

Trabajo Fin de Grado
Grado en Ingeniería de Tecnologías Industriales

Un estudio de identificación por tarareo para cante flamenco

Autor: Paula Serna García

Tutor: Jose Miguel Díaz Báñez



Trabajo Fin de Grado
Grado en Ingeniería de Tecnologías Industriales

Un estudio de identificación por tarareo para cante flamenco

Autor:

Paula Serna García

Tutor:

Jose Miguel Díaz Báñez

Catedrático de Universidad

Dpto. Matemática Aplicada II
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2019

Trabajo Fin de Grado: Un estudio de identificación por tarareo para cante flamenco

Autor: Paula Serna García

Tutor: Jose Miguel Díaz Báñez

El tribunal nombrado para juzgar el Trabajo arriba indicado, compuesto por los siguientes miembros:

Presidente:

Vocales:

Secretario:

Acuerdan otorgarle la calificación de:

Sevilla, 2019

El Secretario del Tribunal

*A mi madre, por hacer de puente
ente una generación que no sabe
leer ni escribir, a una ingeniería*

*A mi abuelo, por su sabiduría y
educación emocional, sin mentar
una sola palabra, con su ser y
estar*

A mi abuela, por sus orígenes

Agradecimientos

La realización de este trabajo no hubiera sido posible sin la ayuda, paciencia y dedicación que el profesor doctor José Miguel Díaz Báñez ha depositado en él, por este motivo quisiera aprovechar estas líneas para expresarle mis agradecimientos por darme la posibilidad de realizar este proyecto con él.

Agradecer también a Nadine Kroher, por la colaboración y sus conocimientos de gran ayuda en todo el proceso.

Inmensamente agradecida a mi madre, Amalia García Pantoja y a mi familia, por su apoyo incondicional y por haber creído en mí y en mi futuro, animándome en todos aquellos momentos que lo he necesitado.

A José Gallardo y profesionales que han participado en este proyecto, por su tiempo, su predisposición, y por mostrarme su pasión por el flamenco.

Y, por último, a mis amigas/os, compañeras/os y toda aquella persona que se haya cruzado en mi camino, y que me haya tratado con amor como motor de mi avance y evolución.

Paula Serna García

Sevilla, 2019

Resumen

El flamenco como entidad musical tiene su base en la voz cantada, llamada “cante” en el argot flamenco, donde predomina una ornamentación barroca que suele ser improvisada. Esto provoca una serie de retos tecnológicos a la hora de automatizar las tareas de estudio. En el presente trabajo se aborda el diseño e implementación de un sistema de identificación de melodías mediante consultas realizadas sobre grabaciones de cante flamenco obtenidas en un trabajo de campo realizado con cantaores en Sevilla y Jerez. Se propone una estrategia que usa un algoritmo de similitud melódica basado en el método de Needleman-Wunsch. Los resultados obtenidos muestran la competencia del método que resulta ser el primero que aborda la tarea de *query by humming* para cante flamenco.

Abstract

Flamenco as a musical entity is based on the singing voice, called “cante” in the flamenco argot, where a baroque ornamentation predominates and it is usually the improvised part of the performance. This musical aesthetics causes a series of technological challenges. This paper deals with the design and implementation of a query by humming method for flamenco singing using a data base with recordings obtained in a field work carried out with singers in Seville and Jerez. A strategy using a new melodic similarity algorithm based on the Needleman-Wunsch method is proposed. We discuss the results obtained with our approach, the first one that addresses the query by humming problem for flamenco singing.

Agradecimientos	ix
Resumen	xi
Abstract	xiii
Índice	xv
Índice de Tablas	xvii
Índice de Figuras	xix
1 Introducción	1
1.1 <i>Tecnología musical</i>	1
1.2 <i>Identificación por tarareo</i>	5
1.2.1 Dificultades	6
1.2.2 Trabajos previos	7
1.3 <i>Motivación</i>	8
1.3.1 El proyecto Cofla	10
1.4 <i>Objetivos</i>	12
2 Identificación por tarareo para flamenco	17
2.1 <i>El cante flamenco, sus clasificaciones y características</i>	17
2.2 <i>Metodología</i>	20
2.3 <i>El corpus Cofla</i>	21
3 Transcripción automática	23
4 Función de similitud	27
4.1 <i>Algoritmo Needleman Wunsch</i>	27
4.1.1 Needleman Wunsch para dos secuencias	27
4.1.2 Needleman-Wunsch para similitudes melódicas	29
5 Base de datos de consultas	35
6 Experimentos y discusión de los resultados computacionales	39
7 Conclusiones	47
7.1 <i>Trabajo futuro</i>	47
Referencias	49
Anexos	53
<i>Anexo I: Términos musicales</i>	53
<i>Anexo II: Biografía de los participantes en las consultas realizadas</i>	55

ÍNDICE DE TABLAS

Tabla 1. Resultado de la mejor ruta de alineación del patrón de consulta realizado por Eva Rubichi en la transcripción automática de una Seguiriya de Manuel de Torres	35
Tabla 2. Recopilación de los datos de las consultas	37
Tabla 3. Medida de los aciertos sobre las 120 consultas	41
Tabla 4. Agrupación de los participantes según estudios y sus rendimientos en las 37 consultas	42
Tabla 5. Agrupación de los participantes según acompañamiento y sus aciertos en posición top 1 en la muestra	42
Tabla 6. Ranking de los participantes con más peso dentro del porcentaje de aciertos top 1	43
Tabla 7. Ranking de los estilos de cante flamenco que componen los aciertos top 1	44
Tabla 8. Porcentajes de aciertos de los participantes según la aportación en la muestra	45
Tabla 9. Rendimientos de los sujetos conforme con sus colaboraciones en la muestra	46
Tabla 10. Canciones pertenecientes a la base de datos que el algoritmo extrae como resultados de la identificación de la consulta grabada número 62	47

ÍNDICE DE FIGURAS

Figura 1. Arquitectura básica de un sistema de consulta mediante tarareo	5
Figura 2. Notación numérica de los patrones rítmicos ternarios del flamenco	20
Figura 3. Escala modo Frifio de Mi	20
Figura 4. Carátula del disco "100 AÑOS DE FLAMENCO" contenido en la base de datos con la que se compara	23
Figura 5. Estructura del modelo de transcripción automática	26
Figura 6. Matriz de similitud S cuyas celdas contienen la puntuación obtenida utilizando la ecuación (1)	29
Figura 7. Ruta de alineación óptima de dos secuencias con el algoritmo Needleman-Wunsch	30
Figura 8. Transición no adyacente desde $S(i+2, j+1)$ hasta $S(i, j)$	33
Figura 9. Transición adyacente desde $S(i+1, j+1)$ hasta $S(i, j)$	33
Figura 10. Ejemplo de detección de un patrón de consulta en una transición automática de una canción perteneciente a la base de datos	34
Figura 11. Ejemplo de los datos MIDI del contorno melódico que genera la transcripción automática de un Tango	39
Figura 12. Distribución de los porcentajes de participantes en los aciertos top 1	43
Figura 13. Distribución de los diferentes estilos de canto flamenco en los aciertos top 1	44
Figura 14. Comparativa de los rendimientos individuales de los sujetos según su participación con el rendimiento del sistema (η)	45
Figura 15. Contraste entre los rendimientos de los estilos según la contribución en la muestra y el rendimiento general (η)	46
Figura 16. Pentagrama indicador de algunos conceptos musicales	57

1 INTRODUCCIÓN

En la actualidad, se tiene acceso a una gran cantidad de información a través de Internet, que permite enviar, compartir y descargar todo tipo de contenido con el uso de ordenadores personales, smartphones y Smart TV. El volumen de los contenidos multimedia aumenta y la necesidad de almacenarlos en bases de datos especializadas surge, así como nuevos métodos de recuperación.

En los sistemas de recuperación de información, el objetivo es lograr que cualquier tipo de usuario encuentre la información que busca de la manera más fácil e intuitiva. Uno de los sistemas más tradicionales y conocidos, es el uso de Google, que recupera la información mediante consultas de palabras claves.

En los sistemas de recuperación de información multimedia se han producido grandes avances que permiten recuperar la información con nuevas técnicas más intuitivas. Existen sistemas que permiten identificar canciones mediante contenido musical e incluso identificar similitudes entre contenidos musicales para ofrecer al usuario información que se adecua a sus gustos, como es el caso de la aplicación Shazam, sin embargo, esta *app* no identifica melodías.

La similitud de la música es un aspecto fundamental en el desarrollo de sistemas de recuperación de la información musical (MIR), así como desarrollar métodos automáticos para cuantificar la similitud de la música abordando parte del problema de dar sentido a la información digital.

La similitud de la música es, sin embargo, un término ambiguo, ya que involucra diferentes facetas musicales como la instrumentación, tonalidad o ritmo, y también depende de lo cultural. Hay muchos factores involucrados en los juicios de similitud de la música, y algunos de ellos son difíciles de medir.

Una tarea muy relacionada con la recuperación de la información multimedia mediante contenido musical, y que aborda la similitud entre este contenido, es la identificación de canciones a partir de melodía cantada, tarareada o silbada. Su aplicación puede ser útil tanto para los desafíos que presenta la automatización de análisis computacionales de redes de influencia musical, como por su atractiva aplicación para usuarios finales. La identificación o consulta por tarareo es una solución natural al problema de reconocimiento de piezas desconocidas.

1.1 Tecnología musical

Las tecnologías para el análisis de contenido musical tuvieron su aparición a finales de los años 90, en respuesta a la necesidad de conocer, organizar y explotar la información musical comprendida en un archivo sonoro, y como objetivo a la automatización de los procesos de descripción y explotación de tales contenidos, creándose desde entonces una comunidad definida por investigadores, desarrolladores de software y empresas, que tienen este objetivo común. Su evolución se debe, en gran parte, a la disposición que la comunidad ha tenido a grandes colecciones de archivos sonoros musicales.

Décadas atrás, el desarrollo de dichas tecnologías era impensable, ya que el acceso al contenido musical requería la intervención humana para anotar, de manera estructurada dicho contenido. Las grandes colecciones de archivos sonoros musicales eran exclusivamente patrimonio de instituciones culturales y educativas, y de empresas comerciales. Además, el interés por acceder a la información del contenido musical solamente formaba parte de musicólogos, bibliotecarios, compositores, intérpretes, aficionados a la música, etc.

La comunidad tecnológica trabaja con un contenido musical definido como la información referente a una obra musical que está representada en ella misma. Se analiza una señal sonora extrayéndole una serie de descriptores de la pieza, que definen el contenido como una representación de la obra. Algunos de estos descriptores incluyen la estructura, la melodía, el ritmo, la instrumentación y la armonía, incluso conceptos como la emoción que tiende a transmitir, la energía o el género musical. Dentro de esta comunidad hay numerosas disciplinas involucradas, tales como, procesado de señal, musicología psicoacústica y cognición musical, informática, ciencias de la información y de la computación, etc.

El objetivo de las tecnologías de procesado del contenido musical es el de describir automáticamente el contenido de señales de audio musicales y proporcionar medios para editarlo, visualizarlo y explotarlo con fines diversos mediante herramientas computacionales. Aunque una parte del contenido puede etiquetarse de forma manual, como, por ejemplo, el año de composición, el lugar de grabación o el nombre del autor.

Actualmente, las tecnologías permiten automatizar la transcripción a notación musical de un archivo sonoro musical, permitiendo contar con una partitura digital de su contenido musical. En archivos monofónicos esta tarea está bastante conseguida, mientras que en archivos polifónicos se complica y aún está en vía de desarrollo. Probablemente la aplicación comercial más avanzada en este aspecto es Melodyne. Pero convertir un archivo sonoro en una partitura no garantiza la extracción de todo su contenido, dado que la notación musical solo caracteriza un subconjunto suyo, y también se considera que muchos documentos sonoros musicales pueden pertenecer a culturas musicales diferentes de las occidentales, por lo que, además se necesitan tecnologías de descripción para integrar metadatos o informaciones relacionadas con el contenido musical, de manera que se pueda recibir, categorizar o experimentar por parte de un oyente.

Los descriptores que se extraen a partir de dichas tecnologías son todos aquellos datos que puedan constituir una extensión del contenido del archivo sonoro con el fin de caracterizarlo, como, por ejemplo, tener un tempo de 100 pulsaciones por minuto o contener sonidos de trompeta. Dichos descriptores se calculan automáticamente a partir del análisis del archivo sonoro, siempre sin despreciar la probabilidad de error intrínseco en ellos, de hecho, la precisión de los mejores algoritmos oscila entre el 60% y 80%, dependiendo de los descriptores que se calculen. Algunos descriptores que constituyen el contenido del archivo sonoro, no se pueden calcular de forma automática, por lo que se requiere la anotación manual realizada por un oyente experto, como, por ejemplo, haber sido compuesto en el siglo XIX o transmitir una sensación de tristeza.

Los descriptores se pueden clasificar según su nivel, bajo, medio o alto, dependiendo de la distancia a la que estén sus grados de abstracción de los conceptos que un oyente o experto musical pudiese utilizar. Los descriptores de alto nivel también se denominan descriptores semánticos y son los más difíciles de extraer automáticamente y los más susceptibles a errores. También se puede distinguir los descriptores según el nivel temporal, denominados instantáneos, a los que se refieren a un punto concreto de la grabación, y globales, a los que hacen referencia a la obra completa o a un fragmento amplio de ella.

La mayoría de los descriptores se calculan a partir de representaciones de la forma de onda sonora basados en el teorema de Fourier. Es posible describir a bajo nivel el contenido frecuencial de un segmento sonoro muy corto a partir de operaciones como la denominada transformada de Fourier a corto plazo. Con la combinación de secuencias de dichas micro-descripciones, se puede obtener descriptores de más alto nivel, como, notas musicales o segmentos estructuralmente relevantes.

Algunos sistemas incorporan conocimientos musical y sonoro mediante modelos semánticos que capturan las similitudes y diferencias entre conceptos, como, por ejemplo, qué tienen en común las piezas de jazz y qué las hacen diferentes de las de folclore. Para la elaboración de dichos modelos se utilizan colecciones previamente anotadas por humanos, para que un sistema de aprendizaje automático extraiga los rasgos más relevantes para caracterizar un concepto, así como, los rasgos que permiten discriminarlo frente a otros conceptos relacionados.

En los últimos años, se destaca el estudio y la creación de técnicas para obtener automáticamente descriptores de las diferentes facetas musicales: timbre, ritmo, melodía, armonía, tonalidad y estructura. Aunque las representaciones no equivalen a los conceptos más utilizados en teoría musical, y la precisión de las técnicas desarrolladas sea limitada, constituyen una ayuda computacional a la catalogación manual, y en concreto, al procesamiento masivo de colecciones sonoras de miles o millones de archivos.

A continuación, se destacan algunas de estas técnicas según el descriptor a extraer:

- **TIMBRE**

Las técnicas de descripción de timbre se basan en analizar la señal sonora y extraer su representación frecuencial en un periodo de tiempo, que no es más que la intensidad relativa de cada frecuencia audible en cada instante. El análisis a lo largo de la señal da como resultado la evolución temporal frecuencial, denominada espectrograma. Las técnicas de análisis computacional de timbre describen las cualidades del sonido y estiman qué instrumentos están representados.

Actualmente, el etiquetado automático de archivos musicales polifónicos requiere la elaboración de modelos tímbricos correspondientes a los instrumentos que se desea identificar.

- MELODÍA

La melodía es un aspecto muy importante para caracterizar una pieza y, por ejemplo, reproducirla mediante tarareo. La descripción melódica está relacionada con la transcripción, aunque proporciona una descripción más amplia y relacionada con la interpretación o las características melódicas. Los sistemas artificiales aún no pueden descifrar la melodía predominante en un fragmento musical polifónico con un grado de fiabilidad aceptable, por lo que, el ámbito más adecuado para trabajar con descriptores melódicos es el de los instrumentos monofónicos o las polifonías simples.

Las técnicas para una transcripción melódica siguen tres pasos. En primer lugar, se extrae la frecuencia fundamental (relacionada con la altura) y la energía (relacionada con la intensidad). Existen numerosas propuestas con métodos para estimar la frecuencia fundamental de una señal sonora, pero todavía no es un problema resuelto para todos los instrumentos y condiciones, incluso en sonidos monofónicos. En cuanto a la energía, se extrae directamente de la señal de audio, sin margen de error, permitiendo detectar la aparición de los instrumentos en diferentes tesituras. En segundo lugar, la transcripción lleva a cabo una segmentación de notas basada en detectar variaciones de los descriptores calculados anteriormente y eliminando las variaciones que no son debidas a cambio de nota. La segmentación de notas es un problema complejo y aún en vías de investigación. Finalmente, cada nota queda descrita por su altura, duración y energía.

A partir de la transcripción melódica se puede calcular otros descriptores, como la tesitura, la distribución de notas e intervalos o los patrones melódicos utilizados, que servirán para caracterizarla. La descripción melódica automatizada tiene diversas aplicaciones, como la búsqueda de melodías por tarareo (query by humming), el análisis expresivo, el análisis comparativo de estilos o la transcripción en sí misma.

- TONALIDAD

La descripción tonal se realiza a partir de otros descriptores obtenidos en una misma señal sonora, por lo que, los métodos computacionales de descripción tonal caracterizan las relaciones entre las notas de una pieza musical, además se relaciona con el estilo y el carácter de la obra. Por otra parte, posibilita encontrar relaciones y similitudes entre diferentes piezas, información crucial para la identificación de versiones de una determinada composición.

Como el primer paso al trabajar con una grabación es la extracción de información sobre las notas presentes en un fragmento sonoro, para evitar problemas, las técnicas de descripción tonal utilizan los descriptores de bajo nivel denominados chroma. Estos descriptores aproximan el valor de la intensidad relativa de cada semitono de la escala temperada, es decir, aproximan los acordes de un fragmento musical. Para determinar la tonalidad de la que se trata, los descriptores de chroma se comparan con los perfiles tonales mayores y menores obtenidos a partir de experimentos con oyentes o derivados de la teoría musical.

- RITMO

Los métodos computacionales de descripción rítmica se relacionan con aspectos temporales de la música y con la organización de los eventos musicales en el tiempo, y se derivan de métodos de descripción de partituras.

En primer lugar, se extraen los parámetros más relevantes de la señal sonora, como pueden ser la evolución de la energía, tanto global, como distribuida en las diferentes bandas del espectro de frecuencia. A continuación, se analiza la repetición de dicha energía, que determina los pulsos existentes, y se estudia su variación a lo largo del tiempo y los patrones que forman los diferentes pulsos coexistentes (métrica).

Las áreas de trabajo fundamentales son la detección de pulsos (beat), la detección de tempo (beats per minute), la extracción de patrones rítmicos (métrica binaria o ternaria) y la descripción de características expresivas (desviación respecto a una partitura o notación inexpresiva); centrándose en algoritmos de detección de eventos a partir de los parámetros de la señal, algoritmos de cálculo de periodicidad y métodos de descripción métrica.

Actualmente, los métodos desarrollados funcionan bien con música muy rítmica y cuya métrica es estable en toda la pieza, pero en estilos métricos variables aún no resultan fiables del todo.

- ESTRUCTURA

Los métodos de descripción estructural automatizados se centran en la segmentación automática de grabaciones musicales y en la búsqueda de fragmentos repetidos dentro de una pieza musical, mediante la investigación de cambios bruscos o recurrencias periódicas de los parámetros relacionados con la intensidad, armonía, melodía o ritmo. A partir de la definición de los delimitadores estructurales, se puede realizar tareas como la navegación rápida a lo largo del archivo, o la comparación de segmentos de la misma clase dentro de una colección de archivos.

Los sistemas de descripción automática existentes pueden asignar etiquetas semánticas a un archivo sonoro con una alta fiabilidad, tal y como la mayoría de los oyentes podría realizar. Para ello, los sistemas requieren de técnicas de aprendizaje automático que detectan relaciones entre los descriptores musicales (timbre, tonal, melódica, rítmica y estructural) y las diferentes categorías a aprender, para que esto sea posible, tienen que disponer de una buena colección de ejemplos para cada categoría que el sistema tiene que asignar. Los sistemas actuales están capacitados para asignar etiquetas de diferentes tipos: emocionales, según el género, energía de la música o tipo de instrumentación. Para las etiquetas con una cierta variabilidad interpersonal, se desarrollan sistemas personalizados que aprenden la manera específica que cada usuario tiene para categorizar la música.

Las tecnologías presentadas han desarrollado diferentes métodos de búsqueda de información que exploran la descripción del contenido musical, tales como:

- QUERY BY HUMMING/SINGING

Mediante el tarareo o canto de una melodía por parte de un usuario, se realiza una búsqueda con el fin de encontrar música que contengan melodías similares.

- QUERY BY TAPPING

Mediante la pulsación rítmica de una tecla, el usuario crea una secuencia que se utiliza como búsqueda, dicha búsqueda está basada en el ritmo.

- QUERY BY PLAYING

El usuario introduce su petición de búsqueda mediante una interfaz musical proporcionada por el sistema.

- QUERY BY EXAMPLE

El usuario proporciona al sistema un archivo sonoro con un fragmento musical que hace de modelo para buscar otros fragmentos similares en una colección de música predefinida.

Además de estos métodos novedosos de búsqueda de información, también se han propuesto nuevos métodos para la visualización de colecciones y resultados de una búsqueda. Las colecciones pueden visualizarse como mapas geográficos clasificados mediante características semánticas o rasgos de bajo nivel, estos mapas se denominan auto-organizativos (SOM), y sirven para organizar colecciones musicales y categorías de estas en dos dimensiones. Las clasificaciones se presentan mediante nubes de etiquetas repartidas por toda la superficie del mapa, combinando texto con diferentes efectos visuales transmisores de información. Las etiquetas pueden asignarse tanto automáticamente, como con criterios personales y grupales de una comunidad de oyentes, también puede darse la combinación de ambas asignaciones. Una vez etiquetados personalmente unos cuantos ejemplos, los algoritmos de clasificación automática pueden sugerir etiquetas o etiquetar nuevos archivos.

Las tecnologías de análisis del contenido musical de archivos sonoros ofrecen complementos adicionales a otras opciones de organización y búsqueda de información. Dichas tecnologías requieren la adopción de nueva terminología, la aceptación de un cierto grado de imprecisión, y la necesidad de corrección de datos por parte del usuario. Aunque a día de hoy aún queda mucho por explorar y progresar en este campo, existen aplicaciones desarrolladas y enmarcadas a la evolución futura en el estudio de este ámbito.

La aplicación más desarrollada y en explotación comercial, consiste en la identificación de una determinada pieza musical que está registrada en diferentes soportes, mediante el cálculo de una huella digital. Esta tecnología es utilizada por las sociedades de gestión de derechos de autor para automatizar las estadísticas de reproducción pública de los temas musicales bajo su protección.

También se han desarrollado técnicas de detección de versiones, no explotadas comercialmente, que consisten en la búsqueda de similitud de una pieza musical dada, en vez de realizar una búsqueda de identidad física en el sonido, realiza una búsqueda de similitud que varía según los tipos de versiones. Aunque si se conocen algunos reproductores musicales como Sony, Philips, Bang & Olufsen o Yamaha, que utilizan criterios de similitud que explotan descriptores tímbricos, rítmicos, tonales y de género, para la organización de colecciones musicales personales. También existen algunos distribuidores de música en internet, como Jamendo, Musipedia, Midomi, Audiobaba o Muffin, que combinan el uso de estas tecnologías con la búsqueda basada en metadatos textuales asignados manualmente, para ayudar a explorar sus colecciones. Pero aún no se conoce aplicaciones específicas para bibliotecas o centros de documentación sonora y musical.

Como avance novedoso en las aplicaciones de estas tecnologías, también se destaca la reciente aparición de interfaces de programación de aplicaciones (API) ofrecidos por Canoris y Echonest, que facilitan el cálculo de descriptores vía web y su integración dentro de un sistema diseñado a medida. Esta novedad puede suponer un gran cambio, ya que permite la creación de entornos experimentales de descripción, acceso y organización de contenidos propios.

Los sistemas actuales presentan limitaciones importantes en cuanto a efectividad y eficiencia si las expectativas se centran en la automatización de dichos sistemas, aunque los avances experimentados en la última década son importantes y las promesas de cara al futuro resultan muy seductoras.

1.2 Identificación por tarareo

Los sistemas de identificación por tarareo son los encargados de reconocer una melodía por su contenido mediante el canto o tarareo realizado por un usuario. Este proceso se lleva a cabo mediante una estructura general que se compone por subprocesos, cada uno de ellos formado por módulos que aseguran su correcto funcionamiento (Figura 1). A continuación, se presenta y desarrolla detalladamente dicha estructura.

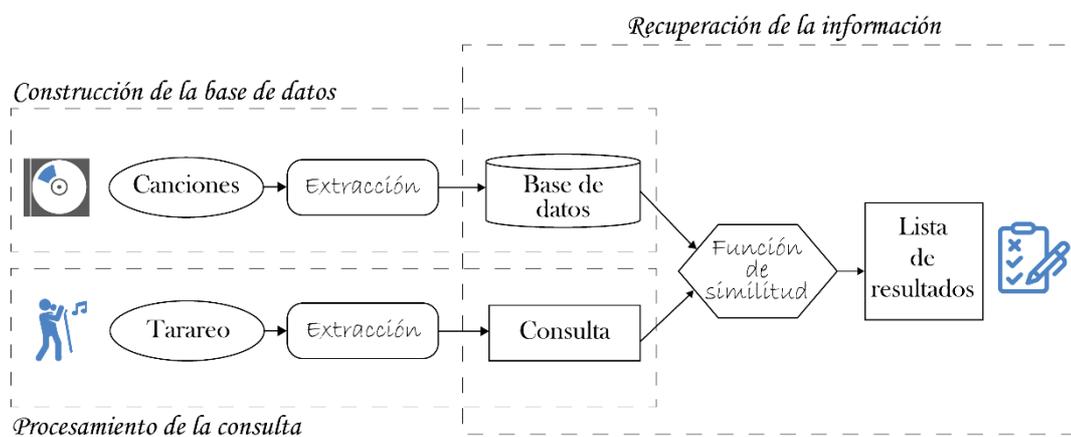


Figura 1: Arquitectura básica de un sistema de consulta mediante tarareo.

Para la construcción de la base de datos, primero se necesita una recopilación de las canciones que servirán de comparación con el canto realizado por el usuario, es decir, tienen que ser canciones originales que estén grabadas en alguna discografía oficial, para así generar una buena base de datos. Opcionalmente, puede incluirse las canciones idénticas que el usuario vaya a tararear, para que pueda escuchar y verificar que los resultados de su consulta son correctos. El conjunto de canciones puede variar en tamaño.

Tras esto, dichas canciones pasarán por un proceso de codificación, para extraer sus características más representativas y discriminatorias. El proceso de extracción de características es una forma de traducir las canciones a un lenguaje unificado, para luego poder trabajar con esta información de manera computacional. Cada lenguaje estará compuesto por las características que más representen al tipo de música que se quiere extraer.

Una vez obtenida toda la información con la que el sistema necesite trabajar, ya está formada la base de datos. La base de datos debe estar apropiadamente etiquetada, recogiendo toda la información relativa a las canciones (título de la canción, artista, etc.), y que el usuario recuperará como resultado de su consulta.

El proceso de consulta se inicia con la grabación de una canción cantada o tarareada por un usuario. El audio grabado se procesa de la misma manera que las canciones que conforman la base de datos, para que haya una concordancia entre ambos lenguajes, posibilitando así sus comparaciones. Normalmente, las canciones y los tarareos se codifican con un algoritmo, el cual, dada una señal de audio, calculará descriptores representativos coherentes con la información de la base de datos y la consulta, respectivamente.

Para recuperar la información referente a la consulta realizada, el sistema aplica la función de similitud a la consulta contra la base de datos y obtiene la canción identificada o un listado de las posibles canciones, ordenado por mayor grado de similitud. La función de similitud suele ser un algoritmo de comparación, que busca el parecido entre dos canciones a través del lenguaje codificado en la etapa de extracción de características.

1.2.1 Dificultades

La tarea de identificación de melodía por tarareo se enfrenta a problemas relacionados con el tratamiento de los datos de música y tarareo, además de los problemas habituales de recuperación de la información. A continuación, se nombra algunos de los problemas que se pueden encontrar dentro de un sistema de búsqueda por tarareo, y que se deben de tener en cuenta a la hora de realizar dicho sistema.

- La base de datos puede variar en tamaño, cuanto mayor sea el tamaño de ésta, mayor será el tiempo de comparación entre la consulta y las canciones en la base de datos, esto afecta directamente al tiempo de respuesta deseado.
- La consulta puede ser realizada por cualquier tipo de usuario, es decir, puede que no cante la canción con exactitud, por lo tanto, puede haber imprecisiones respecto a la pieza de música original, y con esto aparezca un tono no adecuado, una duración de la nota equivocada o una clave incorrecta.
- También puede que el usuario cante o tararee sólo un fragmento de melodía perteneciente a cualquier instante de una canción completa porque sea lo que recuerde, y con ello, aparezca distintas longitudes de melodías a comparar.
- El ambiente en el que se realiza la consulta puede ser ruidoso, esto crea distorsión en la señal de audio obtenida.
- Finalmente, el problema de la similitud melódica no está resuelto y existen muchos algoritmos, sobre todo para representación simbólica.

A continuación, se mencionan algunas propiedades deseables para que dichos problemas afecten lo menos posible al sistema.

1. El sistema debe trabajar con grandes volúmenes de datos.
2. El tiempo de respuesta debe ser reducido.
3. La tonalidad empleada en el tarareo no debe afectar al resultado de la consulta.
4. El usuario podrá tararear a cualquier velocidad.
5. El usuario debe poder tararear cualquier fragmento de una melodía.
6. La función de similitud debe permitir errores en el tarareo.
7. Los resultados deben ser precisos.

Si se realiza un sistema teniendo en cuenta estas características, se aumenta el poder discriminativo en el proceso de búsqueda y se consigue un sistema con mayor rendimiento. Por otra parte, para disminuir la problemática del tiempo de respuesta en relación con el volumen de la base de datos, se propone que el sistema tenga una base de datos estructurada y que la función de similitud realice la correspondencia de forma válida; con esto se obtienen unos resultados más eficientes y en el menor tiempo posible.

Una de las acciones que suelen realizarse, consiste en normalizar la melodía para poder identificar la consulta, pues no todos cantan en la misma altura, ni consumen el mismo tiempo.

Es necesario que la función de similitud entre el tarareo del usuario y las canciones de la base de datos puedan empezar en varios puntos de una canción, así se abordaría el problema de las diferencias entre las longitudes que se comparan. Definir una longitud mínima que deba tener la consulta sería otra medida de prevención a este problema.

Para afrontar el problema de la precisión en los resultados se puede tomar los resultados como recuperación por similitud, en vez de por equivalencia, si se espera encontrar el resultado en la primera opción, requiere que la precisión de los resultados sea mayor. También es necesario analizar en detalle qué aspectos de una melodía se parecen más al canto o tarareo de un usuario, es decir, qué características son las que se mantienen durante un tarareo.

1.2.2 Trabajos previos

La principal tarea de los sistemas de identificación por tarareo es la búsqueda de una melodía cantada por el usuario en una base de datos melódicos. La tarea de detectar melodías en una misma pieza musical ha recibido mucha atención de la comunidad de investigación a lo largo de la última década surgiendo así distintos enfoques para enfrentar este problema.

Desde la primera propuesta en [1], tanto por razones prácticas como técnicas, la base de datos se compone de música codificada (generalmente MIDI) en lugar de audio (wav, mp3, etc.), para reducir la información y posibilitar que haya un menor costo de procesamiento. Aún no existían técnicas automáticas de extracción de la melodía de una grabación para compararla con la melodía cantada por el usuario.

Las siguientes propuestas pueden clasificarse respecto a la representación de la melodía y respecto a la técnica de búsqueda, dos enfoques en los que se encuentran las diferencias entre los sistemas propuestos.

El primer enfoque y más tradicional está basado en la comparación de notas, es el que siguen los sistemas [1] y [2], que consiste en el fundamento de que una forma natural de comparar melodías es a través de las notas que las componen y consiste en transcribir la señal de voz a una secuencia de notas y buscar las mejores ocurrencias de ese patrón en una base de datos de melodías. La diferencia entre ambos sistemas es que en el [1] la transcripción de la señal se hace con valores de tono (frecuencia), mientras que en el sistema [2] además se introduce el valor de la duración.

El segundo enfoque, más reciente, utiliza una comparación de series temporales de frecuencia fundamental, utilizados en los sistemas [3], [4] y [5], que consiste en la comparación de melodías a través de series temporales de frecuencia fundamental (F_0), esto se hizo debido a que los errores en la transcripción automática deterioran el desempeño del sistema. La diferencia entre estos sistemas propuestos es que en el [3] integra la segmentación de forma automática, mientras que en los sistemas [4] y [5], la fragmentación temporal se realiza imponiendo al usuario que cante un fragmento de melodía previamente definido.

Posteriormente, se han realizado otras propuestas de sistemas. El [6] aplica el segundo enfoque, en el que la función de similitud hace una comparación de la melodía teniendo en cuenta la distancia de Levenshtein, y representa las melodías mediante la técnica del contorno melódico. El [7] propone una combinación de ambos enfoques, en el que se hace un análisis melódico para ajustar la altura de las notas y la función de similitud implementada para la comparación de los tarareos está basada en el algoritmo LDTW (Local Dynamic Time Warping), el cual restringe a una banda de cierta anchura la expansión de la dependencia. El [8] construye su base de datos con canciones estructuradas por índices y aplica una función de similitud basada en la comparación lineal con la restricción de la distancia Manhattan.

Hay algunos sistemas más recientes que se pueden encontrar, como es el caso del [9], que crea un gran sistema donde combina sistemas semejantes, cuya diferencia está en la sustitución de la representación de la melodía que se realiza con diferentes descriptores, utilizando la armonía, el tono y la línea de bajos en cada uno de ellos, para luego hacer una fusión de las diferentes representaciones y obtener resultados más fiables que si se hicieran por separado.

En definitiva, las investigaciones sobre el sistema de búsqueda por tarareo intentan mejorar el resultado aportando nuevas técnicas o combinando las ya existentes. Las características de los sistemas propuestos es que casi todos los enfoques intentan detectar el tono, algunos agregan el ritmo, y finalmente la mayoría convierten el tono en notas o secuencias. Dichos tonos son detectados por diferentes técnicas, como la autocorrelación, la heurística o por métodos estadísticos. Algunos enfoques más recientes, hacen coincidir las consultas directamente con la base de datos, como es el caso de la técnica de DTW (Dynamic Time Warping).

A lo largo de las investigaciones sobre un sistema de identificación por tarareo, cada estudio ha realizado un experimento en el que se ha utilizado diferentes tipos de música, desde música clásica ([2]), hasta música hindú ([6]), pasando por Los Beatles ([5] y [7]), o combinando dichos géneros con más [4], creando una variedad musical dentro de la base de datos.

1.3 Motivación

El Flamenco es un arte declarado Patrimonio Inmaterial de la Humanidad, grandioso título otorgado por La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) en el año 2010, reconociendo así el gran puzzle de valores culturales que este género pieza a pieza ha ido recopilando a lo largo de nuestra Historia. Con este reconocimiento, hay un compromiso por parte de las instituciones públicas de proteger, estudiar y difundir este arte, en concreto la Junta de Andalucía, ya que es en el sur de España donde nace, se cultiva y tiene su mayor actividad. Y sin duda alguna, hay un respeto e interés culto por parte del resto del mundo, siendo objeto de estudios sociológicos, etnológicos, de musicología, etc.

El Flamenco ha sido y sigue siendo un arte de estudio, tanto en el ámbito de la historia de su origen y evolución, como en el de la música, abarcando el cante, el baile y la guitarra. De ahí surge la flamencología, disciplina que se dedica a indagar y estudiar en su totalidad el arte Flamenco. Podríamos considerar este interés como una ciencia reciente, datada en 1955 con el lanzamiento del libro “Flamencología: toros, cante y baile” del autor Anselmo González Climent y desde entonces no ha parado de crecer.

Tanto es el interés por conocer este arte, que se han hecho estudios para comprenderlo desde un punto de vista matemático. Siendo la primera vez que se enlaza Flamenco y matemáticas, pero no música y matemáticas, ya que este enlace binario viene de hace más de 2500 años de la mano de Pitágoras, siendo los antiguos griegos los que establecieron las teorías musicales que hoy día conocemos.

Acercar el Flamenco a las matemáticas, en concreto, a métodos teóricos científicos, ha derivado en la utilización de sistemas de simulación computacional, y con ello, se ha vuelto a ampliar su campo de estudio, aunque es una metodología muy joven, por lo tanto, el Análisis Computacional del Flamenco es algo muy reciente, iniciada por el grupo de investigación Cofla¹. Esto se debe a que, el objeto de estudio en los proyectos que han abordado el tratamiento computacional de la música, por lo general, lo han realizado con música occidental y con otros fenómenos musicales, como la música pop, el rock, el jazz, etc. Los sistemas computacionales que se han desarrollado hasta el momento pueden ofrecer resultados relativamente efectivos para este tipo de música, puesto que existe conocimiento de los códigos que rigen su creación e interpretación.

El Flamenco se trata de una música de transmisión oral, campo musical en el que apenas se ha utilizado herramientas computacionales para su estudio, y cuando esto ha ocurrido es común que no se obtengan los resultados esperados. Esto se debe a que las particularidades específicas que presentan estos repertorios, generalmente son diferentes a las músicas académicas occidentales. Como le ocurre al Flamenco, que se compone de expresiones musicales con mecanismos que operan de forma distinta, siendo algunos de sus rasgos característicos la libertad interpretativa y la improvisación.

La música flamenca también tiene su espacio en las nuevas tecnologías, aunque es una música tradicional, tiene su hueco en aplicaciones móviles dedicadas a servir al público como guía de eventos y actuaciones, como forma didáctica musical o como forma de difusión.

Décadas atrás, el flamenco era objeto de estudio exclusivamente en el ámbito de las Humanidades, pero tan solo hace unos años atrás, ha despertado la curiosidad en el ámbito de la Tecnología, siendo objeto de estudio de la comunidad científica, requiriendo la colaboración del resto de campos que se dedican a su investigación, como la historia, la musicología o la antropología.

La música flamenca se caracteriza por ser una música de tradición oral, y como tal, su estudio tecnológico debe hacerse desde el punto de vista de la Etnomusicología Computacional, campo que desarrolla herramientas computacionales teniendo en cuenta el arraigo cultural de la música con el entorno donde se desarrolla. Además, la Etnomusicología Computacional está englobada en el campo interdisciplinar MIR (Musical Information Retrieval), cuyo objetivo es describir el contenido musical de señales de audio, así como

¹ www.cofla-project.com

proporcionar medios para visualizarlo y explotarlo con fines diversos, tales como la búsqueda, la navegación, la recomendación o el análisis.

Actualmente, el objetivo de la comunidad científica es la creación de tecnologías mediante el uso de algoritmos, como herramienta complementaria, para el estudio analítico de las estructuras musicales del flamenco.

Las distintas técnicas algorítmicas y procedimientos matemáticos para el estudio del flamenco constituyen una herramienta útil para enfocar estudios multidisciplinares del arte flamenco, como consecuencia en beneficio de su docencia, su preservación como herencia cultural, así como en la divulgación de la cultura andaluza y la ciencia, y además como avance en los siguientes aspectos:

- ✓ Orígenes y evolución del flamenco.
- ✓ Estudios culturales y musicológicos.
- ✓ Agrupación y clasificación de los diferentes estilos.
- ✓ Búsqueda de modelos melódicos.
- ✓ Creación de datos masivos de música flamenca.
- ✓ Sistemas de extracción para la comprensión de la naturaleza del arte flamenco.
- ✓ Desarrollo de plataformas públicas para facilitar el acceso a toda la población interesada.
- ✓ Sistemas de recomendación musical.
- ✓ Herramientas para la docencia del flamenco.

El flamenco es una de las principales industrias de la Comunidad Autónoma Andaluza con un gran potencial en el sector turístico a nivel mundial. El uso de tecnología destaca la necesidad de descubrir el flamenco en ámbitos como el comercial, turístico, como herencia cultural y en el marco académico andaluz. Tanto es así, como el avance tecnológico en los últimos años, que recientemente se ha inaugurado el 'I Máster Interuniversitario en Investigación y Análisis del Flamenco'.

El avance tecnológico musical viene dado por el desarrollo de tecnologías que son capaces de proporcionar al usuario herramientas de interacción fáciles y significativas con colecciones digitales de música. La disposición de dichas colecciones musicales es gracias a que en los últimos años haya habido un crecimiento enorme en la cantidad de material sonoro a través de sistemas de almacenamiento y redes informáticas. Las técnicas utilizadas, actualmente, son empleadas en otras músicas e implementadas en sistemas comerciales de navegación en grandes bases de datos musicales, recomendación musical, transformación de sonido, etc. La cantidad de trabajos publicados reflejan el tremendo crecimiento de dichas colecciones y la necesidad de realizar búsquedas dentro de su contenido de forma eficiente y efectiva. En el mundo del flamenco, las colecciones de grabaciones de cante son escasas, por lo que es el primer problema al que se enfrenta un investigador de este arte.

Una de las áreas de mayor interés dentro de la investigación de tecnología musical, es la descripción automática de señales de audio, ya que permite proporcionar automáticamente una descripción de las características más relevantes de una pieza, como son el ritmo, la armonía, la instrumentación, la estructura y la melodía. En el mundo del cante flamenco, al no existir colecciones de transcripción, una línea clave de investigación es la transcripción automática del cante. La tarea de la transcripción automática es extraer una representación simbólica del cante directamente desde la señal de audio de grabaciones. La representación es una secuencia de notas que forman el contorno melódico donde se omite una cierta cantidad de fenómenos, tales como vibrato, microtono o ligaduras. La transcripción automática del cante flamenco presenta multitud de dificultades debido a la presencia de la guitarra, al uso de ornamentación microtonal y melismas, y al hecho de que los cantaores no usan una afinación estable.

La transcripción automática extrae la notación musical de un archivo sonoro, pero esto no garantiza la extracción de todo el contenido albergado en él, además, el flamenco tiene convenciones musicales diferentes a las occidentales, por lo que, a parte de las tecnologías de transcripción, se necesita el uso de tecnologías de descripción, que buscan integrar informaciones relacionadas con el contenido musical y el contexto en el que dicho contenido se puede recibir, categorizar o experimentar por parte de un oyente.

En el estudio de cante flamenco aún queda mucho camino por recorrer y son escasos los productos comerciales que incluyen la música flamenca. Tampoco existe ningún sistema comercial de recomendación, ni un sistema de búsqueda automática basado en el contenido del audio que esté centrado en flamenco, lo que

justifica la necesidad de desarrollar el área de tecnología y flamenco para extender tales productos de análisis automáticos.

El Flamenco como entidad tiene su esencia llamada cante, que constituye una forma de interpretación vocal donde predominan su carácter melismático² e improvisatorio, es más, la voz es, en origen, solista e instrumento de sí misma, y hasta la aparición de la guitarra se bastaba sola o con la ayuda de los nudillos o las palmas para interpretar. Ella misma debe ser capaz de producir sentimientos, melodías y armonía para comunicarse con el público.

Es el cante flamenco, el principal objeto de estudio en este trabajo, junto con las distintas herramientas computacionales aplicadas dentro del sistema de búsqueda anteriormente descrito. Enlazar dos ámbitos tan distintos y, sobre todo, que tome algún sentido, es una tarea bastante ardua. A continuación, se describen las principales tareas abordadas en el proyecto Cofla.

1.3.1 El proyecto Cofla

COFLA (Análisis Computacional de la Música Flamenca) es un proyecto de investigación de técnicas de recuperación de información de música que apoya el análisis, la caracterización y generación de la música flamenca [39].

Este proyecto nace en 2009, apoyado por la Junta de Andalucía y dirigido por el Dr. José Miguel Díaz Báñez, profesor en el departamento de matemáticas aplicadas II de la Universidad de Sevilla y tutor de este trabajo fin de grado, en colaboración con profesorado de prestigio de esta misma Universidad, de la Universidad Politécnica de Madrid, Universidad Pompeu Fabra de Barcelona, Universidad McGill de Montreal (Canadá), Universidad de Piraeus de Grecia, entre otras, y cada vez son más los interesados que se suman a este gran proyecto.

El análisis científico que lleva a cabo este grupo combina conocimientos sobre antropología, musicología, psicología de la música, matemáticas, algorítmica, recuperación de información musical y procesamiento de señales de audio, todo ello enfocado de una manera interdisciplinaria para una máxima comprensión de la música flamenca.

Se mueven en el campo de la Etnomusicología Computacional, que es el estudio de la música a través de las personas, y es el caso del Flamenco un arte transmitido oralmente, cuyas canciones pasan de generación en generación (persona a persona) donde toma gran importancia la improvisación, y nos muestra la huella cultural de la etnia gitana en Andalucía; ofreciendo herramientas tanto para la caracterización expresiva del cante flamenco, como para la descripción automática de piezas flamencas, estudiando su melodía, armonía, timbre y ritmo, para apoyar la investigación comparativa de piezas de flamenco.

El grupo de investigación COFLA es pionero en el estudio sobre computación y música flamenca. Con el uso de programas informáticos especializados en la música flamenca pretende alcanzar los siguientes objetivos:

- 1) Facilitar el acceso eficiente a grandes colecciones de datos musicales que faciliten el conocimiento, divulgación y enseñanza del arte flamenco.
- 2) Enriquecer de información semántica, como, por ejemplo, descriptores de alto nivel, a las bases de datos que existen actualmente en el flamenco.
- 3) Ofrecer una herramienta para la visualización robusta de big data en música flamenca.
- 4) Alcanzar nuevas audiencias e interesados en el flamenco, dando “vida digital” al género.
- 5) Hacer partícipe a ciudadanos, artistas, comunidades y entidades asociadas en la preservación y administración de la herencia cultural andaluza, para que colaboren compartiendo material y colecciones personales, para su explotación y análisis, permitiendo la preservación del material que, de otra forma, se perdería con el tiempo.

El proyecto COFLA ha abordado varias tareas de investigación, realizando algunos avances en el área estudiada, así como se marcan retos por conseguir y poseen líneas abiertas de investigación.

² Técnica de cambiar la altura musical de una sílaba de la letra de una canción mientras se canta.

Uno de los problemas existentes en el estudio del cante flamenco es la falta de colecciones de grabaciones, tal y como se ha mencionado anteriormente. Para abordar esta problemática, el grupo COFLA ha creado dos grandes colecciones accesibles: TONAS y CORPUSCOFLA [22]. El primero contiene 72 grabaciones de martinets y deblas con sus correspondientes transcripciones automáticas con correcciones. El segundo está compuesto por 1500 grabaciones comerciales tomadas de las antologías del flamenco y puede considerarse como una muestra representativa del flamenco del siglo XX. Cada audio va acompañado con metadatos añadidos manualmente (información editorial, artistas, etc.). Además, se destaca la creación de una base de datos, FLABASE, que incluye datos bibliográficos y musicales extraídos de recursos y sitios en internet.

Otro problema, ya mencionado, en la investigación del cante flamenco es el de la transcripción automática de una señal de audio de manera eficiente. Abordando esta problemática, el grupo COFLA ha creado dos sistemas de transcripción automática para cante flamenco, SMSTools y CANTE, que constituyen un punto de partida para la transcripción automática de flamenco. CANTE está dando mejores resultados para el caso de cante flamenco ornamentado, su código está disponible en lenguaje Python y constituye el punto de partida de la mayoría de los algoritmos que se están desarrollando en esta área.

COFLA hace de unión entre las matemáticas y la música en el ámbito del Flamenco ayudándose de la tecnología a través del campo computacional soportado por métodos matemáticos, es decir, hace un estudio del Flamenco a través de las matemáticas utilizando herramientas para su comprensión, además generan material de apoyo para la enseñanza y la investigación, mediante los conocimientos que adquieren con el análisis de las peculiaridades del cante flamenco y procesando actuaciones flamencas.

Por lo tanto, la utilidad de este estudio en general se puede enfocar de dos formas didácticas. La primera, usando el Flamenco para conocer conceptos matemáticos, como en el trabajo publicado en 2013 en la Gaceta de la Real Sociedad Matemática Española, donde se expone una serie de actividades que se podrían hacer dentro de las aulas para la divulgación de conceptos matemáticos a través de la música flamenca. O bien, se puede enfocar de forma aplicada, usando las matemáticas para conocer un poco más el Flamenco, desde este análisis han creado varias aplicaciones relacionadas con el MIR.

Las tareas que engloban el estudio realizado por el proyecto Cofla son:

- Transcripción automática: es la extracción de la frecuencia de las notas musicales predominantes que emite un cantaor/a de flamenco, mediante la que define el contorno melódico (donde sube, donde baja, donde hace melisma, donde ornamentación, etc). Es un programa informático que elige las notas fuertes en sonido MIDI (piano) creando el contorno de melodía de un cante flamenco.
- Identificación del cantaor/a: es el reconocimiento del rostro del cantaor/a a través de un video flamenco, fijándose en las facciones de la cara. Sirviendo como herramienta para la organización de grandes colecciones de música flamenca mal etiquetadas o faltantes.
- Búsqueda de un patrón: mediante un programa informático, un método algorítmico, se hace una búsqueda del patrón o modelo melódico de un determinado palo dentro del cante flamenco, analizando las canciones de este palo dentro de una discografía de flamenco. De manera, que se puede comparar dicho patrón con cualquier otro audio de otro cantaor/a que interprete el mismo palo, identificando si se parece o ajusta más o menos al patrón encontrado anteriormente.
- Segmentación automática: Consiste en la división de un audio por partes, es decir, en detectar las diferentes partes de un audio, identificando percusión, ornamentación, palmas, voz, etc.
- Clasificación de estilos: es el reconocimiento de las variantes que se encuentran dentro de un mismo palo a través de un audio de cante, muy útil para la clasificación de los estilos, siendo una complicada tarea a simple escucha, ya que hay una similitud melódica entre ellas.

El objetivo general del proyecto es diseñar unos algoritmos para descifrar componentes y propiedades del cante flamenco, como la evolución de la música o la agrupación según los palos, siendo el objetivo final la creación de estas aplicaciones a nivel de usuario, para que cualquier investigador, docente o aficionado de la música flamenca pueda serle de utilidad.

Otra utilidad del estudio científico de la música flamenca, es la de entender la evolución y preferencia musical de los aficionados al Flamenco, para preservar la herencia cultural facilitando el acceso al material de Flamenco que se va generando, para la enseñanza dentro de los conservatorios, para facilitar el acceso internacional a plataformas sobre Flamenco de las personas extranjeras interesadas en este arte, etc.

Además, analizar la música flamenca con programas informáticos permite procesar una gran cantidad de datos en pocos segundos, siendo una forma de extraer información rápidamente, sirviendo no solo para el estudio de la música en su evolución, origen, sus características propias y transcripción, si no, para crear programas informáticos para la divulgación, la docencia e incluso para la comercialización de la música flamenca.

Una de las actividades relevantes que se crea en 2009 es el Primer Congreso Interdisciplinar de Investigación y Flamenco (INFLA)³, como un espacio universitario que constituye una apertura y acercamiento a la investigación científica en torno a la música flamenca, para dar a conocer el objetivo de este proyecto.

1.4 Objetivos

Este trabajo se encuentra en el marco del proyecto COFLA y se centra en la aplicación de un sistema de búsqueda por consulta tarareada, que permitirá analizar y caracterizar la música flamenca. Desde un punto de vista global, los resultados obtenidos en este trabajo serán analizados, y contrastados con los que se hayan obtenido en otras investigaciones pertenecientes a otros ámbitos dentro del proyecto COFLA.

En este trabajo se hará uso tanto de herramientas matemáticas y algorítmicas, como de herramientas tecnológicas ya desarrolladas y utilizadas por los componentes del grupo COFLA, que han sido facilitadas. Se debe subrayar la importancia de estas dos áreas, las matemáticas junto con la ingeniería y sus aplicaciones tecnológicas, en el desarrollo de este trabajo.

Uno de los objetivos del sistema de búsqueda por consulta tarareada que se presenta en este trabajo, es el estudio, comprensión y análisis de las herramientas utilizadas y facilitadas por el grupo COFLA, desde la transcripción automática hasta el algoritmo de comparación, todo ello enmarcado en el ámbito del cante flamenco.

Para realizar dicho sistema, se ha detectado la necesidad e importancia de la utilización de grandes colecciones de música flamenca, y que, a día de hoy, es un problema debido a la falta de estas, por lo que, para ello se ha tenido que marcar como objetivo, la creación de una colección de cante flamenco como trabajo de campo, para su utilización como consultas tarareadas.

El objetivo principal de este trabajo es la realización de experimentos con el sistema creado, como prueba de éste, y el análisis de su comportamiento en base a los resultados obtenidos. Los resultados darán muestra del procedimiento del sistema, considerando así si se trata de una herramienta útil o con necesidades de mejoras futuras.

Hay que destacar, que el sistema va a trabajar con diferentes versiones de una pieza musical de cante flamenco, por lo que, el estudio se centra en las similitudes entre los diferentes estilos que enmarcan el cante flamenco. Este estudio es el primero que se realiza dentro de este ámbito, por lo que, la incertidumbre del comportamiento del sistema y los resultados inciertos, están latentes.

³ <http://congreso.us.es/infla/infla2018/>

2 IDENTIFICACIÓN POR TARAREO PARA FLAMENCO

La consulta por cante o tarareo de música flamenca se puede considerar un caso especial de la tarea de identificación por tarareo. Como se ha visto en el capítulo 2, sección 2.5, los estudios hasta la fecha han realizado experimentos con todo tipo de músicas, pero es la música flamenca un género no explorado en este ámbito.

Como pionero en este campo se encuentra el grupo de investigación COFLA, como se ve reflejado en el capítulo 3, el cual trabaja con el cante flamenco desde un enfoque matemático y computacional, introduciendo métodos de transcripción, representación de la melodía y extracción del patrón melódico, entre otros.

En este proyecto se presenta un sistema de identificación por tarareo en el cante flamenco. El objetivo de este sistema es el de la identificación de melodías mediante consultas de tarareos o cantes flamencos realizadas por usuarios. En concreto, el interés es reconocer los palos o tipos de cantes flamencos a partir de interpretaciones personales. Además, las interpretaciones son realizadas por los usuarios, que puede cantar o tararear a capela o con un instrumento (siempre la guitarra) como acompañamiento.

2.1 El cante flamenco, sus clasificaciones y características

El Flamenco se crea, desarrolla y transmite de forma diferente e independiente de la música occidental o clásica. Al tratarse de una música de tradición oral, ha desarrollado sus propias reglas e incluso su propia jerga o lexicón. Por lo tanto, las prioridades, evaluación y gusto por una interpretación no coinciden necesariamente con las valoraciones en otras músicas.

El cante nace de la voz popular y aparece de forma individual interpretado por un cantaor o cantaora, para transmitir al oyente en un acto comunicativo y de forma poética, temas de la vida, la muerte, el dolor, la separación, el gozo o de cualquier sentimiento que lo esté ahogando, y es en ese acto de coger aire cuando brota.

Es en el cante flamenco donde no se busca una voz limpia, belleza de tono, y resumiendo, las perfecciones de lo que se suele entender como una buena voz para el canto, ya que el canto es una cosa, y otra muy distinta el cante. La capacidad expresiva y emocional, el buen compás o ritmo y la vibración puesta en las interpretaciones son los valores a tener en cuenta; y una voz ingrata, corta, puede rendir en flamenco excelentes resultados expresivos. Cada interpretación de cante es muy diferente según el intérprete que la realice.

Puede ir acompañado de una guitarra que, sobre el compás propio de cada palo, dibuja variaciones llamadas falsetas, a modo de breves cambios que sirven de adorno al cante. También puede ir acompañado de palmas, o por varios cantaores, guitarristas y bailaoras, formando los llamados cuadros flamencos, reunidos para el mismo objetivo e interpretando al unísono, a diferencia de un grupo coral. Aunque las mejores condiciones suelen darse en reuniones pequeñas y en un ámbito de cercanía con un solo cantaor acompañado exclusivamente por el compás que se forma entre sus nudillos y una superficie de madera.

El cante tiene una estructura basada en el compás, que consiste en un ciclo rítmico armónico, el cual, permite la identificación inmediata del estilo interpretado. El compás acompaña siempre al cantaor, dándole éste su personalidad con su timbre de voz y caudal pulmonar, haciendo varios cambios armónicos y rítmicos en comparación con el compás básico, pero siempre conservando la estructura general de este último.

Dos elementos fundamentales del cante flamenco son el ritmo y la melodía.

RITMO

Los cantes flamencos están ajustados a un metro rítmico (compás) que se repite periódicamente, que no es más que una distribución de acentos en una secuencia corta de pulsos. Resulta crucial para el cantaor ajustar las frases a ese ciclo periódico de acentos.

Por otra parte, toda vez que el intérprete tiene interiorizado el compás, puede realizar juegos de tiempo usando recursos como silencios, fraseos largos (alargar la duración), etc., que enriquecen el colorido rítmico de la ejecución.

La rítmica flamenca presenta tres tipos de compases o distribuciones de acentos:

- **Binario:** compás de 2/4 o de 2 tiempos. Este compás se subdivide en 4 partes rítmicas, que tienen el acento fuerte cada 2 o 4 tiempos. Otro compás binario, pero de subdivisión ternaria es el 6/8 o 12/8.
- **Ternario:** compás de 3/4 o de 3 tiempos. Acentuando el primer tiempo del compás.
- **De 12 tiempos:** es la suma y alternancia de los anteriores (de 2 y 3 tiempos). Es el más representativo del flamenco.

En la Figura 2, se representan dos patrones ternarios del flamenco con una notación numérica, donde los acentos fuertes se representan en negrita y los débiles sin señalización. Se ha etiquetado cada compás con un palo flamenco que lo utiliza, aunque no son exclusivos del estilo etiquetado.

[1 2 3 4 5 6 7 8 9 10 11 12] – *fandango*
 [1 2 3 4 5 6 7 8 9 10 11 12] – *soleá*
 [1 2 3 4 5 6 7 8 9 10 11 12] – *bulería*
 [1 2 3 4 5 6 7 8 9 10 11 12] – *seguiriya*

Figura 2: Notación numérica de los patrones rítmicos ternarios del flamenco.

MELODÍA

Muchas melodías en el flamenco están basadas en la escala modo Frigio de Mi (Figura 3), que se caracteriza por tener un semitono entre la 1ª y 2ª nota y la 5ª y 6ª. Este modo está muy presente en la música folklórica tradicional de la Península, pero el flamenco tiene sus propias peculiaridades, por lo que la melodía, en la música flamenca, se caracteriza por alterar la 3ª nota de la escala, de modo que se aumenta un semitono. Por lo tanto, el modo Frigio de Mi se convertiría al modo de Mi flamenco.

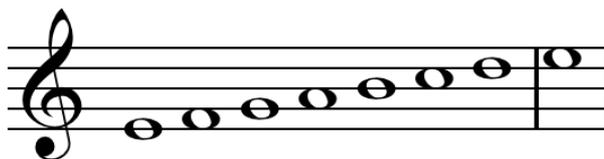


Figura 3: Escala modo Frigio de Mi.

Las características de la melodía flamenca son las siguientes:

- Melismas
- Ornamentaciones
- Utilización de cuartos de tono (intervalos más pequeños que la unidad del semitono)

La melodía es uno de los elementos más importantes en los cantes flamencos, ya que determina la diferencia entre unos y otros, es decir, cada cante tiene su estilo, por lo tanto, dependiendo del tipo de melodía y sus distintas inflexiones, puede ser de un estilo u otro, o de un autor u otro.

Aunque existe un patrón melódico establecido en cada cante, la interpretación de las melodías presenta variaciones sobre el patrón, que dependen de las facultades o preferencias estéticas del cantaor o de la escuela de cante a la que pertenece el intérprete. Se puede oír el mismo cante (con la misma letra) interpretado de forma diferente por Antonio Mairena o Camarón de la Isla. Esto se debe al uso personal del fraseo (organización de frases), ornamentaciones y al timbre, que le dan al cante flamenco una estética fácil de reconocer y diferenciar con otras músicas cantadas.

Desde el punto de vista de la tecnología musical, la ornamentación del flamenco dificulta la separación automática de las notas principales (patrón melódico) y los adornos (melismas).

Se conoce como “palo” a un estilo flamenco concreto, existiendo variantes de un mismo palo. Existen muchas maneras de clasificar los palos en distintos grupos, según su origen musical, según su procedencia geográfica, según su carácter serio o festero, y según sea su métrica.

A continuación, se introducen los palos del flamenco que se han utilizado y sus características más significativas para facilitar su comprensión dentro de un contexto tecnológico musical. Para más información puede consultarse⁴

- SEGUIRIYAS

Es uno de los estilos más representativos e importantes del Cante Jondo (los más graves y dramáticos).

Derivan del antiguo grupo de las Tonás, se independizan de este grupo con la incorporación del acompañamiento de la guitarra.

La letra es una estrofa de cuatro versos de 5, 6 y 7 sílabas, excepto el tercero que es de 11 sílabas, con rima asonante en los pares y a veces en los impares.

Lleva un compás de 12 tiempos, con los acentos en las posiciones 1, 3, 5, 8 y 11.

La línea melódica se desarrolla en torno a los cuatro acordes de la cadencia andaluza, siempre sobre el compás de 12 tiempos.

- SOLEÁ

Es uno de los pilares más importantes del Cante Jondo.

La letra de la soleá puede ser de estrofa de 4 versos octosílabos con rima asonante o consonante en los versos pares y a veces en los impares, de estrofa de 3 versos octosílabos con rima consonante o asonante en los versos impares, o estrofa de 3 versos con el primero pentasílabo y octosílabos los restantes.

Lleva compás de 12 tiempos, sucesión de 2 compases ternarios y 3 binarios, con los acentos en las posiciones 3, 6, 8, 10 y 12.

La melodía se basa en el modo Frigio de Mi.

Se puede diferenciar dos tipos de soleá:

- Soleá clásica: suele ser de tempo lento y acompañada por arriba, es decir, sobre Mi Menor.
- Soleá por bulerías: el tempo suele ser un poco más rápido y se acompaña por medio, es decir, sobre La Menor.

⁴ www.flamencopolis.com

- TANGOS

Forman parte de los cantes básicos, y configuran otro grupo genérico en el que se incluyen los Tientos y Tanguillos.

La letra es una estrofa de 3 o 4 versos octosílabos con rima asonante o consonante en los versos impares. También puede ser asonante o consonante en los pares. Se caracteriza por su compás binario.

- FANDANGOS

La letra se compone de 6 tercios melódicos. La estrofa contiene 4 o 5 octosílabos, con rima asonante o consonante en los pares de los que se repite uno o dos versos hasta llegar a los 6 tercios.

El compás que suelen tener los fandangos es ternario, pero gracias al proceso de aflamencamiento que ha sufrido este tipo de cantes se ha quedado sin compás, por su carácter más libre.

- BULERÍAS

Proceden del aligeramiento del compás de la soleá.

La letra es una estrofa de 3 o 4 versos octosílabos con rima asonante o consonante en los versos impares, y a veces en los pares.

Lleva un compás de 12 tiempos, con especial acentuación, en las posiciones 1, 4, 8, 9 y 11.

Hay que destacar una variante, que son las Bulerías por Soleá, interpretadas con tempo más lento, y se denominan bulerías “al golpe”.

- TIENTOS

Forman parte de la familia de los Tangos, pero con la diferencia de que se representan con un tempo más lento.

La letra es una estrofa de 3 o 4 versos octosílabos, con rima asonante en los versos impares o pares.

Lleva un compás binario, similar al de los Tangos.

- MALAGUEÑAS

Proviene del fandango primitivo malagueño.

El acompañamiento armónico se realiza sobre el acorde de Mi Menor, es decir, en términos flamencos “por arriba”.

Suelen terminar por verdiales.

- ZAMBRA

El compás de la zambra es binario, de 2 tiempos.

- ALEGRÍAS

Se caracteriza por su melodía diferente a otras cantiñas.

La letra es una estrofa de 4 versos octosílabos con rima asonante o consonante en los pares y a veces en los impares.

Su compás es idéntico al de la Soleá, de 12 tiempos, aunque el tempo es bastante más rápido.

2.2 Metodología

El procedimiento que se realiza en este trabajo está basado en la estructura descrita en la Sección 2.2, representada en la Figura 1. Este sistema se particulariza por la utilización de diferentes herramientas para el tratamiento de música flamenca, dentro de los distintos módulos que la compone.

Cada módulo particular de este sistema está descrito en la Sección 4.3 y Capítulos 5, 6 y 7, que se describen a continuación, y en el que se identificarán las diferentes características que hacen a este sistema peculiar.

Las herramientas utilizadas son facilitadas por el equipo de investigación COFLA, con la singularidad de la aportación propia en la construcción de las consultas.

2.3 El corpus Cofla

El enfoque de este sistema se evalúa utilizando una colección de canciones de flamenco recopiladas en una colección llamada CORPUSCOFLA, la cual contiene alrededor de 30 mil archivos de audio (en fichero wav o mp3) de canciones completas sacadas de las bibliografías discográficas que se pueden encontrar en el mercado musical.

En la Figura 4 se puede ver la caratula y el contenido de uno de los discos que componen dicho Corpus.



Figura 4: Caratula del disco "100 AÑOS DE FLAMENCO" contenido en la base de datos con la que se compara.

El Corpus está estructurado por carpetas con los nombres de las antologías de música flamenca que la componen, que a su vez contienen subcarpetas con los CDs que forman cada antología musical, y cada CD integra a su vez, los archivos de los audios elaborados.

Para acceder a un archivo de audio en concreto, hay que seguir el siguiente directorio:

En la colección se encuentran numerosos palos flamenco, como Fandangos, Soleares, Bulerías, etc., también diferentes artistas, voces tanto masculinas, como femeninas, en definitiva, presenta una gran variedad de intérpretes y estilos.

Esta colección de música flamenca es representativa del Flamenco del mundo real y, como tal, los resultados que se obtengan reflejarán lo que se espera obtener en un escenario actual.

Todas las canciones que componen el Corpus serán procesadas por un novedoso algoritmo de transcripción automática, descrito en el Capítulo 5, creando así la base de datos con los archivos generados (MIDI) del procesamiento de los audios transcritos.

Cada archivo está debidamente etiquetado con el título de la canción, nombre del intérprete y, en algunos casos, el estilo que interpreta. Por lo tanto, los resultados que se extraen están en la base de datos, y mostrarán el directorio que hay que seguir para escuchar el audio y así poder contractarlo con la consulta.

3 TRANSCRIPCIÓN AUTOMÁTICA

El procesamiento de los audios de las canciones se hace a través de un algoritmo de codificación para extraer las características más significativas para su posterior utilización en la función de similitud, es decir, se calcula descriptores representativos de los audios, que serán coherentes con la información de la base de datos, de hecho, de esta forma es como se construye la base de datos.

Tanto la base de datos como las consultas, están construidas por la codificación de las canciones y los cantos o tarareos, respectivamente.

Los atributos fundamentales de reconocimiento de canciones son la melodía, el ritmo, el tempo (lento o rápido), la intensidad (suave o fuerte), el timbre y, en algunos casos, la letra. Es en estas dimensiones en las que se es capaz de distinguir una pieza musical de otra. La melodía y el ritmo son las características más distintivas. La melodía de una pieza musical es la sucesión de notas de tono con una duración específica distribuidas u organizadas de forma que siguen un patrón melódico (ritmo).

En este sistema se utilizará como extracción de características un modelo novedoso de transcripción automática, que se centrará en obtener el valor de tono, su inicio y su duración en una melodía predominante del canto flamenco de una pieza musical polifónica y que está publicado en el artículo [10].

La transcripción del canto flamenco presenta una tarea de gran desafío, debido a sus complejas progresiones melódicas, las inexactitudes de la entonación, el uso de un alto grado de ornamentación, como el uso de micro tonales⁵ y la presencia del acompañamiento de la guitarra. Para una mayor comprensión, se va a desarrollar detalladamente el funcionamiento y características de cada parte de este modelo de transcripción automática.

La estructura del modelo contiene 2 etapas fundamentales, como se puede ver en la Figura 5.

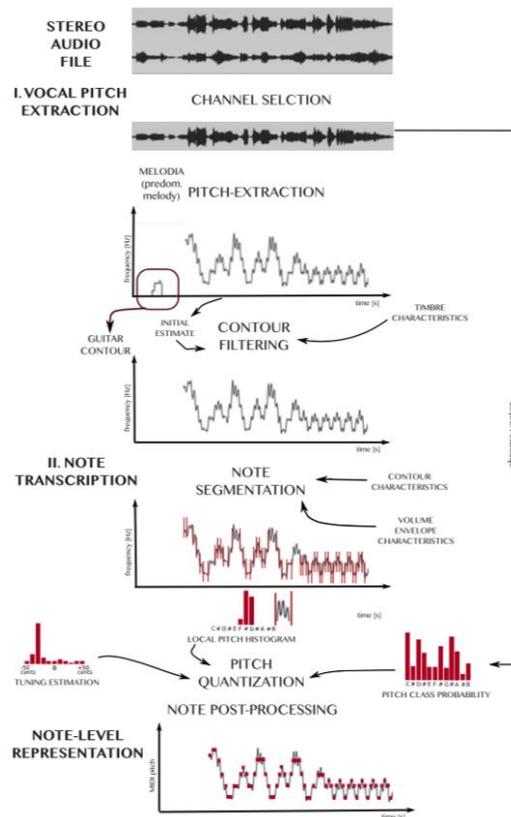


Figura 5: Estructura del modelo de transcripción automática.

⁵ Uso excesivo del vibrato e inestabilidad del timbre.

La primera etapa es la extracción del tono vocal en la que, dada una señal de audio polifónica, extrae los valores de tono predominantes presentes en la melodía de la voz, eliminando los tonos que corresponden a los contornos de instrumentación, en este caso, de la guitarra, que es el único instrumento acompañante en el cante flamenco, aislando así lo más posible la voz. Esta tarea en audios monofónicos se simplifica, ya que la extracción del tono en la melodía de la voz no necesita un aislamiento previo, ya que ésta es la única presente en el audio. Aunque en ambos audios, tanto monofónicos, como polifónicos, aparezcan silencios y sonidos de fondos causados por el ambiente, que necesiten ser eliminados de igual forma. Este procesamiento se lleva a cabo mediante 3 etapas, que son las siguientes:

1. Selección del canal: en esta etapa se usan características espectrales para seleccionar el canal estéreo en el que las voces son más dominante.
2. Extracción de la melodía predominante: en esta etapa se extrae la melodía predominante del canal previamente seleccionado utilizando el algoritmo [11], éste se encuentra disponible en la librería Essentia, y se encarga de estimar candidatos de tono en un nivel de fotograma basado en la suma de armónicos, y los agrupa en contornos continuos de tono y tiempo, utilizando principios de transmisión auditiva. Además, filtra los contornos en función de su prominencia de tono promedio que no forman parte de la línea melódica dominante. En concreto, este proceso es una estimación de la pista de tono correspondiente a la melodía dominante en una grabación de música polifónica, se trata de determinar si la melodía principal está presente en un marco de tiempo determinado o no.
3. Filtrado de contorno: en esta etapa se eliminan las partes del campo del contorno melódico extraído anteriormente, que se encuentran fuera de las secciones vocales. Es decir, elimina segmentos de contorno de guitarra basado en las características espectrales.

Tras eliminar los contornos de guitarra no deseados, se obtiene un conjunto de contornos correspondientes a la melodía vocal. Y comienza la segunda etapa de transcripción de notas, que consiste en la segmentación de dichos contornos continuos de frecuencia en el tiempo, en eventos de nota discretos, caracterizados por el tiempo de inicio, duración y cuantificación del valor de tono. Para ello, este procesamiento se lleva a cabo mediante 3 etapas, que se explican a continuación:

1. Segmentación: en esta etapa se va a identificar los eventos de nota discretos. Para ello, se distinguen dos tipos de inicio de notas, aquellos que coinciden con un cambio de tono (intervalos de inicio), y los comienzos en los que los tonos de dos notas consecutivas son iguales (inicio de tono estable). Para detectar los intervalos de inicio se puede combinar el método de detección basado en máximos locales, y refinarlo mediante el método de filtrado gaussiano derivado al contorno, que detecta otros inicios no descubiertos con el primer método. Para detectar los inicios de tonos estables, se combina el método de detección por decaimiento local repentino en el volumen, y el método de detección por una disminución repentina en el contorno del tono.
2. Etiquetado del tono: en esta etapa se asigna una etiqueta de tono a cada evento de nota resultante. Para ello, se calcula una referencia de sintonización global de toda la canción y luego, se combina las estadísticas de afinación del evento de nota con las probabilidades de clase de afinación global para determinar finalmente, el valor de afinación. Este valor de tono se asigna con métodos estadísticos ya que, hay dos dificultades que enfrentar. La primera es que la afinación de la pista puede desviarse de la referencia, haciendo que incluso segmentos constantes se ubiquen entre dos intervalos de semitonos, es decir, las ornamentaciones no simétricas pueden causar un desplazamiento en el valor de tono medio. Y la segunda, es que la entonación local puede no ser siempre precisa, y el tono estimado puede ubicarse ligeramente por debajo o por encima del valor objeto.
3. Post-procesamiento de nota: en esta etapa se lleva a cabo un proceso para refinar aún más la transcripción de notas obtenidas, aplicando restricciones de música básica en relación al tono y la duración. En este caso, se aplican limitaciones de rango de tono de menos de una octava, que es donde suele localizarse el cante flamenco, y se descartan los segmentos de notas de menos de cinco segundos. El fin en esta etapa es el de reducir la complejidad computacional y eliminar los posibles errores de transcripción.

Destacar que, el método descrito es capaz de transcribir interpretaciones de música flamenca improvisada y fuertemente ornamentadas, logrando mayor precisión que los métodos más avanzados. Actualmente, hay una aplicación en desarrollo llamada Cante, para llevar este método al alcance de la mano de los usuarios, cuyo objetivo es el de poder convertir un archivo de audio en formato wav, a notas musicales en formato MIDI.

4 FUNCIÓN DE SIMILITUD

La función de similitud melódica es la encargada de comparar el contorno melódico del tarareo con cada canción de la base de datos y estimar el grado de correspondencia entre las dos melodías, proporcionando una lista clasificada, de acuerdo con las puntuaciones de similitud obtenidas. Para asegurar robustez, debe tener en cuenta las propiedades de los tarareos que dificultan la identificación. Nuestra propuesta para la tarea de identificación por tarareo es usar una modificación del algoritmo Needleman-Wunsch [12].

4.1 Algoritmo Needleman Wunsch

El algoritmo Needleman-Wunsch está basado en el principio de la programación dinámica y es utilizado en el ámbito de la bioinformática para el alineamiento global de dos secuencias sumamente largas, como es el caso del genoma humano con 3000 millones de pares de bases, dando como resultado un puntaje que indica las similitudes que existen en ambas secuencias.

Han surgido numerosas variantes de algoritmos basados en el algoritmo de Needleman-Wunsch, dependiendo del ámbito al que se quiera aplicar. En el caso de este trabajo, el tema principal es la comparación de secuencias melódicas, por lo tanto, se utiliza una modificación del algoritmo original.

Para hacer esta memoria autocomprensiva se explica el funcionamiento del algoritmo con dos secuencias genéricas y las modificaciones a posteriori para la utilización de dicho algoritmo en el ámbito musical.

4.1.1 Needleman Wunsch para dos secuencias

El algoritmo funciona del mismo modo independientemente de la complejidad o longitud de las secuencias y garantiza la obtención del mejor alineamiento.

El término alineamiento se refiere al hecho de que cuando se combinan dos secuencias de símbolos discretos, el objetivo es alinearlos desde el principio hasta el final de forma óptima mediante una matriz de puntuación, donde se encuentra la mejor ruta de puntos de forma ascendente, y en la que se tiene en cuenta la penalización de huecos al ser formados.

Dadas dos secuencias genéricas formadas por símbolos $X := \{x_1, x_2, \dots, x_m\}$ e $Y := \{y_1, y_2, \dots, y_n\}$, que se quieren alinear, cuyo tamaño sea m y n , respectivamente, la matriz se forma colocando las secuencias X e Y en el eje x e y , respectivamente, y se rellena con una función de puntuación teniendo en cuenta la penalización de huecos

Se define como $f(i, j)$ a la función de similitud local, la cual asigna una puntuación de similitud al par formado por los elementos i -ésimo y j -ésimo de las secuencias X e Y , respectivamente. También se define d , un valor de penalización por no emparejar un elemento de la secuencia con otro produciendo un salto al siguiente, dejando un hueco que se penaliza con este valor. Entonces se forma una matriz de similitud S de tamaño $(m + 1) \times (n + 1)$ comenzando de derecha a izquierda y de abajo hacia arriba. Se inicializa la última columna y fila como:

$$\begin{aligned} S(m + 1, j) &= -(n + 1 - j) \cdot d \\ S(i, n + 1) &= -(m + 1 - i) \cdot d \end{aligned}$$

Y se va completando las demás casillas de la matriz usando la siguiente recurrencia:

$$S(i, j) = \max \begin{cases} S(i + 1, j) - d \\ S(i, j + 1) - d \\ S(i + 1, j + 1) + f(i, j) \end{cases} \quad (1)$$

Como el algoritmo desea obtener el valor y alineamiento óptimos, se va a guardar en la matriz $P(i, j)$ los pares de índices sucesores correspondientes a las coordenadas de la matriz $S(i, j)$ que se alcanza de acuerdo a (1). De esta forma, la ruta de alineación óptima puede rastrearse de manera progresiva siguiendo los sucesores almacenados en la matriz P , desde $P(1,1)$ hasta $P(m, n)$. Este proceso se conoce como *Trace-Back*. Una transición horizontal de una celda a sucesor indica que ambos elementos son diferentes, creando un hueco que se inserta en X ; lo mismo ocurre con una transición vertical, pero esta vez el hueco se crea en Y ; si ocurre una transición diagonal, indica que ambos elementos coinciden. La puntuación de alineación resultante, que cuantifica la similitud entre X e Y está contenida en $S(1,1)$.

Hay que tener en cuenta que el mismo valor óptimo se puede alcanzar mediante diferentes rutas, por lo que puede haber varias soluciones óptimas. Además, el algoritmo también se puede formular en la dirección inversa, por lo que la matriz S se llena de arriba hacia abajo y de izquierda a derecha, y la mejor ruta se obtiene mediante retroceso.

A modo de ejemplo, sean dos secuencias con símbolos:

$$\begin{aligned} X &= \{A, B, C, D, B\} \\ Y &= \{B, C, B\} \end{aligned}$$

Supongamos que la función de similitud es:

$$f(i, j) = \begin{cases} 1.5, & \text{si } X(i) = Y(j) \\ 0, & \text{en demás casos} \end{cases}$$

y la penalización de hueco es de $d = 1$. La matriz de similitud sería:

	A	B	C	D	B	
B	2.0	3.5	1.5	0.5	0.5	-3.0
C	-0.5	1.0	2.0	1.5	0.5	-2.0
B	-1.5	-1.5	-0.5	0.5	1.5	-1.0
	-5.0	-4.0	-3.0	-2.0	-1.0	0.0

Figura 6: Matriz de similitud S cuyas celdas contienen la puntuación obtenida utilizando la ecuación (1).

Realizando el proceso *Trace-Back* se tiene:

	A	B	C	D	B	
B	2.0	3.5	1.5	0.5	0.5	-3.0
C	-0.5	1.0	2.0	1.5	0.5	-2.0
B	-1.5	-1.5	-0.5	0.5	1.5	-1.0
	-5.0	-4.0	-3.0	-2.0	-1.0	0.0

Figura 7: Ruta de alineación óptima de dos secuencias con el algoritmo Needleman-Wunsch. Las flechas indican las transiciones de los sucesores al analizar la matriz, y las celdas rojas forman la ruta de alineación.

Obteniendo la ruta óptima de alineación:

A B C D B
- B C - B

con un valor óptimo del alineamiento de 2.0.

4.1.2 Needleman-Wunsch para similitudes melódicas

El algoritmo que se presenta ha sido propuesto en la tesis doctoral [14] y extiende los conceptos de comparación de secuencias abarcados en la sección anterior, con el fin de adaptarlos a la comparación de secuencias musicales.

El objetivo es detectar la aparición de un patrón melódico en un flujo de notas significativamente más largo, obteniendo una lista de prioridad de subsecuencias candidatas.

Para ello, el método introduce de forma sistemática las siguientes novedades:

- Extrae de forma iterativa las ocurrencias del patrón de referencia, ordenándolas con respecto a la puntuación de similitud.
- Introduce restricciones en el final de frase.
- Asegura que no varíe los cambios de tono.
- En función de las diferencias interválicas, formula los costes de transición entre los nodos de la matriz de similitud.

Como se ha visto anteriormente, la melodía es una secuencia de notas que se puede definir mediante varios conceptos. En este método, las notas se traducen mediante su altura, inicio y su duración.

En la primera etapa, vamos a trabajar únicamente con las secuencias de alturas extraídas, para posteriormente,

refinar los resultados eliminando los alineamientos correspondientes a un excesivo alargamiento de tiempo.

En tal caso, se define $A := \{a_1, a_2, \dots, a_m\}$ y $B := \{b_1, b_2, \dots, b_n\}$ como las secuencias de altura de la transcripción automática y el patrón de búsqueda, respectivamente, donde los elementos que las componen, a_i y b_j , son valores de altura en algún formato simbólico (MIDI o similar). Es esta etapa, se ignora la duración de las notas, de manera que la función de similitud local no está definida por el valor único de altura que las secuencias forman (puntuación de similitud), si no, por el intervalo que estos valores generan. Por lo tanto, se define:

$$I_B(j_2, j_1) = b_{j_2} - b_{j_1}$$

sujeto a $1 \leq j_1 < j_2 \leq n$, siendo el intervalo formado entre el valor de altura j_1 y j_2 del patrón de búsqueda, que no son necesariamente adyacentes. Y de manera similar, se define:

$$I_A(i_2, i_1) = a_{i_2} - a_{i_1}$$

sujeto a $1 \leq i_1 < i_2 \leq m$.

La función de costo queda definida como:

$$f(x, y) = \begin{cases} 1, & \text{si } x = y \text{ (Mismo semitono)} \\ -1, & \text{si } |x - y| = 1 \text{ (Un semitono)} \\ -\infty, & \text{si } |x - y| > 1 \text{ (Más de un semitono)} \end{cases}$$

Es decir, para intervalos iguales, la función recompensa con una puntuación de $+1$, para intervalos que se diferencian de un semitono, penaliza con una puntuación de -1 , y prohíbe las diferencias entre intervalos de más de un semitono, de ahí la puntuación de $-\infty$.

El algoritmo busca una secuencia de A cuyos índices sean ascendentes pero no necesariamente adyacentes, de tal forma que la secuencia de intervalos resultante coincida en puntuación con la subsecuencia de intervalos óptima.

Para resolver este problema desde un punto de vista de programación dinámica, se coloca la secuencia A en el eje horizontal, y la secuencia de B en el eje vertical, formando así la matriz de similitud S de tamaño $(m + 1) \times (n + 1)$, donde se inicializan la última fila y columna a cero. Se trabaja de manera ascendente y de derecha a izquierda, como en el algoritmo de la sección anterior. Cada nodo (i, j) de la matriz va a acumular una puntuación $S(i, j)$ que se calcula de la siguiente manera, donde $i < m$ y $j < n$:

$$\begin{aligned} h &= \max_{j+1 \leq k \leq j+G_h} S(i+1, k) + f(I_A(i+1, i), I_B(k, j)) \\ v &= \max_{i+1 \leq l \leq i+G_v} S(l, j+1) + f(I_A(l, i), I_B(j+1, j)) \\ S(i, j) &= \max\{h, v\} \end{aligned}$$

Las dos primeras ecuaciones imponen que el mejor sucesor del nodo (i, j) reside en la siguiente fila $(i + 1)$ o en la siguiente columna $(j + 1)$. Los parámetros G_h y G_v son enteros positivos que definen el radio de búsqueda en los ejes horizontal y vertical, respectivamente, es decir, controlan la longitud de hueco horizontal y vertical. En otras palabras, controlan cuántos valores de altura se pueden omitir horizontal o verticalmente cuando se busca el mejor sucesor del nodo.

Las coordenadas del mejor sucesor de cada nodo se almacenan nuevamente en una matriz P . Entonces una vez procesada toda la matriz S , se selecciona la puntuación más alta acumulada en la primera columna E_1 y el seguimiento de avance se revela en la matriz P , generando la mejor ruta de alineación hasta llegar a la última columna E_2 de la matriz S .

Sin embargo, la ruta puede ser rechazada si no termina en la última columna E_2 o comienza en la columna E_1 , constituyendo así los parámetros que representan las restricciones de punto final del procesamiento de alineación, es decir, se permite la omisión de las notas E_1-1 y E_2-1 como máximo de los puntos finales izquierdo y derecho del patrón, respectivamente. Al rechazar una ruta, se repite desde la segunda puntuación más alta hasta que se detecte una ruta válida (hasta última columna E_2) o hasta que todos los nodos de la primera columna (E_1) se procesen como candidatos de los puntos de partidas hacia la mejor ruta.

Si se quiere que el algoritmo devuelva más de una ocurrencia de patrones, el procedimiento se repetirá tantas veces como ocurrencias que quieran que se revelen, dando como resultado el número de rutas igual al número de ocurrencias que se desee.

Como ejemplo, consideremos dos secuencias en formato MIDI:

$$A = \{59, 58, 57, 61, 59\}$$

$$B = \{63, 62, 65, 63, 63\}$$

A proviene de la transcripción automática de un cierto cante y B es el patrón de búsqueda.

Hay dos posibilidades de transición en la matriz de similitud S , de forma adyacente o no necesariamente adyacente, tal y como se ha visto anteriormente, y que a continuación reflejan las figuras 10 y 11.

- Transición desde $S(i+2, j+1)$ hasta $S(i, j)$; a simple vista se comprueba que se trata de una transición no adyacente, ya que no se desplaza directamente hacia algún nodo de su alrededor, si no, que lo hace hacia un nodo dos posiciones más de su origen. Este movimiento se produce porque tanto el intervalo de la transición $I_A = 61 - 58 = 3$ (vertical) como el intervalo de la melodía de consulta (patrón) $I_B = 65 - 62 = 3$ (horizontal) son iguales. Al tratarse del mismo semitono, la función de similitud será $f(x, y) = f(3, 3) = 1$.
- Transición desde $S(i+1, j+1)$ hasta $S(i, j)$, siendo una transición adyacente ya que se desplaza de un nodo hacia el siguiente más cercano. En este caso, el movimiento se produce en un intervalo de transición $I_A = 58 - 57 = 1$ (vertical), mientras que el intervalo del patrón de búsqueda es $I_B = 65 - 62 = 3$ (horizontal), por lo que no son coincidentes, desviándose más de un semitono cuya función de similitud queda definida como $f(x, y) = f(3, 1) = -\infty$, ya que $|3 - 1| = 2 > 1$.

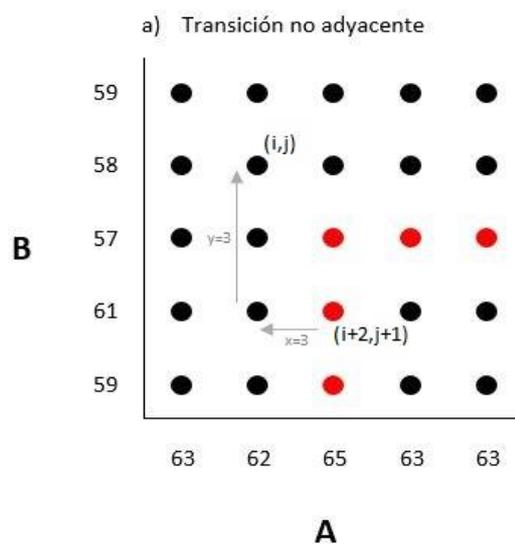


Figura 8: Transición no adyacente desde $S(i+2, j+1)$ hasta $S(i, j)$, siendo el rango de búsqueda de $G_h = G_v = 3$, como indican los nodos rojos.

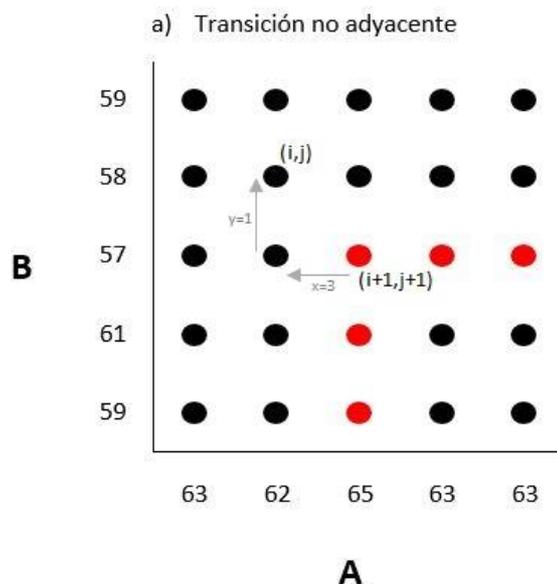


Figura 9: Transición adyacente desde $S(i + 1, j + 1)$ hasta $S(i, j)$, con un desplazamiento horizontal de un nodo y 3 de intervalo (x), y un desplazamiento vertical de un nodo y 1 de intervalo (y).

Se observa, que el algoritmo no solo busca que ambos intervalos de cada secuencia sean coincidentes, sino que, además tiene en consideración que el intervalo se divida en varios subintervalos, sólo teniendo en cuenta que los intervalos de la primera y última nota sean idénticos en ambas secuencias, como se puede ver en la figura 8, en el que la sucesión de secuencias de nota $\{62, 65\}$ se corresponde con la secuencia $\{61, 57, 58\}$ en el patrón de consulta de la transcripción automática, en una transición de ejecución donde una nota en el medio (57) se inserta de manera que no afecta al resultado final.

Toda la ejecución del algoritmo da como resultado la detección del patrón de consulta en la transición automática, ejecutando el proceso se extrae la ruta de alineación coincidente, como se muestra en la figura 10 y la tabla 1.

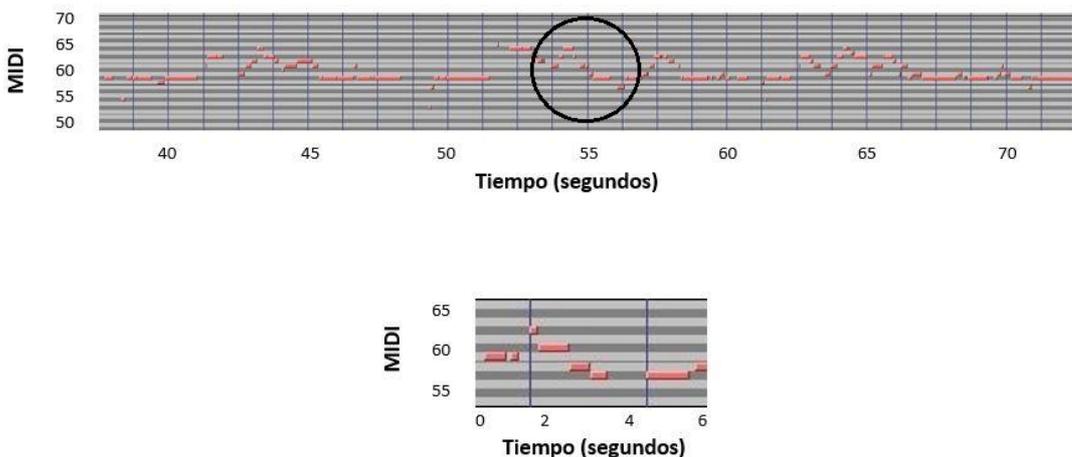


Figura 10: Ejemplo de detección de un patrón de consulta en una transcripción automática de una canción perteneciente a la base de datos, cuya ruta de alineamiento está marcada por el círculo.

A		B	
Altura	Duración	Altura	Duración
57	0.27	61	0.50
57	0.54	61	0.50
56	0.15	60	1.00
56	0.59	60	0.50
56	0.22	60	0.50
60	0.26	64	0.50
56	0.18	-	-
60	0.17	-	-
56	0.52	60	1.0
58	0.17	-	-
55	0.17	59	0.50
-	-	56	0.50
51	0.16	55	0.50

Tabla 1: Resultado de la mejor ruta de alineación del patrón de consulta realizado por Eva Rubichi en una transcripción automática de una Seguiriya de Manuel de Torres, en el que se observa la inserción de huecos en las notas no coincidentes.

Hasta ahora sólo se ha trabajado con la representación de alturas de las notas, pero hay una restricción dentro del algoritmo en la que se trabaja con la duración de dichas notas y sirve como restricción a la variabilidad en la duración, desechando cualquier ruta en la que la diferencia de inicio de las notas alineadas y formadas por ambos ejes se excedan en un umbral de estiramiento predeterminado igual a 3 o 1/3. Este procesamiento es equivalente al ya conocido como DTW (Dynamic Time Warping), método que se dedica al análisis de series de tiempo, cuya aplicación se ha destinado al reconocimiento automático de voz, entre otras aplicaciones [13][14].

5 BASE DE DATOS DE CONSULTAS

Para evaluar el sistema de identificación mediante tarareo propuesto, se utilizan dos colecciones de música flamenca. La base de datos CORPUSCOFLA, anteriormente descrita en el apartado 4.3, y una segunda colección que se trata de una base de datos formada por consultas recopiladas exclusivamente para evaluar dicho sistema.

Las consultas contienen un total de 37 canciones interpretadas por 7 cantaores de características diferentes. Estas grabaciones fueron realizadas en un trabajo de campo realizado en Jerez de la Frontera en un ambiente íntimo y personal, para simular un escenario lo más realista posible.

La duración de los archivos de audio (.mp3) varía de 38 a 345 segundos y la duración promedio de consulta es de 105 segundos. Debido a que el tiempo de las grabaciones se asemeja al promedio de tiempo que puede tener una grabación de cualquier cante flamenco que aparezca en un disco en la realidad, por lo tanto, de las canciones de la base de datos a comparar, se lleva a cabo un pre-procesamiento de los audios de consulta.

Los intérpretes que han colaborado en la recopilación de las consultas son de ambos géneros, masculino y femenino, y tanto profesionales dedicados al cante flamenco, como aficionados de este arte: Juan Zarzuela, Lela Soto, Pepe Alconchel, Eva Rubichi, Raúl Gonzalez, Angélica Montañés y Álvaro Linares. Un total de tres mujeres y cuatro hombres, la mayoría con una gran experiencia musical en el ámbito del flamenco. De todos ellos se detalla una biografía en el Anexo I.

Los audios de los cantes realizados por cada uno de ellos se dividen en audios más cortos de forma manual y como procesamiento posterior, tomando como criterio que cada audio comprenda una frase de la canción o bien, las frases que ocupen entre los descansos que ocurren a lo largo de la interpretación, ocupando alrededor de un minuto por cada audio dividido. En concreto, la duración de los audios cortados oscila de 11 a 98 segundos, con una duración promedio por consulta de 36 segundos.

Anterior a este procesamiento se descartaron 8 audios, por lo que se realizaron 45 grabaciones; estos se eliminan debido a la mala calidad, apareciendo numerosas perturbaciones de ruidos externos o equivocaciones musicales, quedando así los 37 audios sobre los que trabajar.

El corte de los audios grabados se realiza con la finalidad de que el procesamiento posterior sea más fácil, y es necesario ya que las consultas deben de ser de un tamaño mucho menor a las canciones con las que se compara. Al buscar un patrón más pequeño dentro de una canción completa menos coincidencias se deben dar, por lo tanto, menos dificultad a la hora de ser procesados.

También debemos destacar que tiene que haber un mínimo de notas para crear un contorno de melodía apto para poder trabajar con él. Por ello, los audios no pueden ser ni tan cortos que no se pueda crear un patrón de consulta, ni tan largos que ocupen el tiempo completo de la canción, girando los tiempos de los cortes de audios en torno a un minuto. Esto dio lugar a un aumento en el tamaño de la base de datos de las consultas con un total de 102 audios.

El registro que se realizó para la evaluación conforman un conjunto de canciones cantadas que no se corresponden necesariamente en versión con las canciones de la colección de la base de datos. Esto se debe a que se dio total libertad en la elección del cante por parte de los intérpretes. Por lo tanto, para cada consulta va haber más de una respuesta correcta dentro de la colección de la base de datos. Esto no disminuye el rendimiento de recuperación, ya que la búsqueda es sobre el estilo flamenco interpretado en la consulta. Si el resultado de alguna búsqueda coincide con exactitud a alguna versión de las canciones comprendidas en la base de datos sería considerado como un resultado de éxito total.

Los datos del conjunto de las consultas cantadas se recopilan en un archivo Excel, en el que aparece diferentes columnas con las características necesarias para la identificación de cada audio. La columna A contiene la etiqueta que lleva asignada cada audio completo grabado, la columna B contiene el nombre del autor/a de los audios de consulta realizados y que se han nombrado con anterioridad, la columna C contiene el nombre de la canción interpretada por cada cantaor/a, la columna D contiene el palo flamenco por el que canta dicha canción, la columna E contiene el intérprete original u autor que da origen a la canción, la columna F contiene la tipología del audio grabado, es decir, si va acompañado por guitarra (Poli) o se realiza a capella (Mono), y

las columnas G, H e I, contienen las etiquetas que se les asignan a los audios ya divididos y procesados en la transcripción automática. Estas etiquetas son nombradas de forma idéntica con la diferencia de la tipología del archivo, para los audios cortados (.wav), para el contorno de su melodía (.mid), para la transcripción con los valores de cada nota que forman la melodía (.notes), respectivamente.

Los archivos a los que se refieren las columnas G, H e I, se reúnen junto a los archivos .mp3 en la carpeta “Audios cortados”, ya que se trata de los mismos audios, pero en diferente formato. Esta carpeta junto a la de “Audios grabados”, están contenidas en otra cuyo nombre es “Queries”.

INTÉRPRETE	AUDIO	DIVISIÓN	ESTILO
Juan Zarzuela	28	4	Soleá
	31	4	Seguiriya
	33	3	Alegría
	34	2	Taranto
	36	2	Fandango
	37	3	Malagueña
	38	2	Malagueña
	39	3	Tiento
Lela Soto	40	3	Soleá
	41	3	Alegría
	42	4	Bulería
	43	2	Tango
	44	3	Fandango
	46	5	Tango
Pepe Alconchel	47	3	Malagueña
	48	2	Fandango
	49	2	Fandango
	50	3	Malagueña
	51	2	Fandango
	52	3	Malagueña
Eva Rubichi	54	3	Bulería
	55	3	Zambra
	56	3	Bulería

	57	3	Rumba
	58	4	Tiento
	59	3	Bulería
	60	2	Taranto
	61	4	Bambera
	62	3	Seguiriya
	64	3	Malagueña
Raúl González	67	2	Levántica
	70	3	Fandango
	65	3	Tango
Álvaro Linares	68	-	Bulería
	66	-	Bulería
Angélica Montañés	69	2	Bulería
	72	-	Tango

Tabla 2: Recopilación de los datos de las consultas.

Lo descrito anteriormente se resume en la tabla 2. En la que se observa que hay un total de 37 grabaciones de consulta, siendo la menor cantidad de 2 en el caso de Álvaro Linares, y la mayor de 9 en el caso de Eva Rubichi, con una media de 5 grabaciones por intérprete. También se puede ver el número de divisiones de cada una de ellas, que dan lugar a los 102 audios que forman la base de datos de las consultas. Las divisiones van de no realizarse ninguna como en el caso de las grabaciones 66, 68 y 72, ya que son tan cortas de tiempo que no contenían silencios entre las frases cantadas, a realizarse hasta 5 divisiones dentro de una grabación como es el caso del audio 46. Dando lugar a la menor cantidad de consultas registradas de 4 y la mayor de 28, con una media de 15 consultas por sujeto. Por último, se observa la variedad de estilos flamencos que se han interpretado, desde una Bulería hasta un Fandango, pasando por una Malagueña, Tientos, Tangos, Seguiriyas, incluso una Rumba o Bambera. Esto indica que hay una gran diversidad en la tipología del estudio, y con ello numerosas posibilidades de resultados, en los que se analizará la coincidencia de los estilos con los que componen la base de datos con la que se compara.

En este estudio, se usa como patrones de consultas las melodías definidas manualmente y que representan la variedad de estilos de cante de la música flamenca, tratando de recuperar sus ocurrencias ornamentadas y modificadas en transcripciones automáticas de representaciones reales que se lleva a cabo en la fase descrita en el capítulo 5.

Una vez que se recopila la base de datos de las consultas, se almacenan como documentos de la colección en una representación que sea accesible para el algoritmo de similitud, de esto se encarga la transcripción automática que genera archivos MIDI que serán guardados en la base de consultas como representaciones de secuencia melódicas.

0.156735, 0.505034, 60
0.667574, 0.316372, 61
0.989751, 0.240907, 60
1.23646, 0.386032, 60
1.70086, 0.269932, 61
1.99982, 0.217687, 60
2.22331, 0.19737, 60
2.42649, 0.261224, 61
2.69351, 0.22059, 61
2.91991, 0.386032, 60
3.38431, 0.272834, 61
3.73551, 0.188662, 61
3.92998, 0.15093, 60
4.44952, 0.24381, 61
4.78041, 0.20898, 60
5.12, 0.287347, 61
5.46249, 0.19737, 61
5.66567, 0.165442, 60
5.83692, 0.2322, 58
6.8615, 0.258322, 57
7.12562, 0.226395, 58
7.35782, 0.211882, 60
7.57551, 0.20898, 61
7.79029, 0.217687, 61
8.01379, 0.194467, 61
8.21406, 0.673379, 60
8.89905, 0.179955, 60
9.08481, 1.43383, 58
10.5244, 0.1161, 58

Figura 11: Ejemplo de los datos MIDI del contorno melódico que genera la transcripción automática de un Tango.

Las características extraídas por la transcripción automática de las consultas se almacenan en un fichero MIDI (.mid) en el que se encuentra el contorno melódico, y en un fichero Excel (.csv) como se muestra en la figura 11, donde cada dato en la fila representa el instante de inicio de nota, su duración y el valor en formato MIDI de la nota.

6 EXPERIMENTOS Y DISCUSIÓN DE LOS RESULTADOS COMPUTACIONALES

En este capítulo se encuentra el rendimiento del sistema propuesto en una tarea de consulta. El objetivo es detectar ocurrencias de secuencias MIDI anotadas en un corpus de transcripciones automáticas de grabaciones de flamenco.

Habiendo evolucionado a partir de las melodías folklóricas españolas, las canciones pertenecientes a estos estilos se basan en un esqueleto melódico particular para cada uno de ellos, que durante la interpretación están sujetos a cambios melódicos y modificaciones rítmicas en términos de una ejecución expresiva.

El sistema de identificación de melodías por tarareo que se ha descrito proporciona la funcionalidad para comparar e identificar la melodía correspondiente a los tarareos realizados por los sujetos.

Con los datos ya estructurados, seleccionados y procesados, se lleva a cabo con múltiples repeticiones del experimento, la comparación del desempeño de la variante del algoritmo Needleman Wunsch, enviando las consultas realizadas y devolviendo los resultados de la identificación.

Los resultados obtenidos son mostrados de forma ordenada según su proximidad a la consulta realizada, especificando el valor de la función objetivo alcanzada. Se recuperan 10 canciones de la base de datos con la que se compara, mostrando el directorio que ubica a cada una de las canciones.

Una vez recibido los resultados en una lista ordenada de mayor a menor grado de similitud y mostrada en la pantalla de comandos de Matlab, se procesan extrayendo la característica del estilo de cante flamenco al que pertenecen, considerándose como acierto la coincidencia de estilos entre la canción de consulta y la canción de cada resultado de los 10 que se muestran.

Tanto los resultados como la puntuación vinculada a cada uno de ellos se recogen en el Excel (.xlsx) "Información audios", aumentando así su número de columnas en dos (score and result), y el número de filas en 10 por cada audio de las 102 consultas recopiladas, obteniendo un total de 1020 resultados.

La métrica utilizada para el análisis de los resultados obtenidos en el proceso de identificación de las melodías mediante el tarareo serán distintas a las utilizadas en otros sistemas de recuperación de la información. La medida de tasa de éxito superior a una posición x es la más utilizada y se define como el porcentaje del estilo de la consulta que se encuentra entre las primeras x posiciones de los resultados obtenidos. En el caso de este experimento, la x toma valores de 1, 3, 5, 8 y 10, es decir, se evalúa el sistema teniendo en cuenta el porcentaje de audios de las consultas coincidentes en estilo al de las canciones de los resultados que se identifican en primera posición, el obtenido entre las tres primeras posiciones, entre las cinco primeras posiciones, entre las ocho primeras y el obtenido entre las diez primeras posiciones.

X	ACIERTOS	%
1	22	18
3	45	38
5	54	45
8	65	54
10	71	59

Tabla 3: Medida de los aciertos sobre las 120 consultas.

En la tabla 3 se muestran los resultados obtenidos. La primera columna contiene el número de posiciones a tener en cuenta, la segunda columna, el número de consultas en las que en sus x primeros resultados hay algún acierto, tomando como acierto el estilo de dicha consulta, y en la tercera columna se muestra el porcentaje de consultas con aciertos según el tamaño de resultados definido por x (si hay algún acierto entre las x primeras posiciones de los resultados).

Como se puede observar, los resultados obtenidos de la identificación van mejorando cuando mayor es el número de resultados a tener en cuenta, es decir, aumentar el tamaño de la lista de resultados aumenta la posibilidad de encontrar la respuesta correcta, con un aumento del 41% entre la tasa de acierto de los 1 primero y los 10 mejores. Pero estos, no son todo lo bueno que se esperaba obtener, ya que su se quiere identificar el estilo de una consulta en un escenario lo más realista posible, habría que fijarse en la primera posición de los resultados (top 1) y éste toma el valor del 18% de acierto, una cantidad bastante baja.

Otra forma de analizar los resultados es reducir la muestra a las 37 consultas iniciales, tomando como acierto la coincidencia de estilo que aparezca en el top 1 de los resultados en alguna de las partes en las que se ha dividido el audio inicial. Es decir, si se tiene un audio dividido en cuatro partes, y solo se tiene en cuenta el resultado obtenido en primera posición (top 1), si en alguno de esos cuatro resultados hay un acierto, se da como válida la actuación del sistema con respecto a la consulta no dividida.

Ahora se tiene una muestra de 37 consultas, de las cuales hay 16 aciertos top 1 y esto da un rendimiento del sistema (n) del 43%.

Se tiene un total de 7 individuos que han participado en la muestra. Al dividir dichos participantes en dos grupos, de tal manera, que en uno estén los que tienen mayor conocimiento musical y en el otro queden los que menor conocimiento musical tienen, los resultados quedan como se muestra en la tabla 4.

		AUDIOS		
		REALIZADO	ACIERTO TOP 1	%
PROFESIONALES	Juan Zarzuela			
	Lela Soto			
	Pepe Alconchel	29	13	45
	Eva Rubichi			
AFICIONADOS	Raúl González			
	Álvaro Linares	8	3	38
	Angélica Montañés			

Tabla 4: Agrupación de los participantes según estudios y sus rendimientos en las 37 consultas.

Por lo que el sistema actúa de la misma manera independientemente del nivel musical de los participantes, ya que el porcentaje es muy parecido, con un 45% de rendimiento en el grupo profesional y un 38% en el grupo aficionado. Ambos distan un 7% de diferencia, un porcentaje muy bajo y que prácticamente se puede despreciar.

Se realiza otra división de los participantes en dos grupos, pero esta vez diferenciándolos en si llevan o no acompañamiento de guitarra. Los resultados obtenidos se muestran en la tabla 5. Al igual que en el caso anterior, el sistema actúa de forma independiente del acompañamiento de instrumentación musical, obviando el 7% de diferencia entre el grupo con guitarra y el grupo a capella.

		AUDIOS		
		REALIZADO	ACIERTO TOP 1	%
A CAPELLA	Juan Zarzuela	8	3	38
	Lela Soto			
	Pepe Alconchel			
	Eva Rubichi			
	Raúl González	29	13	45
	Álvaro Linares			
	Angélica Montañés			

Tabla 5: Agrupación de los participantes según acompañamiento y sus aciertos en posición top 1 en la muestra.

Es importante destacar que el comportamiento del sistema es el mismo tanto para los participantes aficionados que tienen poca formación musical, como para las consultas que llevan acompañamiento instrumental, con un 38% en el rendimiento de cada caso.

Sin embargo, al realizar la división de los participantes según su género (femenino o masculino), se observa que el sistema se comporta mejor con una voz femenina con un 50% de rendimiento en el grupo de las mujeres frente a un 37% en el grupo de los hombres.

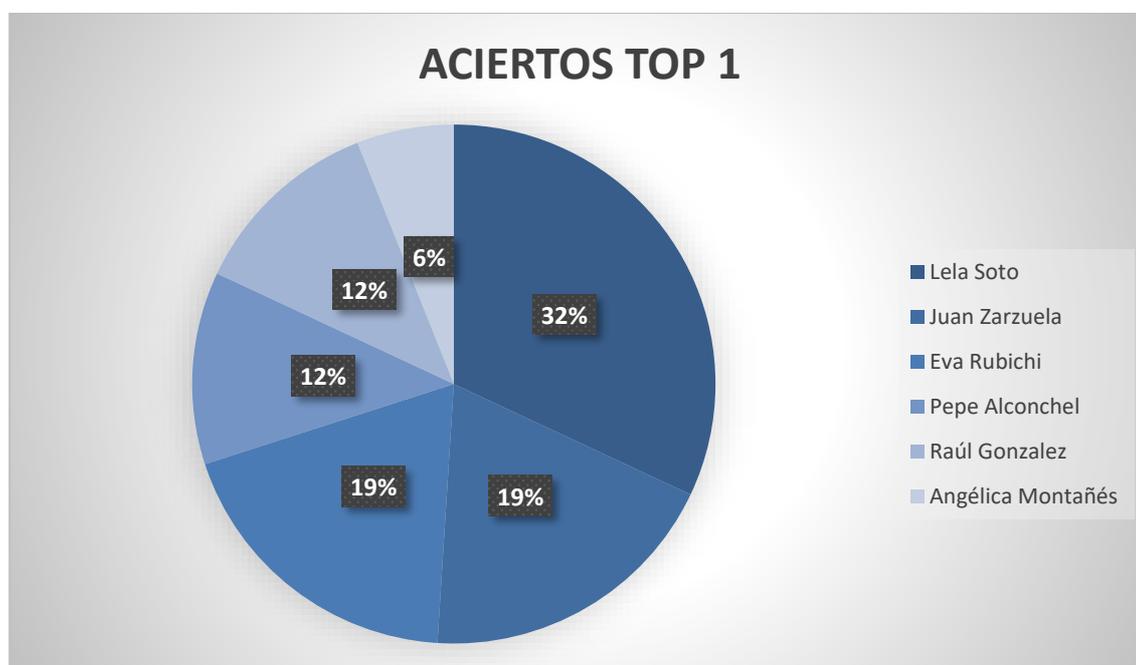


Figura 12: Distribución de los porcentajes de participantes en los aciertos top 1.

RANKING

1.	Lela Soto
2.	Eva Rubichi
3.	Juan Zarzuela
4.	Pepe Alconchel
5.	Raúl González
6.	Angélica Montañés

Tabla 6: Ranking de los participantes con más peso dentro del porcentaje de aciertos top 1.

Por otro lado, se tiene un total de 16 aciertos en los que se analiza el porcentaje que le corresponde a cada uno de los 7 participantes, tal y como refleja la figura 12. De este análisis se obtiene un ranking que muestra una lista ordenada de los participantes que han tenido más aciertos top 1, tal y como se ve en la tabla 6, en la que la primera posición es ocupada por la cantaora Lela Soto, esto indica que el sistema ha funcionado mejor con ella, mientras que con la cantaora Angélica González, el sistema ha tenido el peor comportamiento, ocupando la última posición en la tabla. También se observa que no aparece el cantautor Álvaro Linares, ya que éste no ha tenido ningún acierto en los audios aportados.

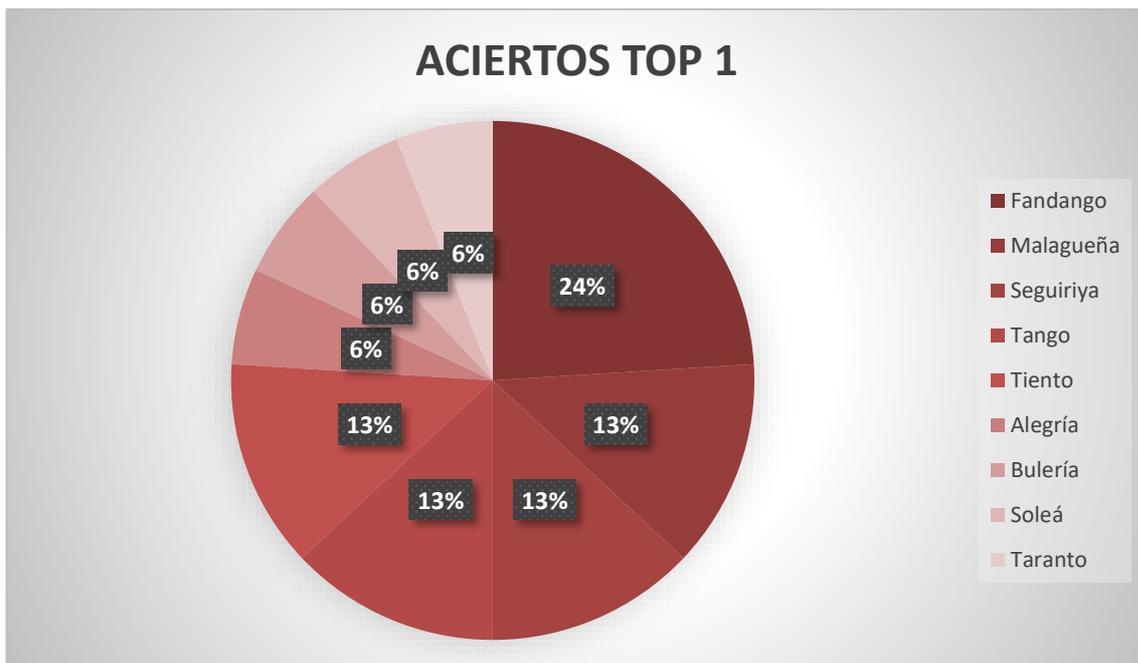


Figura 13: Distribución de los diferentes estilos de canto flamenco en los aciertos top 1.

RANKING

1. Fandango
2. Malagueña
3. Seguiriya
4. Tango
5. Tiento
6. Alegría
7. Bulería
8. Soleá
9. Taranto

Tabla 7: Ranking de los estilos de cante flamenco que componen los aciertos top 1.

De la misma forma, se analizan los aciertos teniendo en cuenta esta vez el estilo de cante que aparecen en la muestra. El resultado obtenido se refleja en la figura 13, mostrando los porcentajes de acierto que le corresponde a cada palo flamenco. De igual manera, se genera un ranking en el que la primera posición de la tabla 7 es ocupada por un Fandango, siendo el estilo de cante con el que el sistema se comporta mejor, y en el último lugar de la tabla 7 se encuentran el Taranto compartiendo mismo porcentaje con la Alegría, la Bulería y la Soleá, siendo los estilos flamencos que salen por parados en la evaluación del sistema. En esta lista no tiene lugar la Zambra, la Rumba, la Bambera, ni la Levantica, que son los estilos de cantes que completan la muestra, pero en este caso no han tenido aciertos, por los que se eliminan del ranking de aciertos.

PARTICIPANTES	%
Lela Soto	83
Raúl González	67
Juan Zarzuela	38
Angélica Montañés	33
Eva Rubichi	33
Pepe Alconchel	33

Tabla 8: Porcentajes de aciertos de los participantes según la aportación en la muestra. Los porcentajes se muestran ordenados de mayor a menor.

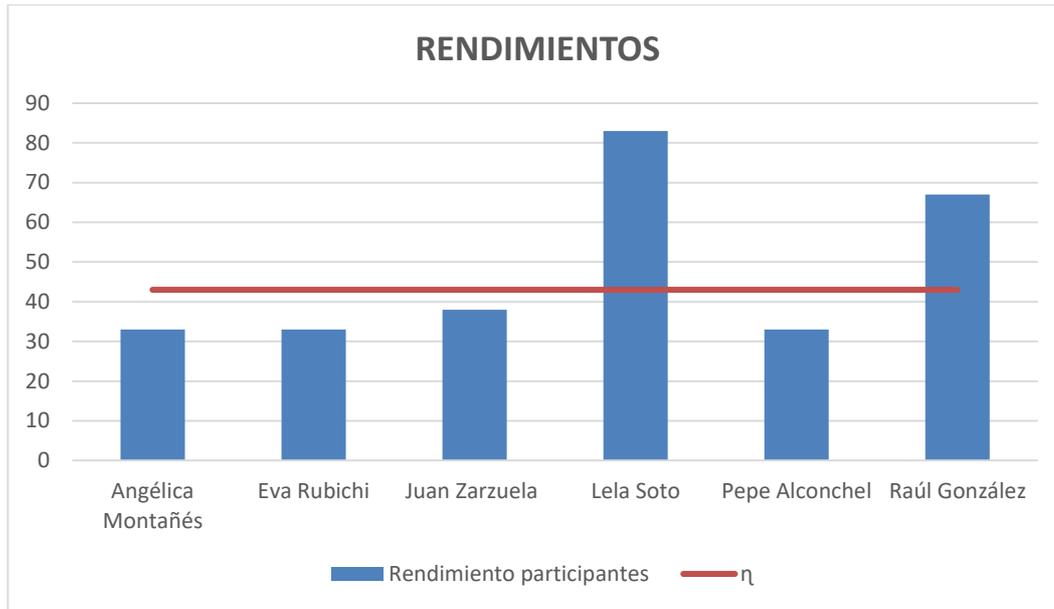


Figura 14: Comparativa de los rendimientos individuales de los sujetos según su participación con el rendimiento del sistema (η).

Por otro lado, se analizan los rendimientos de los participantes que conforman la lista de aciertos según la participación de cada uno de ellos dentro de la muestra. Los resultados que se obtienen aparecen en la tabla 8 y se comparan con el rendimiento del sistema, como muestra la figura 14. En este gráfico se observa que la cantora Lela Soto tiene un rendimiento del 83%, muy por encima del rendimiento general, esto indica que el sistema es con la que mejor que ha funcionado, le sigue Raúl González con un 67% en rendimiento, sobrepasando al general e indicando un buen comportamiento del sistema, y por último, por debajo del rendimiento general se encuentra Juan Zarzuela con un 38%, Pepe Alconchel, Eva Rubichi y Angélica Gonzalez con un 33%, que aunque esto indique que el sistema ha tenido un peor comportamiento con ellos en cuanto a los demás, se acerca al 43% establecido como el general, por lo que no se puede afirmar que el sistema haya tenido un mal funcionamiento con ningún participante.

ESTILOS	%
Seguiriya	100
Tiento	100
Fandango	67
Alegría	50
Soleá	50
Tango	50
Taranto	50
Malagueña	33
Bulería	14

Tabla 9: Rendimientos de los sujetos conforme con sus colaboraciones en la muestra.

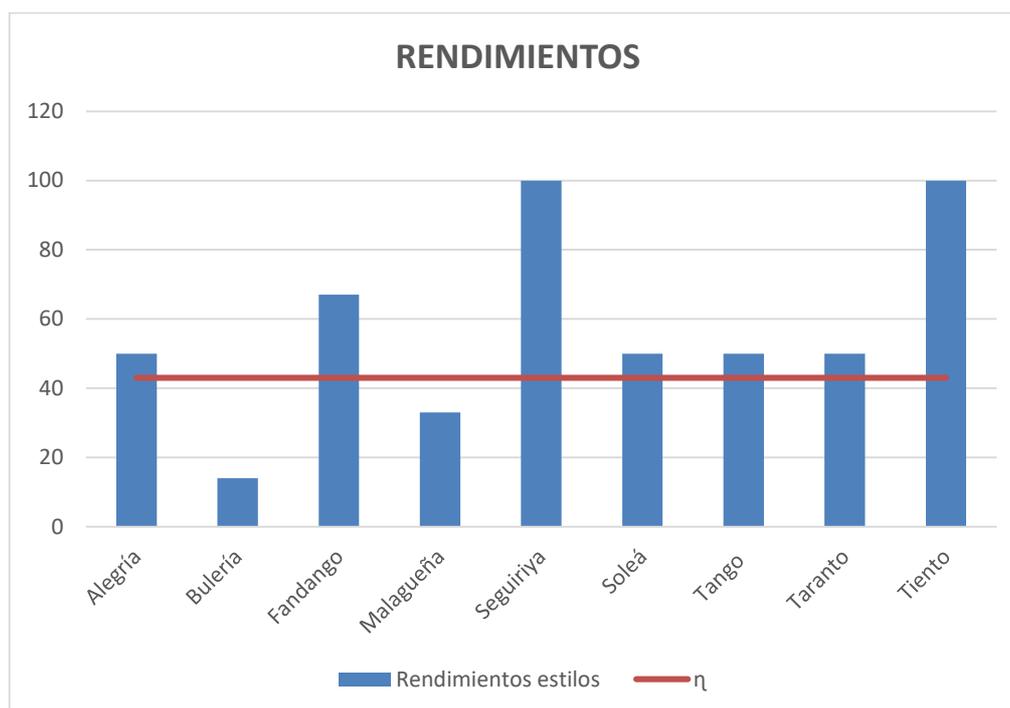


Figura 15: Contraste entre los rendimientos de los estilos según la contribución en la muestra y el rendimiento general (η).

De la misma manera se analizan los rendimientos de los palos flamencos que constituyen los aciertos respecto a la contribución en la muestra, como refleja la tabla 9. En la figura 15 se aprecian dichos rendimientos respecto al general del sistema, en el que se destaca la Seguiriya y la Soleá como los cantes mejor procesados con un 100% en sus rendimientos, mientras que la Bulería es el estilo por procesado con un 14%. Los demás estilos, aunque estén por encima y por debajo del rendimiento general, se acercan mucho a este, por lo que el sistema ha tenido un buen comportamiento, sin necesidad de destacar ninguno de ellos.

Una explicación razonable para el comportamiento del sistema con respecto a todos los casos está relacionada con la cantidad de variación a la que se somete un patrón durante la ejecución. Los cantantes tienden a aplicar recursos más expresivos y personales, por lo que aparece una dependencia de su origen, de la cadencia andaluza, de la prolongación de algunas notas como ornamentación, de la rapidez de ejecución y del acompañamiento, esto resulta en una mayor variación de rendimiento.

CANTAOR/A	ESTILO	RESULTADOS
Eva Rubichi	Seguiriya	1 Que Dobles Fatigas (Seguiriyas de Manuel Torre)
		2 Que Dobles Fatigas (Seguiriyas de Manuel Torre)
		3 Antonio Rengel - Siguiriyas
		4 Antonio Rengel - Soleares
		5 Jose Salazar - cambiaron los tiempos (siguiriyas)
		6 Se Me Acabo el Gusto (Seguiriyas de Su Padre Manuel)
		7 Enrique Morente - Siguiriya
		8 HistoriaDelFlamenco\Testimonios Flamencos CD 36\Track 5
		9 Se Me Acabo el Gusto (Seguiriyas de Su Padre Manuel)
		10 Siguiriyas Cruzadas De Curro Dulce Y De Diego El Marurro

Tabla 10: Canciones pertenecientes a la base de datos que el algoritmo extrae como resultados de la identificación de la consulta grabada número 62.

Por último, es interesante extraer el audio con mayor número de aciertos, que en este estudio pertenece a un cante por Seguiriya ejecutado por Eva Rubichi, tal y como muestra la tabla 10. En los resultados se observa la coincidencia en 9 de ellos, sin embargo, aparece una Soleá de Antonio Rengel que, aunque se califique como no relevante en esta tarea, demuestra el potencial de este sistema para descubrir similitudes y estructuras ocultas en el contexto de grandes estudios de extracción.

7 CONCLUSIONES

Dada la amplitud y complejidad del problema de identificación de melodías mediante tarareo, el presente trabajo es apenas una aproximación al tema. Aun así, se han abordado aspectos importantes de la identificación como la extracción de las características y su representación mediante transcripciones de rendimiento automáticas, y la función de comparación implementada por el algoritmo Needleman-Wunsch, obteniendo de esta forma un sistema completo.

Se presentó un sistema prototipo que se evaluó para el caso concreto de música flamenca con la creación de una base de consultas de diferentes estilos y géneros a comparar con una colección que contiene una amplia recopilación de música representativa en el ámbito del cante flamenco, demostrando las capacidades y limitaciones del sistema.

Al principio de realizar el proyecto se desconocía el funcionamiento de los sistemas de identificación de melodía por tarareo y como estos podían comportarse para el caso de los diferentes estilos que se pueden encontrar en la música flamenca, pero el estudio realizado de los mismos ha permitido adquirir los conocimientos necesarios e implementar nuestro propio sistema de identificación.

Como resultado final del trabajo, se ha evaluado el sistema de identificación de las melodías mediante consultas de cante flamenco, cuyos resultados ponen en conocimiento los siguientes aspectos:

1. El rendimiento del sistema aumenta en proporción a la cantidad de muestras de consultas recogidas.
2. El comportamiento del sistema no se ve afectado ni por la experiencia musical del cantante, ni por la instrumentación que lo acompaña, esto se debe a que las consultas son definidas por la transcripción automática que contempla este aspecto, eliminando cualquier variación de ruido o sonidos que no pertenezcan a la voz cantada.
3. El sistema muestra mejores resultados si se ejecuta con una voz femenina que con una masculina.
4. El sistema es sensible al cantante y al estilo de cante realizado, tal y como se muestra en los rendimientos individuales.
5. El sistema identifica diferentes estilos en una misma ejecución, relacionando estilos aparentemente dispares, por lo que es un potencial descubridor de estructuras ocultas.
6. El rendimiento general del sistema no es óptimo pero prometedor. Esto se debe a que el cante flamenco tiene un alto grado de ornamentación, por lo que los estilos interpretados son dependientes del cantante que los realiza.

Hay que tener en cuenta que los resultados obtenidos han sido una primera aproximación que irá mejorando con el tiempo y la experiencia.

7.1 Trabajo futuro

Las extensiones futuras de este trabajo podrían dirigirse hacia la mejora en la obtención de las características de las melodías a más bajo nivel para tener toda la información más detallada, que se utilizaría en la extracción y transcripción de dichas características, de forma que se obtenga con un mayor poder discriminativo.

También resultaría interesante hacer un estudio de los parámetros a utilizar en el algoritmo, eligiéndolos de tal manera que se optimicen dentro del esquema de programación dinámica.

La integración del dominio rítmico en el algoritmo de alineación sería una extensión futura interesante. También, la sustitución del algoritmo por otros algoritmos de alineación de secuencias en el ámbito musical, volviendo a evaluar el sistema para la realización de un estudio de comparación en rendimientos de recuperación, con el fin de validar cuales algoritmos son de utilidad a gran escala.

Un procedimiento alternativo para la tarea en cuestión, búsqueda por tarareo, sería el uso de redes neuronales y deep learning. En la tesis [14] se muestra el interés de estas técnicas para ciertas tareas como el reconocimiento de cantantes en videos.

Otro aspecto interesante que podría abordarse es la indexación o clusterización de la base de datos, ya que, en nuestro estudio, las canciones se agruparían por los estilos de cante flamenco, haciendo una búsqueda más concreta y selectiva, obteniendo resultados más exactos. De esta forma, el tamaño de la base e incluso el coste de la función de similitud no penalizarían tan gravemente las respuestas, habiendo una estructura en la base de datos, permitiendo que la recuperación de la información fuese más selectiva.

Otra posible indexación sería mediante el autoajuste de las consultas cantadas, observando si el cantante mantiene la misma afinación de referencia en toda la consulta, si no fuese así, la misma nota podría estar representada por frecuencias muy diferentes dentro de una sola consulta, cambiando el contorno de la melodía. Por lo que dicho desajuste podría observarse si se divide manualmente los sujetos en dos grupos: “Buena afinación” y “Mala afinación”, viendo como esto afecta en el rendimiento del sistema.

En cuanto a la base de datos, sería interesante el aumento de esta mediante la incorporación de los audios de consultas que resulten identificados, creando así un sistema de retroalimentación que facilitaría la recuperación de la información mejorando en cuanto a rendimiento.

El sistema propuesto en este trabajo, con la incorporación de todas estas posibles mejoras, podría tener una aplicación futura de gran éxito, mediante su adaptación a una aplicación informática (Android) de manera pública, posibilitando el acercamiento y aumento de conocimientos en el ámbito del cante flamenco.

Si bien todavía hay mucho trabajo por hacer en esta área, los resultados presentados aquí sirven como una prueba de concepto y se espera que conduzcan al futuro desarrollo de soluciones de sistemas de identificación por tarareo de alto rendimiento totalmente automatizadas.

REFERENCIAS

- [1] A. Guias, J. Logan, D. Chamberlin and B.C. Smith, "Query by humming-musical information retrieval in an audio database", In ACM Multimedia, 1995.
- [2] R.J. McNab, L.A. Smith, I.H. Witten, C.L. Henderson and S.L. Cunningham, "Toward the digital music library: tune retrieval from acoustic input", Proc. ADM Digital Libraries, 1996, pp.11-18.
- [3] D. Mazzoni and R. Dannenberg, "Melody matching directly from audio", ISMIR, 2001.
- [4] R.B. Dannenberg and N. Hu, "A comparison of melodic database retrieval techniques using sung queries", Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, 2002, pp.301-307.
- [5] Y. Zhu and D. Shasha, "Warping Indexes with Envelope Transforms for Query by Humming", International Conference on Management of Data, San Diego, California, 2003.
- [6] P., Gaurav, C. Mishra, and P. Ipe, "TANSEN: A System for Automatic Raga Identification", IICAI, 2003.
- [7] E. López and M. Rocamora, "Tararira: Query by singing system", Proceedings of the Second Annual Music Information Retrieval Evaluation eXchange, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 2006.
- [8] U., Erdem, et al., "Creating data resources for designing usercentric frontends for query-by-humming systems", Multimedia systems, vol.10, no.6, 2005, pp.475-483.
- [9] J. Salomon, J. Serrá and E. Gómez, "Tonal representations for music retrieval: from version identification to query-by-humming", 2012.
- [10] N. Kroher and E. Gómez, "Automatic transcription of flamenco singing from polyphonic music recordings", 2016.
- [11] J.S Salomon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics", IEEE Transactions on Audio, Speech and Language Processing, vol.20, no.6, 2011, pp.1759-1770.
- [12] S.B. Needleman, and C.D. Wunsch, "A general method aplicable to the search for similarities in the amino acid sequence of two proteins", Journal of molecular byology, 1970, pp.443-453.
- [13] S. Theodoridis and K. Koutroumbas, "An Introduction to Pattern Recognition: A MATLAB Approach", Academic Press, Burlington, USA, 2010.
- [14] N. Kroher, "Flamenco music information retrieval. PhD thesis", 2018.

- [15] P. Herrera Boyer y E. Gómez Gutiérrez, "Tecnologías para el análisis del contenido musical de archivos sonoros y para la generación de nuevos metadatos", Universidad Pompeu Fabra y Departamento de Sonología, Cataluña, 2017.
- [16] A. Kotsifakos, P.P. Papadimitriou, J. Hollmén, D. Gunopulos and V. Athitsos, "A Survey of Query-By-Humming Similarity Methods", Academic Press, 2012.
- [17] J. M. Díaz-Báñez y N. Kroher, "Tecnologías para el análisis multidisciplinar del flamenco", Universidad de Sevilla, 2019.
- [18] M. Rocamora, P. Cancela and A. Pardo, "Query by Humming: Automatically building the database from music recording", Uruguay, 2013.
- [19] B. Stasiak, "Adaptive Approach to DTW-based Query-by-Humming System", Institute of Information Technology, Poland, 2014.
- [20] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff and A. Klapuri, "Automatic music transcription: challenges and future directions", 2013.
- [21] E. Gómez and J. Bonada, "Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing", 2013.
- [22] N. Kroher, J.M. Díaz-Báñez, J. Mora and E. Gómez, "Corpus COFLA: a research corpus for the computational study of flamenco music", 2016.
- [23] O. Nieto and M.M. Farbood, "Identifying polyphonic patterns from audio recordings using music segmentation techniques", 2014.
- [24] A. Pikrakis, S. Theodoridis and D. Kamarotos, "Recognition of isolated musical patterns using context dependent dynamic time warping", 2003.
- [25] J. Song, S.Y. Bae and K. Yoon, "Query by humming: Matching humming query to polyphonic audio", 2002.
- [26] J.M. Díaz Báñez, "On math problems in the study of flamenco singing", *La Gaceta de la Real Sociedad Matemática Española*, vol.13, 2013, pp.513-541.
- [27] R. Typke, "Music retrieval based on melodic similarity", *Academy of Music Utrecht*, 2007.
- [28] M. Mongeau and D. Sankoff, "Comparison of musical sequences", *Computer and the Humanities*, vol.24, 1990, pp.161-175.
- [29] D. Müllensiefen and K. Frieler, "Optimizing measures of melodic similarity for the exploration of a large folk song database", *Proceedings of the 4th International Conference on Music Information Retrieval*, Baltimore, Maryland, 2004.
- [30] A. Uitdenbogerd and J. Zobel, "Melodic matching techniques for large music databases", *ACM Multimedia*, 1999.
- [31] N. Adams, M. Bartsch, J. Shifrin and G. Wakefield. "Time series alignment for music information retrieval", 2004.

- [32] M. Rocamora and P. Herrera, "Comparing audio descriptors for singing voice detection in music audio files", Brazilian Symposium on Computer Music, 2007, pp.187-196.
- [33] M. Antonelli and A. Rizzo, "A correntropy-based voice to midi transcription algorithm ", IEEE Workshop on Multimedia Signal Processing, 2008, pp.978-983.
- [34] E. Gómez, J. Bonada and J. Salamon, "Automatic Transcription of Flamenco Singing from Monophonic and Polyphonic Music Recordings", Interdisciplinary Conference on Flamenco Research, 2012, pp.191-198.
- [35] L. Song, M. Li and Y. Yan, "Automatic Vocal Segments Detection in Popular Music", Conference on Computational Intelligence and Security, 2013, pp.349-352.
- [36] N. Kroher, A. Pikrakis, J. Moreno and J.M. Díaz Báñez, "Discovery of Repeated Vocal Patterns in Polyphonic Audio: a Case Study on Flamenco Music", European Signal Processing Conference, 2015, pp.41-45.
- [37] K. Lemström, "String Matching Techniques for Music Retrieval", University of Helsinki, 2000.
- [38] Shazam. Home page, 2019. <http://www.shazam.com/>.
- [39] Computational Analysis of Flamenco Music. Home page, 2019. <http://www.cofla-project.com/>.
- [40] R.B. Dannenberg and N. Hu, "Pattern discovery techniques for music audio", Journal of New Music Research, 2003, pp.153-163.
- [41] R.B. Dannenberg and N. Hu, "Understanding search performance in query-by-humming systems", Proceedings of the 5th International Conference on Music Information Retrieval, Barcelona, 2004.
- [42] R.A. Mora, J. Ramírez and M. Elizondo, "Heuristic for solving the multiple alignment sequence problem", Revista de matemática: Teoría y Aplicaciones 18(1), México, 2011, pp.121-136.
- [43] Conceptos musicales. Home page, 2019. <http://www.flamencopolis.com/>.

ANEXOS

Anexo I: Términos musicales

Es importante tener en cuenta algunos conceptos musicales que aparecerán a lo largo de este trabajo y que servirán para entenderlo con mayor facilidad. A continuación, son definidos los principales conceptos musicales.

- **Armonía:** Es la técnica de la construcción y relación entre diferentes sonidos o notas que se emiten al mismo tiempo (acordes).
- **Clave:** Es un indicador de la altura de la música escrita, asignando una determinada nota a una línea del pentagrama, que se toma como punto de referencia para establecer las posiciones del resto de notas. Se representa mediante figuras musicales específicas.
- **Compás:** Es la estructura métrica que llevan las canciones, indicando la división del pentagrama en partes iguales de tiempos. Se representa al lado de la clave, mediante una fracción, cuyo numerador indica el número de tiempos que habrá en cada división, y el denominador indica la figura musical que define cada unidad de tiempo.
- **Duración:** Es el tiempo que se mantienen las vibraciones producidas por un sonido, que depende de la aportación de energía que realiza la fuente emisora. Y es representado por los segundos que contenga la onda del sonido, a la vista se representa mediante figuras musicales.
- **Intensidad:** Es la cualidad que diferencia un sonido suave de un sonido fuerte, lo que normalmente se llama “volumen” del sonido. Depende de la amplitud (tamaño) de la onda sonora, es decir, de la fuerza con la que sea ejecutado el sonido y de la distancia del receptor que lo percibe.
- **Intervalo:** Es la distancia que hay entre dos sonidos, esta distancia se mide por la diferencia de altura que hay entre dos tonos musicales oídos sucesiva o simultáneamente.
- **Melodía:** Es un conjunto de sonidos organizados de una forma específicamente estructurada a lo largo del tiempo para crear una expresión musical coherente. En su sentido más literal, es una combinación de alturas y ritmo.
- **Nota:** Es el elemento más básico y primordial del sonido y de la música. Una nota musical es un sonido con una determinada frecuencia, que es definida por el tono y la duración. Es representada en el pentagrama, donde la altura y la clave definen el valor de tono, y la forma de la figura define la duración.
- **Ritmo:** Es la organización de los elementos débiles y fuertes del sonido que marca una sucesión regular. Las duraciones de las notas y los silencios definen el ritmo que se presenta a lo largo de una melodía.
- **Silencio:** Es la ausencia de sonido que existe entre dos notas. Está representado por figuras musicales que indican la duración de una nota sin ejecución.

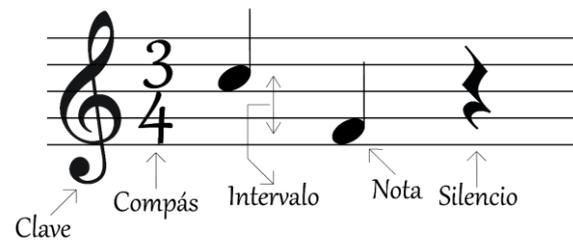


Figura 16: Pentagramas indicador de algunos conceptos musicales.

- **Tempo:** Es la velocidad a la que se interpreta una pieza musical.
- **Timbre:** Es el matiz característico de un sonido, es decir, es la cualidad que permite diferenciar un sonido de otro. Es utilizado en muchas ocasiones como criterio de identidad, permitiendo distinguir los tipos de instrumentos.
- **Tonalidad:** Define la nota sobre la que descansa una melodía. Se forma a partir de la relación que existe entre los sonidos de una escala con una nota principal.
- **Tono:** Es la cualidad que indica el número de vibraciones por segundo, es decir, la frecuencia de vibración a la que se manifiesta una determinada onda sonora. Se podría denominar como la “altura” del sonido.

Anexo II: Biografía de los participantes en las consultas realizadas

JUAN ZARZUELA

Nacido en 1980 en Jerez de la Frontera, conocido en esta ciudad como “Zarzuelita”, comienza bailando flamenco a los 6 años en la Escuela de baile flamenco de Angelita Gómez.

Se inicia en el cante a partir de los 11 años, realizando su primer recital en la peña flamenca ‘El Garbanzo’ (Jerez). Y desde entonces ha pisado diversos escenarios de distintas entidades flamencas dentro y fuera del país.

En 1993, consigue el primer premio en el Concurso de cante por Bulerías de Jerez que organiza la peña flamenca ‘Chato de la Isla’ (Cádiz). Le sigue el primer premio en el Concurso de cante flamenco de Lora del Río (Sevilla). También obtuvo el primer premio en el Concurso nacional de cante de Antequera como mejor cantaor joven. Posee también dos segundos premios en el Concurso de cante que organiza la peña flamenca de Ubrique (Cádiz) y un premio especial a la mejor saeta por seguirillas en el Concurso joven de saetas que organiza la peña flamenca ‘Torre Macarena’ (Sevilla). Tiene el honor de haber recibido el premio Carlos Cano en el festival de música de Alhama de Granada en reconocimiento a su obra “Aromas”.

En cuanto a festivales se refiere, destacamos los Viernes Flamencos de Jerez, Bienal de Sevilla, Reunión de Cante Jondo de la Puebla de Cazalla, Festival de Jerez, Festival de Mairena del Alcor, Velá de las Nieves, y un largo etc. Fuera del país ha llevado su cante a ciudades como París, Ámsterdam, Antwerpen y Lier (Bélgica), Ginebra y Lousan, Salonique (Grecia), así como diversas ciudades de Japón.

Realiza su primer trabajo discográfico con la guitarra de Diego del Morao y la producción de Moraito Chico.

En 1998, forma parte de la compañía flamenca de Carmen Cortés participando en la obra “Yerma”.

En 1999, es apadrinado por el maestro José Menese en el Teatro Albeniz de Madrid. Ese año entra a formar parte en la compañía de danza y baile flamenco de Merche Esmeralda, interviniendo en la obra “Ciclos”. También trabaja como profesor de cante en el IX Concurso Flamenco de Gerardo Nuñez y Carmen Cortés.

En 2001 lanza al mercado un trabajo discográfico en solitario debutando como compositor y autor de sus propios temas. Un trabajo cuyo título es “Entre tú, yo y el aire”.

En septiembre de 2001 comienza a trabajar en el Conservatorio de Danza de Cádiz, como cantaor oficial del centro, labor que compagina con sus actuaciones.

En 2004, saca su último trabajo discográfico titulado “Pura Verdad”, en el que resalta su perfil más jondo de este arte, dejando grabados 20 cantes distintos del amplio abanico que permite el flamenco, presentándolo en una gira por distintas ciudades de Japón, ofreciendo diversos recitales en su regreso al país. Cabe mencionar en este mismo año, su intervención en la Bienal de Flamenco de Sevilla, dentro de la programación más jóvenes más flamencos.

En 2008, entra a formar parte de Zambomba Flamenco Oficial de Jerez como director del espectáculo “Aquellos Años de Gloria”.

Además, ha sido cantaor del Conservatorio Profesional de música de Jerez y de Córdoba, donde además entra a cursar los estudios superiores de Flamencología en el Conservatorio Superior de Música.

En la actualidad, continúa con la docencia, además de sus actuaciones en los distintos espacios

flamencos, peñas, teatros, festivales y demás.

LELA SOTO

Es una de las voces jóvenes con más proyección del panorama flamenco actual. Nacida en Madrid en 1992, y la heredera más joven del legado de la casa de los Sordera. Su padre, Vicente Soto "Sordera", y su madre, Luisa Heredia, bailaora, le han transmitido desde pequeña la esencia y el amor por el Flamenco. Lela es además sobrina de artistas de la talla de José Mercé, José Soto "Sorderita" o Enrique Soto, por parte de padre, y de Ray Heredia o Enrique de Melchor, por parte de su madre.

Desde su infancia comienza a cantar en reuniones familiares y eventos, donde deja constancia de su hermosa voz y natural afición. Además de sus raíces flamencas, Lela Soto impregna en su música un flamenco vanguardista, y en su voz también se aprecian influencias de Jazz, Blues, Salsa o R&B, a través de artistas como Whitney Houston o Beyoncé, por las que tiene plena admiración.

Desde hace unos años, ha participado en grabaciones y giras de algunos artistas españoles muy importantes, como Alejandro Sanz (grabación y gira de "No es lo mismo"), Pitingo (gira de "Soulería"), Niña Pastori (grabación y gira de "Amame como soy"), Navajita Plateá, Diego del Morao, Jesús Méndez o en varios espectáculos y grabaciones de su padre y de la casa de los Sordera.

En solitario, ha grabado varios temas que han sido difundidos con éxito a través de plataformas digitales. Ha puesto música a la marca "Ron Dos Maderas" de las Bodegas Williams & Humbert y ha participado en el CD Homenaje a Lola Flores editado en 2017 por el Ayuntamiento de Jerez, interpretando "Torbellinos de Colores". En concierto, ha recorrido escenarios y peñas flamencas en distintos lugares de España y el extranjero.

En febrero de 2016, inicia una gira junto a su padre promovida por la diputación de Almería. Y en el verano de este mismo año, actúa en los Viernes Flamencos de Jerez y en la Biental de Flamenco de Sevilla. También en el mes de noviembre del mismo año, se presenta en solitario en Madrid, la ciudad donde la vio nacer, actuando en la Sala García Lorca.

En febrero de 2017, actúa en el prestigioso World Music Festival de Copenhage (Dinamarca). Este mismo año, obtuvo el primer premio en el Concurso Talento Flamenco 2017, que organiza la Fundación Crintina Heeren.

Ha participado en distintos programas de Radio y Televisión, destacando los 3 temas que grabó, acompañada de guitarra y percusión, en el Foro Flamenco de Canal Sur TV, donde interpretó Malagueñas, Bamberas y Tientos Tangos.

El cante de Lela Soto lleva la tradición flamenca inyectada en la sangre, aportando al género su dimensión artística y contemporánea de la vida, pues también es una joven de su tiempo, a la que por supuesto no le es indiferente lo que sucede a su alrededor. Sin duda, una voz que nos anticipa el flamenco del mañana, pero que afortunadamente disfrutamos hoy.

PEPE ALCONCHEL

José Alconchel García, nacido en Arcos de la Frontera (Cádiz) en 1949, aunque su vida la desarrolla desde muy joven en Jerez de la Frontera por motivos laborales.

Cantaor payo (no gitano), de siempre le atrajo el cante flamenco y es por lo que estudia, conoce e interpreta todos los estilos del cante, llegando a fundar una peña flamenca con su nombre artístico, Pepe Alconchel.

En el año 1989, hace su primera grabación bajo el título “Con la fuerza del sentimiento”. Este mismo año, la Conferencia Andaluza de Peñas Flamencas le incluye en la grabación que hacen para el recuerdo del primer Concurso Regional Andaluz, grabando alegrías, seguirillas y bulerías en representación de la provincia de Cádiz.

En 1998, graba en solitario “Peña y alegría en mi cante” con letras de Antonio Murciano.

En el año 2001, el Ayuntamiento de Arcos de la Frontera le incluye en las 2 grabaciones que hace, una bajo el título “Cantes y Cantaores de Arcos” y otra con el título “Cantes Cabales”, acompañado a la guitarra por Miguel Chamizo.

En abril de 2002, da recitales en las ciudades de Reims, Lille y Grenoble, pasa también por “La Peña Flamenca de París” y por el Teatro de la música; ese mismo año vuelve a pisar Lille donde con grandes y prolongados aplausos le recuerdan la primera actuación.

En 2008, obtuvo el primer premio “Uva de Oro” en el XIV Concurso Nacional de Cante que se celebraba en el teatro Vico de la ciudad de Jumilla (Murcia), con cantes por Seguirillas, Soleá, Mineras y Cartageneras, premio que obtuvo por primera vez en el año 1997. Ese mismo año, 2008, también realizó el programa “Nuevos ciclos de cante en la Bulería y la tertulia de Pepe Alconchel”, ofreciendo un interesante recital de cante en el Pago de San José (Jerez) con la guitarra de Pascual Lorca.

En 2009, gana el primer premio en el Certamen de Saetas de San José del Valle en la IV edición del Concurso Flamenco de Saetas “Villa de Alcalá del Valle”.

En 2010, gana el Certamen de Cante de Coria, recibiendo el primer premio del Concurso de Cante de Coria del Río (Sevilla). Este mismo año, ofreció recitales en el Puerto de Santa María (Cádiz), Logroño y Huesca, apareciendo también en el programa de los encuentros flamencos sin fronteras “Andaluces por el mundo”. También este año presentó su nuevo disco “1ª Antología del Fandango” acompañado por Chamizo y Pascual de Lorca a la guitarra, donde muestra un excelente trabajo recordando algunos de los fandangos que nos dejaron historia en el arte del cante flamenco.

En 2013, fue ganador del primer premio en el “XXXVIII Concurso Nacional de Cante por Peteneras” en Paterna del Campo (Huelva).

En 2017, obtiene por tercera vez el premio Nacional de cante por Serranas en la localidad gaditana de Prado del Rey.

Destaca por haber recibido más de 100 galardones en festivales flamencos de todo tipo. Su cante es una interpretación seria del flamenco que convence al que sabe y deleita este arte, y que hace que el público pueda recordar a aquellos grandes maestros.

EVA RUBICHI

Eva María Sánchez Navarro, de profesión cantaora, nacida en Jerez de la Frontera, ligada a una de las ramas más puras del mundo del flamenco, los Agujetas, familia de su marido y acompañante de guitarra, Domingo Rubichi, y por este último es por el que recibe su nombre artístico Eva Rubichi.

Desde muy pequeña se dedica al mundo del espectáculo, ha trabajado cantando para recitales en peñas flamencas como Los Cernícalos, La Bulería, Antonio Chacón, Tío José de Paula, y un largo etc. Ha intervenido en festivales como los Viernes Flamencos, Jueves de Cádiz, Fiesta de la Bulería, Festival de Jerez, Fiestas de Otoño, etc. También ha participado en espectáculos como “Viva Jerez” en el Teatro Villamarta, Zambomba Flamenca de Parrilla de Jerez, etc, todo ello dentro de la ciudad de jerez. A nivel nacional ha actuado en ciudades como Madrid, Bilbao y Barcelona. En Europa, en países como Francia, Suiza, Yugoslavia e Italia. Y a nivel mundial, han tenido el placer de verla actuar países como Japón y EE.UU.

Ha trabajado como cantaora en los cursos del Festival de Jerez, acompañando las clases de maestros como Angelita Gómez, Manolo Marín, Manolete, Rocío Molina, Rafael Campallo, Concha Vargas, Carmen Ledesma o Blanca del Rey.

Actualmente, sigue en activo en peñas flamencas, festivales, grabaciones de televisión, y discografía, y de todo ello le aparece un sentimiento de orgullo por su trabajo. Su manera de cantar corresponde a las numerosas vivencias por las que ha pasado en infinitas fiestas en San Miguel, barrio de la ciudad jerezana que la vio nacer. Es una artista que maneja todos los estilos del cante flamenco y que día a día sigue aprendiendo para profundizar en este arte.

RAÚL GONZÁLEZ

Raúl González López, 30 años, nacido en el barrio de San Jerónimo (Sevilla), apasionado del flamenco, vinculado a este arte gracias a su ciudad, de donde nacen muchos artistas del mundo del flamenco, de los cuales tiene el honor de contar con ellos entre su círculo más cercano de amistades.

Desde muy pequeño le ha gustado cantar flamenquito, pero hace 10 años aproximadamente que se aficionó al “cante puro”, como él define al cante jondo.

Es cantante de flamenco aficionado, no se dedica profesionalmente a ello, lo hace por placer y como forma de expresión.

Comenzó aprendiendo este arte a través de las relaciones con artistas del flamenco, escuchando sus cantes, sus consejos y tras esto, poniéndolo en práctica con toda su pasión, pues el cante le hace sentir desahogo, una satisfacción personal que lo deja en un estado de paz y tranquilidad.

Cuenta que sus inicios fueron en “El Corralón”, un bar que había dentro de los corralones en pleno centro de Sevilla, hoy día desalojado, donde conoció a muchos artistas flamencos, algunos de ellos son hoy día sus amigos. Ahí fue donde entabló sus primeras amistades del mundo del cante y donde comenzó su afición.

Actualmente sigue con la práctica y el aprendizaje dentro de su afición por el cante, escuchando de primera mano a sus amigos tanto artistas, como aficionados con los que se codea.

Uno de los lugares de los lugares de encuentro es el “Lola Cacerola”, un bar/restaurante situado en la calle Castilla, en pleno corazón de Triana, donde se reúnen personas que les gustan el arte flamenco, siempre hay alguien cantado, si no son grupos consolidados, son aficionados que hacen

compás con la barra del bar, la cosa es que nunca se para de cantar. De hecho, es un restaurante espectáculo donde se organizan espectáculos de flamencos de jueves a domingo, y el resto de la semana hay espectáculos de flamenco improvisado en un ambiente familiar, en el que por supuesto, Raúl González participa y pertenece a esta gran familia, y todo ello, a orillas del Guadalquivir, el río como principal espectador.

ANGÉLICA MONTAÑÉS

Angélica Montañés, 25 años, nacida en Sevilla, apasionada del cante flamenco. Vinculada al cante flamenco por sus amistades y del cual al día de hoy es una gran aficionada. Sin ninguna formación profesional de canto, pero practica y aprende gracias a los encuentros en “Lola Cacerola”, en el barrio de Triana pasa las horas haciendo una de las cosas que mas le gusta y poco a poco va conociendo más este arte.

ÁLVARO LINARES

Álvaro Linares, 27 años, nacido en el barrio de Triana (Sevilla), aficionado al cante flamenco desde pequeño, aunque no haya recibido formación, si es un gran aficionado que intenta aprender cada día y practicar el cante en reuniones con su grupo de amistades que comparten la pasión por este arte.