron *Word2vec* un método no supervisado para calcular el espacio vectorial denso y probaron que su modelo logra capturar las relaciones semánticas entre palabras. Esta precisamente es la ventaja de la representación *Word embedding* sobre *bolsa de palabras*, ya que en esta última se trata a las palabras como entes sin relación entre sí.

Para el entrenamiento de *Word2Vec* los autores de (Mikolov, et al., 2013) propusieron dos métodos: *Skipgram* y *Continuous bag of words (CBOW)*. En el primero la optimización de la función objetivo se basa en, dada una palabra, tratar de

predecir el contexto, o sea, las palabras que la rodean. *CBOW* se basa en tratar de predecir la palabra central dado el contexto.

Un trabajo interesante es el realizado por los autores de (Jaech, et al., 2016), en lugar de basarse en palabras para construir el espacio vectorial asignaron vectores densos a cada caracter (*Char2vec*) y usaron esta representación para clasificación sobre un corpus de Twitter. Igualmente existen menciones en la literatura sobre representaciones basadas en oraciones (*Sentence2vec*), párrafos (*Paragraph2vec*), e incluso documentos (*Doc2vec*).

## Capítulo 3. Estado del Arte

### 3.1 Antecedentes de estudios sobre Atribución de autoría

La atribución computacional de autoría es una tarea de clasificación en la cual, dada una lista de autores y una lista de obras es necesario asignar cada obra a su autor correspondiente. La idea principal detrás de la atribución de autoría estadísticamente o computacionalmente apoyada es que al medir algunas características textuales, podemos distinguir entre textos escritos por diferentes autores (Stamatatos, 2009).

La atribución de autoría en textos ha sido ampliamente abordada en la literatura (Koppel, et al., 2009; Stamatatos, 2009). Una de las ramas de MIR es la atribución de autoría en la música, también conocida por atribución de composiciones. Puede afirmarse que la atribución de composiciones es al dominio de la música lo que la atribución de autoría es al dominio de lenguaje natural.

Entre los primeros intentos de modelar el estilo de un compositor pueden citarse (Buzzanca, 2001; Buzzanca, 2002; Pollastri & Simoncelli, 2001). En (Buzzanca, 2001) se muestra un sistema experto basado en reglas. En (Pollastri & Simoncelli, 2001) se procesa información melódica de piezas de compositores bien diferentes entre sí (Mozart, Beethoven, Dvorak, Stravinsky y Los Beatles) para clasificar usando Modelos Ocultos de Markov. Los autores argumentan que incluir música pop de Los Beatles al conjunto de piezas “clásicas” ayuda a incrementar la generalidad de sus resultados. Nuestro parecer es que esto facilita el proceso de clasificación ya que resulta más complicada la clasificación entre compositores de un mismo estilo o época. En (Buzzanca, 2002) se usa una red neuronal con propagación hacia atrás para diferenciar entre el compositor Giovanni Palestrina y otros compositores

contemporáneos con este. Se obtienen buenos resultados pero debe mencionarse que la tarea de asignar obras como pertenecientes o no a un compositor parece más sencilla que asignar a cada autor las piezas que le corresponden.

De acuerdo con (Hillewaere, et al., 2009; Kaliakatsos-Papakostas, et al., 2011) los modelos para resolver tareas de clasificación relacionadas con música (en formato simbólico) pueden agruparse en dos grandes categorías. La primera categoría comprende los modelos basados en características globales o descriptores estadísticos los cuales expresan cada pieza como un vector de características, cada característica o descriptor representa la medición de determinado elemento musical a través de toda la pieza, por ejemplo, frecuencia de intervalos de segunda mayor, altura promedio de las notas etc. Entre los trabajos que forman parte de esta categoría pueden mencionarse (Backer & van Kranenburg, 2005; van Kranenburg, 2006). La segunda categoría se muestra en el epígrafe 3.2.

El investigador (van Kranenburg, 2006) utiliza la transformación de Fisher para proyectaren un espacio de bidimensional un catálogo de fugas bien conocidas de J.S. Bach, su hijo W.F. Bach y su alumno más destacado J.L. Krebs, así como algunas otras fugas cuya autoría es cuestionada. De cada pieza se extraen alrededor de 20 características que modelan las relaciones entre las voces de estas composiciones polifónicas. A continuación se reduce cada vector de características mediante la transformación de Fischer y se visualiza el resultado. El autor arriba a interesantes conclusiones sobre varias fugas atribuidas a J.S. Bach y que según el estudio es más probable que pertenezcan a otro compositor.

Los autores de (Velarde, et al., 2016) usan un conjunto de datos compuesto por 107 piezas de Mozart y Haydn. Representan cada obra como una imagen usando la representación denominada piano-roll (ver subepígrafe 2.4). Así transforman cada pieza musical en una imagen en blanco y negro. El eje X representa el tiempo en la pieza y el eje Y representa cada uno de los valores de altura posibles que pueden tomar las notas. Los pixeles (x, y) toman valor 1 (blanco) para representar que en el tiempo x la nota con altura y tiene sonido y toman valor 0 en otro caso. Los investigadores utilizan filtros gaussianos y ondas de Morlet para modificar cada imagen y comparan el desempeño del clasificador Máquina de soporte vectorial sobre el conjunto de imágenes filtradas y sin modificar obteniendo un resultado de 80.4 % de aciertos con su mejor modelo.

Los investigadores (Hillewaere, et al., 2009) utilizaron el mismo conjunto de datos que (van Kranenburg, 2006) y (Velarde, et al., 2016) y se propusieron determinar cuál de los instrumentos que componen un cuarteto de cuerda contribuye más a diferenciar entre estos dos compositores. Para lograr esto crearon cuatro conjuntos de datos aislando cada uno de los instrumentos de cada pieza. Además crearon un quinto conjunto de datos con la concatenación de los cuatro anteriores. De acuerdo a los resultados obtenidos los investigadores reportan que el instrumento portador de la melodía, en este caso el primer violín, es el que mejor caracteriza el estilo de un compositor. Quizá la causa de este hallazgo radica en que los compositores primero desarrollan la melodía y después introducen los demás instrumentos de acompañamiento, en este caso: segundo violín, viola y cello.

### 3.2 Antecedentes de estudios sobre N-gramas

La técnica de n-gramas ha sido ampliamente utilizada para la representación de información musical debido a su sencillez y buen desempeño general. (Downie, 1999; Uitdenbogerd & Zobel, 1999; Pickens, 2000; Doraisamy, 2005; Hillewaere, et al., 2009) utilizan n-gramas sobre intervalos de notas consecutivas en dirección melódica. (Pickens, 2000; Doraisamy, 2005; Hillewaere, et al., 2009) utilizan intervalos sobre la altura de notas y (Uitdenbogerd & Zobel, 1999; Hillewaere, et al., 2009) usan intervalos sobre el contorno de las melodías con tres operaciones posibles a partir de una nota: subir a una nota más alta, bajar a una nota más grave o quedarse en la misma nota. Una parte de los trabajos estudiados consideran solamente piezas monofónicas (Downie, 1999; Pickens, 2000; Hillewaere, et al., 2009; Naccache, et al., 2008) o convierten piezas polifónicas a melodías monofónicas (Uitdenbogerd & Zobel, 1999; Hontanilla, et al., 2013). En trabajos más recientes se permite el manejo de polifonía en las obras (Doraisamy, 2005; Hontanilla, et al., 2013; Wołkowicz, et al., 2008; Pérez-Sancho, et al., 2009; Ogihara & Li, 2008) y se incluye información rítmica de las piezas (Doraisamy, 2005; Hillewaere, et al., 2009; Wołkowicz, et al., 2008; Hontanilla, et al., 2013). En (Pérez-Sancho, et al., 2009; Ogihara & Li, 2008) se presentan métodos para extraer n-gramas a partir de acordes en lugar de notas.

En (Naccache, et al., 2008) se abordan las tareas de detección de autoría, emociones y género en la música. Los autores usan histogramas (n-gramas de tamaño 1) de tonalidad y duración de notas e intervalos entre pares de notas consecutivas, entre otros, para representar piezas musicales en formato MIDI.

Los autores de (Wołkowitz, et al., 2008) realizan una comparación entre elementos básicos de la escritura musical como notas musicales con elementos del lenguaje hablado como palabras y argumentan que es posible aplicar técnicas de procesamiento de lenguaje natural al lenguaje musical. Crean tres tipos de n-gramas basados en altura, ritmo y combinados y utilizan estos para crear el perfil del autor, luego usan una medida de similaridad tomada de la atribución de autoría para establecer las correspondencias entre los perfiles de las obras y los perfiles de los compositores.

Los investigadores (Hillewaere, et al., 2009) comparan características globales como el average de tonalidad de la pieza con lo que ellos llaman características de eventos que en realidad son n-gramas. Demuestran que para el problema de clasificación de canciones europeas de folk es preferible usar características de eventos debido a que estas contienen información secuencial que las características globales no toman en consideración y que para identificar correctamente las canciones pertenecientes a cada región es necesario capturar la estructura interna de las frases musicales de cada canción.

Los investigadores (Hontanilla, et al., 2013) usan n-gramas de tamaño 2, 3 y 4 sobre melodías extraídas de piezas polifónicas tomadas del conjunto de datos propuesto por (Backer & van Kranenburg, 2005) con peores resultados. Sin embargo, obtienen muy buenos resultados al comparar fugas de los compositores Bach y Shostakovich.

### 3.3 Antecedentes de estudios sobre representación densa

Las ideas del procesamiento del lenguaje natural pueden aplicarse al dominio de la música. En este caso los textos equivalen a composiciones, las palabras a acordes y



se produce entre primer y quinto grado de la tonalidad. Además puede apreciarse la cercanía entre cada tonalidad mayor en verde y su relativa menor en fucsia. El modelo es capaz de detectar estas relaciones de manera automática.

En este trabajo consideramos que la misma ventaja que presenta *Word2vec* en relación a *bolsa de palabras* también la presenta *Chord2vec* en relación a *piano roll* y otras representaciones dispersas. Es por eso que proponemos en nuestro sistema usar un modelo parecido (aunque no igual) a *Chord2vec* que proporcione una representación densa de los elementos (acordes, notas etc.) del lenguaje musical para la fase de codificación. De esta forma, podemos verificar si las relaciones “semánticas” que se producen entre estos elementos ayuda o no a mejorar el resultado del proceso de composición.

Los trabajos consultados en la literatura (Huang, et al., 2016; Madjiheurem, et al., 2016; Herremans & Chuan, 2017; Chuan, et al., 2018) se basan en acordes de la misma manera que *Word2vec* se basa en palabras. Consideramos que este enfoque presenta algunas dificultades intrínsecas, por ejemplo la cantidad de acordes que pueden generarse. Si consideramos que un acorde está formado como mínimo por tres notas y que existen 128 notas distintas de acuerdo con el protocolo *MIDI* esto significa que, teóricamente, pueden formarse más de 2 millones de acordes distintos. Esto puede ocasionar que el tamaño del vocabulario sea demasiado extenso en relación al número de tokens de entrenamiento, y puede provocar un entrenamiento deficiente. Debe considerarse además que los corpus de música simbólica son muy reducidos con respecto a los billones de tokens que poseen corpus de entrenamiento para *Word2vec* como Wikipedia o Google News. Nótese también que

solo consideramos en este cálculo acordes formados por 3 notas, si consideramos acordes de 4 o 5 notas distintas, los cuales son comunes, entonces el tamaño del vocabulario se dispara.

En algunos trabajos (Huang, et al., 2016; Chuan, et al., 2018) este problema no se presenta ya que antes de iniciar el entrenamiento se llevan todas las notas que componen el acorde a una misma clase u octava. De esta forma se reduce enormemente el tamaño del vocabulario pero se pierde mucha información valiosa al reducir el rango de las notas. Consideramos que una representación así pudiera actuar en detrimento del proceso de clasificación.

## Capítulo 4. Propuesta de Modelos

En la Figura 19 se presenta una vista general de los procedimientos realizados:

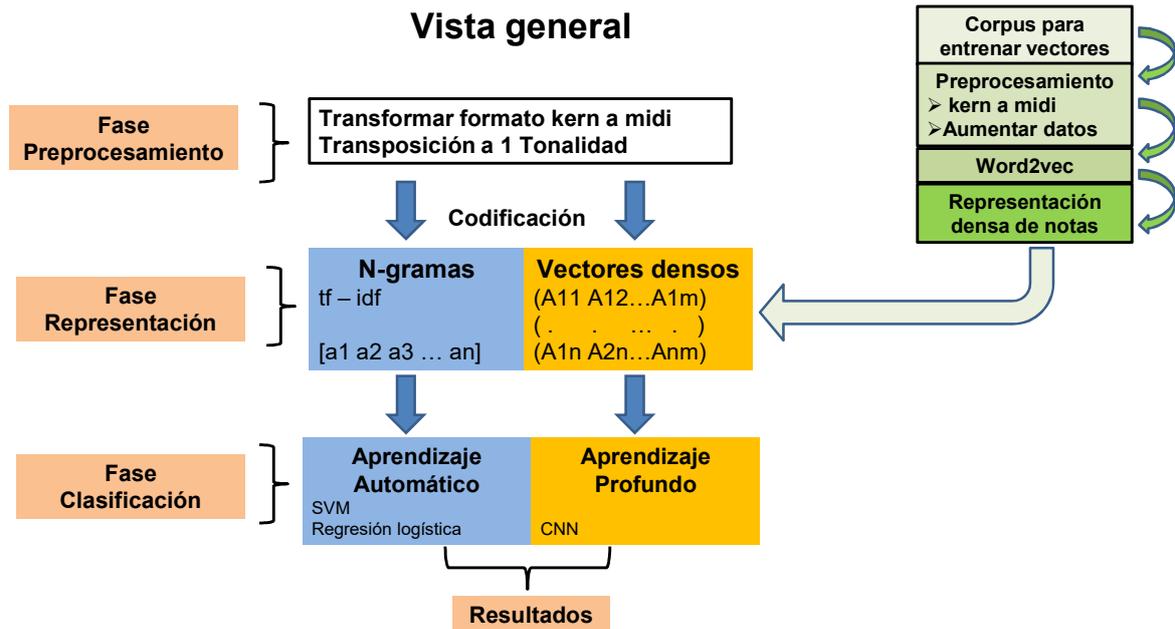


Figura 19. Vista general de procedimientos

### 4.1 Fase Pre-procesamiento

La fase de pre-procesamiento se realiza para todas las obras que serán objeto de análisis. Consta de dos partes. Primero, se extrae de los archivos en formato .kern, la altura de cada nota musical y se convierte la altura de cada nota a su valor entero correspondiente en el formato MIDI. Seguidamente se transponen todas las piezas a una misma tonalidad. En nuestro estudio la transposición se hizo a do mayor coincidiendo con lo realizado por (Kaliakatsos-Papakostas, et al., 2011). Para esto es necesario tener anotada la tonalidad de cada pieza. Debido a las características

del formato musical (cuarteto de cuerdas), es posible encontrar cambios de tonalidad en el interior de la pieza. Hemos sido cuidadosos de tomar en cuenta esos cambios de tonalidad al realizar la transposición. De no considerar este punto se perdería la invarianza de la representación. Como regla general, es necesaria la transposición para que la tonalidad no sea un factor de influencia en la clasificación a la hora de clasificar.

## **4.2 Fase de Representación**

En la fase de representación o codificación se convierten los archivos previamente llevados a MIDI a las dos diferentes codificaciones que proponemos. La representación basada en n-gamas convierte cada uno de los archivos en un vector de características. A su vez, en la representación basada en vectores densos, se codifica cada obra musical como una matriz donde cada fila representa un momento de tiempo de la obra.

### 4.2.1 Representación basada en n-gramas

En los estudios revisados hasta el momento, se ha utilizado previamente la técnica de n-gramas para representar música en notación simbólica. En la mayoría de estos estudios los n-gramas se extraen tomando como base las notas de la línea melódica de cada instrumento o voz<sup>2</sup>. Si comparamos este método con la manera en que las personas escuchan música comúnmente, esto equivaldría a oír cada obra varias veces, centrandó la escucha en un instrumento musical diferente en cada audición.

---

<sup>2</sup> Las voces también se refieren a cada una de las melodías de una obra musical.

En otros estudios se calculan los n-gramas sobre los acordes consecutivos presentes en las piezas. Esto equivale para un ser humano, a escuchar todos los instrumentos o voces simultáneamente. Sin embargo, esta no es la manera usual en que las personas oyen música debido al volumen de la atención requerido para procesar todas las líneas melódicas al mismo tiempo. Normalmente, la atención de los individuos va “saltando” de una voz o instrumento a otro en función de lo que les resulta más interesante. Esto ha sido tomado en cuenta para el proceso de extracción de n-gramas de nuestro estudio.

En nuestro método, se toman como elemento base las notas musicales y se calculan los n-gramas permitiendo saltos de un instrumento a otro (similar a como lo hacen las personas). Esto genera una gran cantidad de n-gramas. Nuestra intuición es que esto pueda ayudar a una representación más exacta de las obras.

#### 4.2.2 Representación basada en vectores densos

##### 4.2.2.1 Pre-entrenamiento de vectores

Es necesario contar con un corpus de música simbólica lo suficientemente grande como para poder entrenar correctamente los vectores densos de cada nota. Hemos utilizado archivos de tres bases de datos distintas: Musedata, Kernscores y Josquin project. En total se cuenta con 4737 archivos en formato .kern de más de 50 autores.

Es necesario especificar que todos los compositores utilizados en las bases de datos pertenecen a épocas anteriores a Haydn y Mozart (cuando más, contemporáneos con ellos). En nuestra investigación, la idea es que los vectores generados con esta

información, representen la base musical de Haydn y Mozart a partir de la cual desarrolló cada uno su propio estilo.

Los archivos son preprocesados de la siguiente manera: se extrae el valor de altura de cada nota y se convierte a su correspondiente valor MIDI y se transpone cada obra a las 12 tonalidades existentes, con el objetivo de que existan más datos para entrenar. Al final se genera un total de 56844 archivos. Para entrenar se utiliza la implementación de Word2vec que proporciona la biblioteca gensim de python. Se generan dos representaciones basadas en notas: MelodicPitch2vec y HarmonicPitch2vec, las cuales explicamos a continuación.

#### Melodic Pitch2Vec

Es una representación de notas en un espacio vectorial n-dimensional. Su objetivo es modelar relaciones entre notas siguiendo el eje tiempo sobre una misma voz o instrumento. El resultado esperado es que notas que ocurran frecuente y consecutivamente en dirección melódica resulten cercanas en el espacio.

#### Harmonic Pitch2Vec

Es una representación de notas en un espacio vectorial n-dimensional. Su objetivo es modelar relaciones entre notas siguiendo el eje espacio sobre todas las voces o instrumentos que confluyen en los mismos instantes de tiempo. El resultado esperado es que notas que ocurran frecuente y consecutivamente en los mismos acordes, resulten cercanas en el espacio.

De esta forma se planea resolver los problemas planteados acerca de las representaciones basadas en acordes ya que el vocabulario que se genere según el

protocolo *MIDI* nunca será mayor de 128 y es posible representar un acorde como combinación de notas en cualquier octava.

### Codificación

Una vez entrenados los vectores densos podemos utilizarlos para representar las piezas de los compositores que queremos clasificar, en este caso, Haydn y Mozart. Para ello se sustituye cada nota en la pieza por su correspondiente vector, ya sea melódico o armónico.

#### 4.3 Fase de Clasificación

Al igual que (van Kranenburg, 2006) se utilizó una validación cruzada de tipo leave one out. Este tipo de validación suele utilizarse cuando trabaja con conjuntos de datos pequeños. Para realizar esta validación se corren  $n$  modelos, donde  $n$  es la cantidad de casos en el conjunto de datos. En cada iteración se deja un caso para prueba y el resto para entrenamiento. El resultado final es el promedio de aciertos totales entre todas las iteraciones.

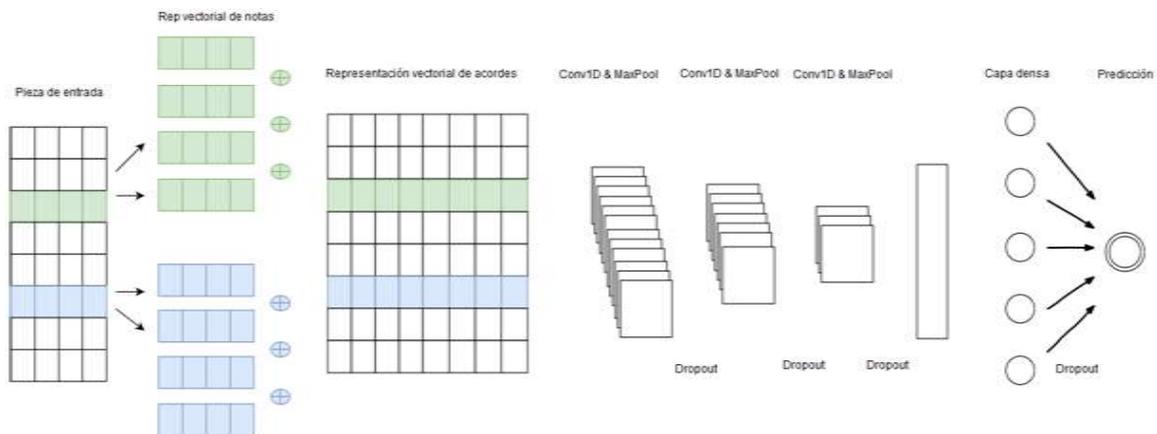
##### 4.3.1 Modelos basados en representación de n-gramas

Una vez que tenemos representada cada obra como un vector, es posible pasar a la fase de clasificación. Se usó la implementación de la biblioteca sklearn de python de los clasificadores: máquinas de soporte vectorial y regresión logística. Se usaron los parámetros por defecto de los clasificadores en todos los casos.

##### 4.3.2 Modelos basados en representación densa

Una vez entrenados los vectores densos, pueden ser utilizados para crear una representación matricial de las obras de los autores que queremos clasificar. Luego de que tenemos representada cada obra como una matriz, es posible pasar a la fase de clasificación. Se usó la implementación de red neuronal convolucional que proporciona la biblioteca keras de python.

En la siguiente figura se muestra gráficamente el modelo de red convolucional. La entrada a la red es una matriz donde cada fila representa un fragmento de tiempo de la obra y cada columna un instrumento con sus correspondientes notas. Cada nota se sustituye por su vector denso (melódico o armónico) y se crea una representación para cada momento de la obra combinando los vectores de cada nota en uno solo. La combinación de vectores puede realizarse mediante concatenación, promedio u otras operaciones encontradas en la literatura. En nuestro caso, usamos la operación de concatenación aprovechando que las piezas de nuestro conjunto de datos tienen un número regular de instrumentos. Se recomienda utilizar otra operación como por ejemplo, la concatenación cuando en nuestro conjunto de datos existen obras con distinto número de instrumentos. Por ejemplo, si queremos comparar un solo de guitarra con una orquesta.



**Figura 20. Modelo de red convolucional utilizado en este trabajo**

Los restantes componentes de la red son tres capas de filtros convolucionales en una dimensión, cada uno de los cuales está seguido por una capa de max pooling y dropout. A esto se le añade una capa densa donde se realiza la clasificación y una capa final con dos valores posibles de salida.

## Capítulo 5. Resultados y Discusión

### 5.1 Modelos basados en n-gramas

Una vez obtenida la representación vectorial de cada pieza mediante las combinaciones de n-gramas explicadas en el epígrafe 4.2.1, se procede a realizar la clasificación. Realizamos experimentos con un rango de n-gramas entre 2 y 5. Probamos la implementación de los clasificadores naive Bayes, regresión logística, árbol de decisión, random forest y máquinas de soporte vectorial que proporciona la biblioteca sklearn de python. Los dos clasificadores de mejores resultados se muestran a continuación. El mejor resultado se logra usando regresión logística con n-gramas de tamaño 4.

Clasificadores	Orden de n-grama	Accuracy
SVM Linear	2	71.9
	3	71.9
	4	73.8
	5	72.8
Regresión logística	2	68.2
	3	72.8
	4	<b>74.7</b>
	5	71.0

**Tabla 5. Comparación de clasificadores basados en n-gramas**

Se observó una mejora en el desempeño de estos algoritmos después de aplicar escalado de máximo absoluto proporcionado por la biblioteca Sklearn de python. Al comparar con el estado del arte, nuestro mejor método de regresión logística (Ver tabla 6) queda un poco alejado de los mejores valores. Al parecer, el hecho de usar combinaciones que producen gran cantidad de n-gramas, generan características que añaden poco valor al modelo de vectores y más bien entorpecen el proceso de clasificación.

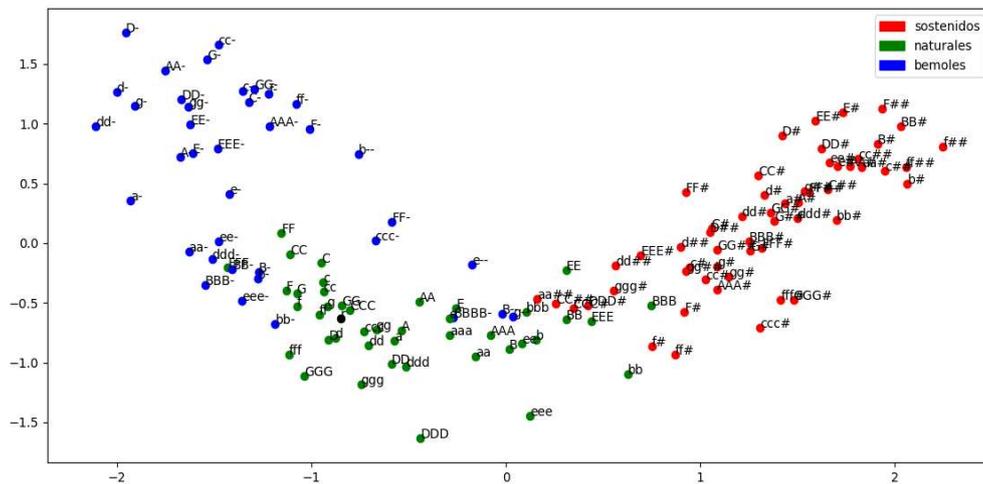
Método	Accuracy
(Lostanlen, 2018)	82.2
(Velarde, et al., 2016)	80.4
(Backer & van Kranenburg, 2005)	79.4
(Velarde, et al., 2018)	79.4
<b>Nuestro método RL</b>	74.7
(Kempfert & Wong, 2018)	<b>85.2</b>
(Herlands, et al., 2014)	80.0
(Hillewaere, et al., 2010)	75.4
(Hontanilla, et al., 2013)	74.7

**Tabla 6. Comparación de modelo basado en n-gramas con estado del arte (en los cuatro últimos trabajos se usó un conjunto de datos diferente)**

Nuestra recomendación es utilizar un método de extracción de las características más relevantes, previo a clasificar. Como trabajo futuro queda verificar si la adición de información rítmica a los n-gramas puede contribuir a elevar la calidad de las características generadas.

## 5.2 Visualización de vectores densos

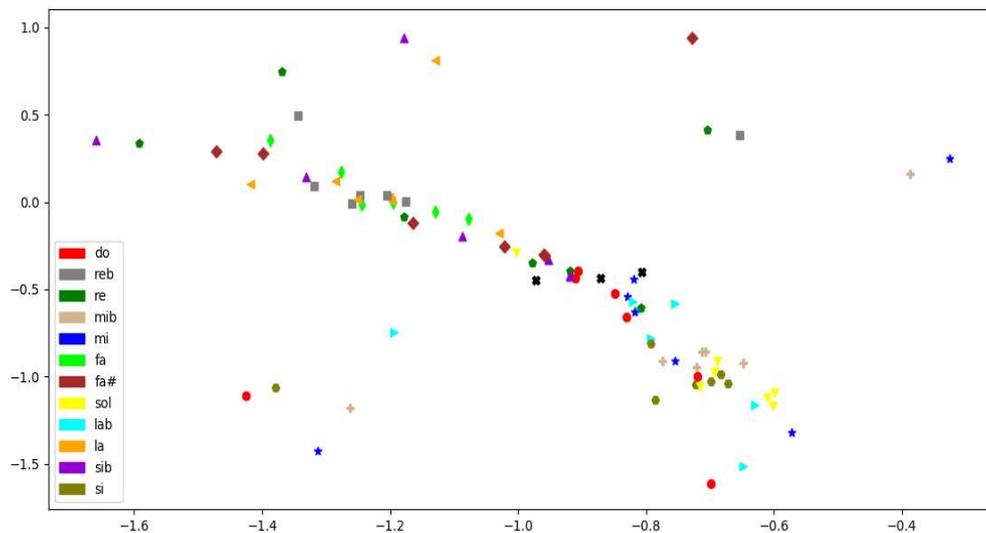
Para verificar que la propuesta de representación densa logra capturar similitudes melódicas y armónicas, hemos entrenado un modelo de vectores “de juguete” de sólo dos dimensiones que pueda ser visualizado gráficamente. En otros casos se utiliza Análisis de componentes principales (Principal Components Analysis PCA) para reducir la dimensionalidad de los vectores más extensos. Todas estas visualizaciones con su correspondiente análisis se muestran a continuación.



**Figura 21. Visualización de vectores densos. Alteraciones**

En la figura 21 se muestra una visualización del modelo word2vec entrenado usando skip gram sobre las notas en formato kern en dirección armónica. Puede observarse que las notas naturales quedan en el centro del gráfico y las notas con alteraciones se encuentran a ambos costados. Esto es debido a que es poco común que aparezcan en la música escrita, notas con sostenidos y bemoles sonando al unísono. Otra característica que resalta es que las notas de una misma clase se

agrupan juntas, por ejemplo Do (C) y La (A). En menor medida se observa la relación armónica de intervalo de quinta justa. Por ejemplo, si nos situamos en Do (C) hacia la izquierda, encontramos muy cerca Fa (F), luego Si bemol (Bb), Mi bemol (Eb), etc. Y hacia la derecha encontramos Sol (G), Re (D), La (A), etc.

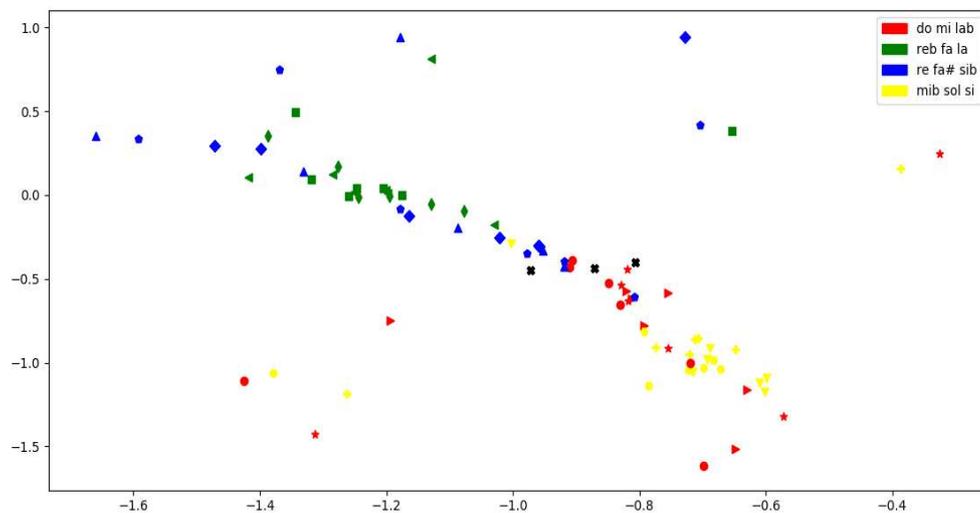


**Figura 22. Visualización de vectores densos. Relación de octavas**

En los restantes gráficos que explicaremos a continuación se ha eliminado la distinción entre bemoles y sostenidos ya que para el problema de atribución de composiciones no es requerida. Se denota el Do sostenido como Re bemol, el Re sostenido como Mi bemol y así sucesivamente.

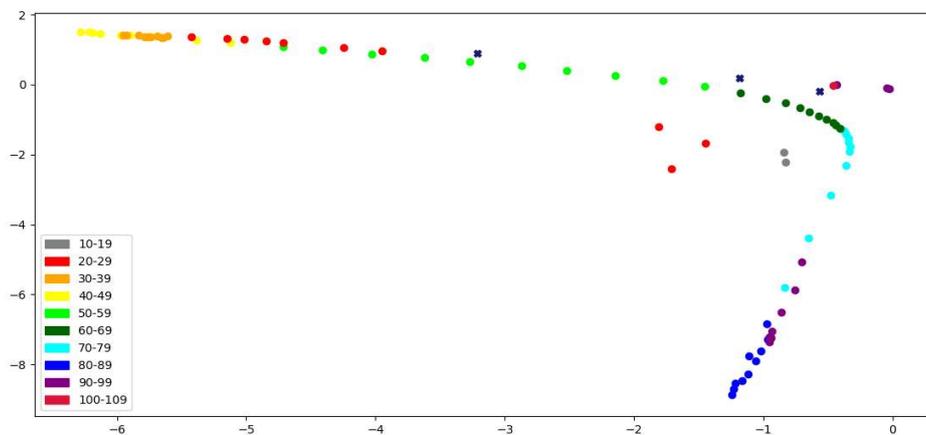
En la figura 22 se han unido las notas por clase para representar la relación armónica de octavas que se produce en el modelo. Las mismas notas en diferentes octavas quedan cercanas entre sí, debido a que están presentes en los mismos contextos (cuando simultáneamente un instrumento o voz, da una octava grave y otro una octava aguda). La relación de octavas puede visualizarse en este gráfico

con mucha fuerza para el caso de las notas Sol (G), Mi bemol (E), Si (B), Re bemol (D). Y en menor medida para el caso de Fa (F), Do (C), La (A), etc. Es interesante notar que vuelven a aparecer algunas relaciones de quinta, por ejemplo Do, (que puede encontrarse entre Fa y Sol), Re (que se encuentra entre Sol y La) y Mi (que se sitúa entre La y Si).



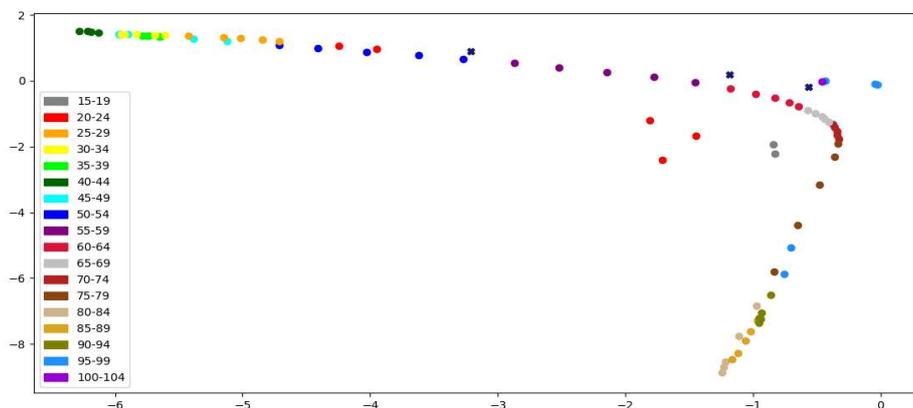
**Figura 23. Visualización de vectores densos. Relación de terceras y sextas**

En la figura 23 se muestra el mismo modelo anterior, pero resaltando las relaciones armónicas de tercera y sexta. Para esto se han agrupado las notas en cuatro clases: la clase de Do, la clase de Re bemol y de Mi bemol. En cada clase existe una relación ya sea de tercera o sexta entre las notas que la componen. Se observa que estas relaciones son muy fuertes ya que puede notarse que los miembros de cada clase por lo general se agrupan en una región específica del espacio. Esto sucede debido a que el intervalo de tercera mayor ocurre en la mayoría de las tríadas comunes a la música occidental (tríadas mayor, menor y aumentada).



**Figura 24. Visualización de vectores densos en dirección melódica (I)**

En la figura 24, se puede visualizar las relaciones melódicas. Las notas se agrupan por su altura en valor entero (notas más graves, valor menor) (notas más agudas, valor mayor). Se observa que las notas se agrupan de manera evidente y que las notas graves quedan hacia un lado y las agudas hacia otro. La causa de este fenómeno es que en dirección melódica son poco frecuentes los intervalos mayores de octava. En la figura 25 se observan las mismas relaciones melódicas pero dividiendo las notas por altura de cinco en cinco.



**Figura 25. Visualización de vectores densos en dirección melódica (II)**

### 5.3 Uso de vectores densos mediante aprendizaje profundo

En el proceso de entrenamiento probamos diferentes configuraciones para la red convolucional. Se experimentó con el número de capas convolucionales y pooling, diferentes valores de dropout y número de neuronas en cada capa. También se experimentó con operaciones de global pooling y diferentes tamaños de filtros actuando simultáneamente. En algunos experimentos reservamos un 10% del conjunto de entrenamiento para validación quedándonos en cada fold con los modelos de menor error de validación y mayor accuracy de validación para realizar la predicción sobre el ejemplo de test. Esto no funcionó al parecer debido al pequeño tamaño del conjunto de datos.

Se probaron ambos modelos de vectores densos logrando el mejor resultado con el modelo que porta información melódica. La configuración final utilizada está compuesta por 3 capas de 32 filtros convolucionales de 1 dimensión con tamaño 5, cada una de estas capas está seguida por una capa de max pooling de tamaño 5, se añade dropout con valor de 0.25 después de cada capa de max pooling; esto genera un vector de representación de tamaño 600 aproximadamente el cuál se clasifica en una capa densa de 32 neuronas, la capa de salida está compuesta por 1 neurona con activación sigmoide. Se utiliza el optimizador adam para el cálculo del gradiente y entropía cruzada binaria para el cálculo del error. En la tabla 7 se muestra este resultado en comparación con el estado del arte. Se observa que nuestro modelo obtuvo el mejor resultado para el conjunto de datos compuesto por 107 piezas de Haydn y Mozart. Es necesario ejecutar nuestro modelo sobre el conjunto de datos utilizado por (Kempfert & Wong, 2018) para poder comparar la efectividad de nuestra

propuesta con la suya. El hecho de que el conjunto de datos utilizado por (Kempfert & Wong, 2018) es mayor (285 piezas en total) puede ayudar a que nuestro modelo de aprendizaje profundo pueda tener mejores resultados ya que los modelos de aprendizaje profundo por lo general tienen mejor desempeño cuando son entrenados con un mayor número de datos.

Método	Accuracy
<b>Nuestro modelo CNN</b>	84.1
(Lostanlen, 2018)	82.2
(Velarde, et al., 2016)	80.4
(Backer & van Kranenburg, 2005)	79.4
(Velarde, et al., 2018)	79.4
(Kempfert & Wong, 2018)	<b>85.2</b>
(Herlands, et al., 2014)	80.0
(Hillewaere, et al., 2010)	75.4
(Hontanilla, et al., 2013)	74.7

**Tabla 7. Comparación del modelo basado en representación densa con el estado del arte (en los cuatro últimos trabajos se usó un conjunto de datos diferente)**

## **Conclusiones y Recomendaciones**

La atribución de composiciones es un problema de clasificación en el cual, dada una lista de compositores y una lista de obras musicales es necesario asignar cada obra a su correspondiente compositor. En el presente trabajo se modela el problema de atribución de composiciones mediante el uso de dos representaciones tomadas del dominio del lenguaje natural.

Se presenta un método novedoso para realizar la extracción de n-gramas a partir de piezas musicales en formato simbólico tomando como base la hipótesis de que las personas escuchan la música concentrando su atención en diferentes instrumentos o voces a medida que transcurre el tiempo de la pieza. A partir de los n-gramas extraídos se construye una representación vectorial de cada pieza y se utilizan clasificadores populares en la literatura como máquinas de soporte vectorial y regresión logística para completar el proceso de clasificación.

Se adapta el algoritmo Word2vec para lograr una representación densa de notas musicales y se usa esta representación como entrada a una red neuronal convolucional. De acuerdo a la literatura revisada esto representa la aplicación por vez primera de aprendizaje profundo al problema de atribución de composiciones. Mediante el uso de visualizaciones se muestra que la representación densa de notas es capaz de atrapar relaciones armónicas y melódicas propias de la teoría de composición musical.

Se comparan los resultados de ambos modelos de clasificación utilizando un corpus de cuartetos de cuerda de Mozart y Haydn, los cuales tienen estilos de composición muy similares. Se evidencia que la diferencia de desempeño entre ambos modelos

es notable con una diferencia de 10 puntos porcentuales de piezas correctamente clasificadas en favor del modelo basado en aprendizaje profundo. Se muestra que el modelo basado en aprendizaje profundo representa una mejora respecto al estado del arte confirmando la hipótesis de que las relaciones armónicas y melódicas presentes en los vectores densos contribuyen a una mejor clasificación.

Como trabajo futuro se recomienda incluir información rítmica en ambos modelos y aplicar una técnica de reducción de dimensionalidad a los vectores generados mediante n-gramas para verificar si esto puede contribuir a una mejora en el proceso de clasificación. Además recomendamos aplicar los modelos aquí presentados a otros problemas de clasificación para validar la eficacia de los mismos en otros conjuntos de datos, así como aplicar la representación densa de notas al problema de generación automática de música con el objetivo de verificar si las relaciones armónicas y melódicas presentes en esta representación pueden contribuir a generar piezas musicales más similares a las de un compositor real.

## Referencias bibliográficas

- Backer, E. & van Kranenburg, P., 2005. On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, 26(3), pp. 299-309.
- Boser, B. E., Guyon, I. M. & Vapnik, V. N., 1992. *A training algorithm for optimal margin classifiers.*
- Boser, B. E., Guyon, I. M. & Vapnik, V. N., 1992. *Support Vector Machines.*
- Briot, J.-P., Hadjeres, G. & Pachet, F., 2017. *Deep learning techniques for music generation-a survey.*:arXiv.
- Buzzanca, G., 2001. *A Rule-Based Expert System for Musical Style Recognition.*
- Buzzanca, G., 2002. *A supervised learning approach to musical style recognition.* p.167.
- Byun, H. & Lee, S.-W., 2002. *Applications of support vector machines for pattern recognition: A survey.* Berlin, Heidelberg, Springer, pp. 213-236.
- Chuan, C.-H., Agres, K. & Herremans, D., 2018. From context to concept: exploring semantic relationships in music with word2vec. *Neural Computing and Applications*, pp. 1-14.
- Chuan, C.-H. & Herremans, D., 2018. *Modeling temporal tonal relations in polyphonic music through deep networks with a novel image-based representation.* pp. 2159-2166.
- Cortes, C. & Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), pp. 273-297.

- Costa, Y. M., Oliveira, L. S. & Silla, C. N., 2017. An evaluation of Convolutional Neural Networks for music classification using spectrograms. *Applied Soft Computing*, Volumen 52, pp. 28-38.
- Doraisamy, S., 2005. Polyphonic music retrieval: the n-gram approach. *ACM SIGIR Forum*, 39(1), pp. 58-58.
- Downie, J. S., 1999. *Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text..* London, Ont.: Faculty of Graduate Studies, University of Western Ontario.
- Herlands, W., Der, R., Greenberg, Y. & Levin, S., 2014. *A machine learning approach to musically meaningful homogeneous style classification.* AAAI Press, pp. 276-282.
- Herremans, D. & Chuan, C.-H., 2017. *Modeling Musical Context With Word2Vec.* pp.11-18.
- Herremans, D., Martens, D. & Sörensen, K., 2016. Composer classification models for music-theory building. En: *Computational Music Analysis.* Springer, pp. 369-392.
- Herremans, D., Sörensen, K. & Martens, D., 2015. Classification and Generation of Composer-Specific Music Using Global Feature Models and Variable Neighborhood Search. *Computer Music Journal*, 39(3), pp. 71-91.
- Hillewaere, R., Manderick, B. & Conklin, D., 2009. *Global Feature Versus Event Models for Folk Song Classification.*
- Hillewaere, R., Manderick, B. & Conklin, D., 2009. *Melodic models for polyphonic music classification.*
- Hillewaere, R., Manderick, B. & Conklin, D., 2010. *String Quartet Classification with Monophonic Models.* pp. 537-542.

- Hontanilla, M., Pérez-Sancho, C. & Inesta, J. M., 2013. *Modeling musical style with language models for composer recognition*. Springer, Berlin, Heidelberg, pp. 740-748.
- Hsu, C.-W. & Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2), pp. 415-425.
- Huang, C.-Z. A., Duvenaud, D. & Gajos, K. Z., 2016. *Chordripple: Recommending chords to help novice composers go beyond the ordinary*. ACM, pp. 241-250.
- Jaech, A. y otros, 2016. *Hierarchical Character-Word Models for Language Identification*. pp. 84-93.
- Kaliakatsos-Papakostas, M. A., Epitropakis, M. G. & Vrahatis, M. N., 2011. *Weighted Markov Chain model for musical composer identification*. pp. 334-343.
- Kempfert, K. C. & Wong, S. W., 2018. *Where Does Haydn End and Mozart Begin? Composer Classification of String Quartets*. arXiv preprint arXiv: 1809.05075.
- Kešelj, V., Peng, F., Cercone, N. & Thomas, C., 2003. *N-gram-based author profiles for authorship attribution*. pp. 255-264.
- Khoo, S., Man, Z. & Cao, Z., 2012. *Automatic han chinese folk song classification using extreme learning machines*. Springer, Berlin, Heidelberg, pp. 49-60.
- Koppel, M., Schler, J. & Argamon, S., 2009. Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, 60(1), pp. 9-26.
- Kotsiantis, S. B., Zaharakis, I. & Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. En: *Emerging artificial intelligence applications in computer engineering*. pp. 3-24.

- Lostanlen, V., 2018. *Eigentriads and Eigenprogressions on the Tonnetz*. arXiv preprint arXiv: 1810.00790.
- Lyu, Q., Wu, Z. & Zhu, J., 2015. *Polyphonic Music Modelling with LSTM-RTRBM*. ACM, pp. 991-994.
- Madjiheurem, S., Qu, L. & Walder, C., 2016. *Chord2Vec: Learning musical chord embeddings*. pp. 1-5.
- Mikolov, T. y otros, 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, pp. 3111-3119.
- Naccache, M., Borgi, A. & Ghédira, K., 2008. *A Learning-Based Model for Musical Data Representation Using Histograms*. Springer, Berlin, Heidelberg, pp. 207-215.
- Ogihara, M. & Li, T., 2008. *N-Gram Chord Profiles for Composer Style Representation*. pp. 671-676.
- Pérez-Sancho, C., Rizo, D. & Inesta, J. M., 2009. Genre classification using chords and stochastic language models.. *Connection science*, 21(2-3), pp. 145-159.
- Pickens, J., 2000. *Pickens, J. (2000, October). A Comparison of Language Modeling and Probabilistic Text Information Retrieval Approaches to Monophonic Music Retrieval*.
- Pollastri, E. & Simoncelli, G., 2001. *Classification of melodies by composer with hidden Markov models*. pp. 88-95.
- Robine, M., Hanna, P., Ferraro, P. & Allali, J., 2007. *Adaptation of string matching algorithms for identification of near-duplicate music documents*. pp. 37-43.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J., 1986. Learning Internal Representations by Error Propagation. En: *Parallel Distributed Processing*.

- Stamatatos, E., 2009. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3), pp. 538-556.
- Uitdenbogerd, A. & Zobel, J., 1999. *Melodic matching techniques for large music databases*. ACM, pp. 57-66.
- van Kranenburg, P., 2006. *Composer attribution by quantifying compositional strategies*. pp. 375-376.
- Velarde, G. y otros, 2016. *Composer Recognition Based on 2D-Filtered Piano-Rolls*. pp. 115-121.
- Velardo, V., Vallati, M. & Jan, S., 2016. Symbolic melodic similarity: State of the art and future challenges. *Computer Music Journal*, 40(2), pp. 70-83.
- Wołkowicz, J., Kulka, Z. & Kešelj, V., 2008. N-gram-based approach to composer recognition. *Archives of Acoustics*, 33(1), pp. 43-55.