

Un algoritmo eficiente para estudiar la similitud melódica de los cantes flamencos

J. C. Rizo-Massia

J. M. Díaz-Báñez

Resumen

80

En este trabajo se estudia la similitud melódica entre cantes flamencos usando el paradigma *Dynamic Time Warping* como medida. Centramos nuestro estudio en un grupo de estilos en particular, las tonás, cantes flamencos a capela, caracterizados por ser cantes *ad libitum* y por su alto grado de complejidad en la ornamentación. Se muestra que esta medida de similitud discrimina correctamente variaciones entre estilos. Con relación a la aceleración de los algoritmos de similitud clásicos, nuestra estrategia calcula una segmentación eficiente del contorno melódico previa a la aplicación de la medida de similitud. Se demuestra que nuestro método obtiene mejores resultados (tanto en eficiencia como en precisión) que otros algoritmos existentes.

1. Introducción

Pese a la existencia del flamenco desde hace más de un siglo y al gran arraigo cultural y social que posee, no ha sido hasta hace relativamente poco cuando se ha querido someter este arte al juicio de la ciencia. Posiblemente, este hecho ha estado motivado por el carácter anárquico y sentimental del propio cante flamenco, ya que, aunque sus formas y estilos están claramente definidos, la tradición oral ha hecho que estos se superpongan y mezclen de tal forma que resulta muy complejo establecer un patrón fijo e inamovible de cada palo.

El flamenco es la música tradicional más importante en Andalucía, y goza además, de gran interés internacional. La música flamenca es fundamentalmente de tradición oral, lo que dificulta, y hace más interesante a la vez, su estudio científico. Por otra parte, la conservación de esta música de generación en generación hace que la melodía juegue un papel crucial en la evolución y clasificación de los distintos estilos del flamenco, siendo este uno de los problemas abiertos más apasionantes del área.

La Teoría Computacional de la Música es un campo de creación relativamente reciente. La música y las matemáticas han estado relacionadas desde al menos 2.500 años, cuando Pitágoras descubrió que los intervalos más consonantes provenían de proporciones simples y pequeñas. Desde entonces, la interacción entre música y matemáticas no ha hecho sino crecer. Y ha sido en las últimas décadas cuando se ha producido una revitalización como consecuencia de la introducción de los métodos computacionales en el análisis de la música. Las técnicas computacionales y modelos matemáticos utilizados en el área de investigación de la Teoría Computacional del flamenco, pertenecen al campo de la Music Information Retrieval (MIR)¹.

81

Por su parte, la clasificación de los estilos flamencos con un procedimiento automático está sujeto a diversas dificultades (incluso manualmente). Sin embargo, aunque dos interpretaciones de un mismo palo puedan sonar muy diferentes para un oído no acostumbrado al flamenco, detrás de cada estilo existe un patrón o esqueleto melódico que tiene memorizado el cantaor, de tal suerte que, cuando interpreta el cante, lo impregna de ornamentación y melismas propios del flamenco y que hacen bastante compleja la detección de tal patrón definitorio del estilo. De cualquier modo, resulta evidente que encontrar un método efectivo y eficiente de cálculo de similitud melódica en el cante flamenco no deja de ser un reto apasionante y complejo.

1 N. Orio, «Music retrieval: a tutorial and review», *Foundations and Trends in Information Retrieval*, vol. 1 (2006), pp. 1-90.

Existen trabajos actuales donde se proponen medidas de similitud melódica en el contexto del flamenco. Tal es el caso de los trabajos del grupo COFLA². Así en uno de ellos³ se propone el uso de una distancia que combina el uso de la *distancia de edición* sobre dos contornos melódicos y una *distancia vectorial* sobre una cadena simbólica de variables musicales previamente definidas para los cantes de tonás.

En este trabajo presentamos un procedimiento novedoso de cálculo de similitud melódica que permite la distinción de distintas variantes del flamenco. Lo aplicamos aquí a los cantes por Tonás, concretamente Debblas y Martinetes, pero puede usarse obviamente a otros estilos menos ornamentados.

2. El método del DTW

82

Este algoritmo de similitud conocido por sus iniciales DTW⁴ se aplica cuando las dos secuencias que se van a comparar pueden variar en tiempo o velocidad, es decir, cuando ambas secuencias tienen diferente longitud. Es muy utilizado en aplicaciones de reconocimiento automático del habla, donde el sistema puede recibir una misma palabra enunciada a diferentes velocidades.

En nuestro caso, las secuencias de entrada serán las dos melodías que queremos comparar. En general, el funcionamiento del algoritmo se basa en la búsqueda de un camino óptimo de coincidencia entre dos secuencias con ciertas restricciones. Las secuencias se transforman de forma no lineal en el tiempo, de modo que son comprimidas o expandidas en el tiempo para que tengan el mismo tamaño, y así poder compararlas

2 <http://mtg.upf.edu/research/projects/cofla>.

3 J. Mora; F. Gómez; E. Gómez; F. J. Escobar-Borrego; J. M. Díaz-Báñez, «Characterization and melodic similarity of a Cappella flamenco cantes», International Society for Music Information Retrieval Conference, ISMIR, 2010.

4 H. Sakoe; S. Chiba, «Dynamic programming algorithm optimization for spoken word recognition. Acoustics, Speech and Signal Processing», *IEEE Transactions*, 26 (1) (1978), pp. 43-49.

punto a punto. Una variante del algoritmo DTW, conocida como *Context-Dependent Dynamic Time Warping*⁵ fue utilizada para la localización de patrones melódicos predefinidos y diferenciadores de estilos en el contexto de los fandangos de Huelva.

El alto grado de complejidad del método DTW, tanto en tiempo como en espacio, crea la necesidad de establecer estrategias o métodos que aceleren el comportamiento del algoritmo DTW original. Se han realizado muchos esfuerzos en este sentido para la búsqueda de similitud de cadenas simbólicas y numéricas bajo el paraguas del algoritmo DTW. Dos de las propuestas más usadas son las que se basan en establecer restricciones en las trayectorias de búsqueda, como son las bandas de Sakoe-Chiba o el paralelogramo de Itakura. Recientemente se ha propuesto un algoritmo más rápido, FastDTW⁶, que obtiene una aproximación mediante la proyección de trayectorias que calcula el método DTW para distintos niveles de reducción.

3. La segmentación. El algoritmo FITTING

83

El algoritmo de aproximación FITTING⁷ fue propuesto para simplificar cadenas ortogonales. Proponemos aquí el uso de este método para segmentar las melodías del cante flamenco. Aplicando este algoritmo sobre la curva melódica, se obtiene una representación discreta de la curva original (segmentación) de mínima longitud para un error o umbral fijado de antemano. Teniendo en cuenta que las máximas variaciones de tono en el cante flamenco están contenidas en una octava musical, la elección de valores de umbral comprendidos entre uno y medio semitono parece adecuada para capturar el esqueleto melódico de la pieza (ver Figura 1).

5 A. Pikrakis; F. Gómez; S. Oramas; J. M. Díaz-Báñez; J. Mora; F. Escobar; et al., «Tracking Melodic Patterns in Flamenco Singing by Analyzing Polyphonic Music Recordings», *ISMIR*, 2012, pp. 421-426.

6 S. Salvador; P. Chan, *FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space*. *KDD Workshop on Mining Temporal and Sequential Data*, 2004, pp. 70-80.

7 J. M. Díaz-Báñez; J. A. Mesa, «Fitting rectilinear polygonal curves to a set of points in the plane», *European Journal of Operational Research*, 130, (1) (2001), pp. 214-222.

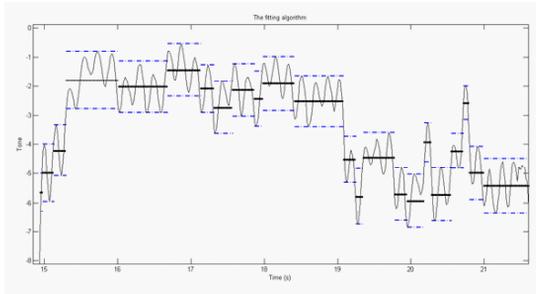


Figura 1: Segmentación de la frecuencia fundamental f_0 realizada por el algoritmo fitting.

A continuación se describe el algoritmo sin entrar en detalles técnicos. Dado un conjunto de puntos $P=\{p_1, p_2, \dots, p_n\}$ en un espacio bidimensional y una tolerancia al error α , se trazan segmentos verticales (ventanas) V_i de longitud 2α centrados en cada punto p_i . La restricción fijada de que cada punto de la curva melódica inicial esté dentro del intervalo V_i es equivalente a decir que la función escalonada (segmentación) intercepta cada uno de estos segmentos de longitud 2α . Barriendo de izquierda a derecha, el algoritmo trata de interceptar tantas ventanas consecutivas como sea posible, antes de empezar un nuevo escalón y repetir el procedimiento.

84

El segmento vertical V_i define un intervalo $Y=[y_i^+, y_i^-]$ donde y_i^+ e y_i^- denotan las coordenadas del punto final superior y inferior de y respectivamente. Recorriendo de izquierda a derecha se mantiene la intersección Δ de los intervalos hasta que se llega a un segmento V_j no interceptado por Δ . En este caso se termina la construcción del segmento, y se empieza un nuevo escalón (segmento horizontal).

El tiempo computacional (rapidez) de este algoritmo es excelente y construye una cadena de segmentos horizontales de mínima longitud que representa la segmentación de la melodía con una tolerancia α .

4. Nuestra estrategia

La ineficiencia del método clásico DTW, tanto en tiempo como en espacio, crea la necesidad de establecer estrategias o métodos que acele-

ren el comportamiento del algoritmo DTW original.

A la hora de enfocar una estrategia para mejorar la complejidad del algoritmo DTW, es posible considerar dos alternativas para diseñar una medida de similitud melódica:

1. Aplicar alguna estrategia de simplificación del DTW, por ejemplo, FastDTW, bandas de Sakoe-Chiba, o alguna otra, y obtener una medida de similitud aproximada asumiendo los errores que ello conlleva.
2. Aplicar el algoritmo clásico del DTW sobre una aproximación de los datos de entrada (de la melodía) calculada mediante algún método de segmentación.

En este trabajo se adoptará la segunda estrategia. Por lo tanto, nuestro objetivo se centra en discretizar la envolvente de la frecuencia fundamental extraída con herramientas existentes (Smstools2, Melody, Mami o similares) en pequeños segmentos o secuencias, de modo que las medidas de similitud no se alteren globalmente. La complejidad temporal de los algoritmos de segmentación utilizados varía desde el orden de la duración de la propia pieza (muy lento), en el caso de usar la segmentación obtenida por el Smstools2, hasta la gran velocidad ofrecida por el método basado en el algoritmo Fitting.

85

A modo de resumen la estrategia combinada que proponemos se llevará a cabo en los siguientes pasos:

1. Obtención de la frecuencia fundamental (que llamamos f_0) de las piezas a analizar mediante métodos de análisis de *pitch*⁸.
2. Pre-tratamiento de la frecuencia fundamental para la eliminación de silencios y posibles datos espurios.
3. Normalización de la frecuencia fundamental a una frecuencia de referencia.

8 E. Gómez; J. Bonada, «Towards computer-assisted flamenco transcription: An experimental comparison of automatic transcription algorithms as applied to a cappella singing», *Computer Music Journal*, 37 (2) (2013), pp. 73-90.

4. Segmentación usando el método elegido.
5. Aplicación del algoritmo DTW y obtención de la matriz de distancias.

Describimos brevemente a continuación las etapas del método:

4.1. Primer paso: obtención de la frecuencia fundamental

Para la obtención de la frecuencia fundamental (f_0) se hará uso de métodos de análisis de *pitch*, con él se obtendrá a partir de la grabación de audio en formato *.wav, en mono, con una frecuencia de muestreo de 44.1 kHz.

4.2. Segundo paso: pre-tratamiento de la frecuencia fundamental

Los datos extraídos poseen silencios y datos espurios que no aportan información y distorsionan la forma del contorno melódico que se va a analizar, por lo que se realizará un pretratamiento de los datos para así tener un perfil melódico correcto. Los silencios y los datos espurios del tipo «valores negativos», han sido eliminados de forma automática, mientras que los datos espurios de tipo «valores excesivamente altos», han sido eliminados de forma manual mediante la inspección del contorno.

4.3. Tercer paso: cuantización y normalización

Al estar los datos de la frecuencia fundamental en Hz, debemos normalizarlos para poder comparar todos los cantes en base a una frecuencia de referencia. La frecuencia de afinación tomada como referencia es 440 Hz, que en el código de notas MIDI representa la nota LA 4ª. Para normalizar la frecuencia fundamental, se realizará una conversión a semitonos, tomando como se ha dicho como frecuencia de referencia 440 Hz.

4.4. Cuarto paso: segmentación

Como se ha comentado en la introducción a la metodología a seguir,

la base fundamental de la estrategia escogida es simplificar los datos de entrada al algoritmo DTW mediante una segmentación del contorno melódico. Dicha segmentación se realizará mediante la implementación en código de Matlab del algoritmo Fitting.

La segmentación obtiene una representación discreta mínima de la curva de contorno melódico para un rango fijado de tolerancia al error α , el cual marca el grado de aproximación de la representación discreta a la real continua, esto es, a menor α más similar es la aproximación a la real, conteniendo mayor número de componentes. A su vez, a mayor α se obtendrá una aproximación con mayor error pero con un menor número de componentes. En este escenario, se ha de encontrar una solución de compromiso entre la eficiencia y la eficacia del algoritmo al establecer un α determinado. Dicho problema se resolverá en la próxima sección, al realizar pruebas con distintos α y determinar los tiempos de computación y los errores que éstos poseen. La resolución mínima que se considera para el parámetro α es de medio semitono, es decir, la ventana se fija con una altura de 1 semitono. En la siguiente figura se representa la segmentación de una Debla interpretada por Antonio Mairena.

87

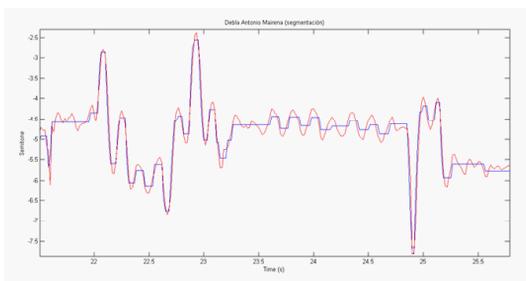


Figura 2: Detalle de la segmentación.

A medida que se disminuye el parámetro α , la aproximación de la función discreta (segmentación) se ciñe mejor a la función continua (contorno melódico), pero por el contrario, como se ha dicho antes, el número de componentes, o segmentos por los que la representación discreta aproxima al contorno melódico, aumenta. La tabla siguiente muestra un ejemplo del cambio de número de componentes de la segmentación de

una Debla interpretada por Antonio Mairena.

α	Nº de segmentos obtenidos
2	70
1	174
0,5	398

Tabla 1: Número de segmentos para cada tolerancia al error.

4.5. Quinto paso: aplicación del algoritmo DTW y obtención de la matriz de distancias

En este paso de la estrategia se llega a lo novedoso del algoritmo SDTW: la aplicación del algoritmo DTW clásico sobre el contorno segmentado de los audios que forman el corpus de cantes elegidos. Con ello, se calculan las distancias o diferencias numéricas entre los cantes del corpus.

5. Medidas de rendimiento

88

Con el fin de evaluar y comparar nuestro algoritmo con otras estrategias usaremos diferentes medidas de rendimiento o bondad. Las describimos brevemente:

- Una representación gráfica de las distancias obtenidas mediante la creación de árboles filogenéticos. Se hará uso de técnicas de Bioinformática⁹ que son capaces de construir un árbol que refleja las relaciones entre sus miembros. Un análisis similar se realizó por primera vez para la música flamenca en un artículo sobre la evolución rítmica del flamenco¹⁰.
- Test de Mantel¹¹ para cuantificar la correlación entre las matrices de distancias para cadenas melódicas no segmentadas y segmentadas.

9 D. H. Huson, «SplitsTree: analyzing and visualizing evolutionary data», *Bioinformatics*, 14 (1) (1998), pp. 68-73.

10 J. M. Díaz-Bañez; G. Farigu; F. Gómez; D. Rappaport; G. T. Toussaint, «El compás flamenco: a phylogenetic analysis», en *Proceedings of BRIDGES: Mathematical Connections in Art, Music and Science*, 2004, pp. 61-70.

11 N. Mantel; R. S. Valand, «A technique of nonparametric multivariate analysis», *Biometrics*, vol. 26, no. 3 (1970), pp. 547-558.

Este análisis de correlación provee como resultado la correlación entre dos matrices y se acerca a 1 en caso de fuerte correlación.

- De otro lado, el uso de las métricas Precision & Recall¹² y F-measure¹³ para cuantificar la bondad, en términos de precisión y efectividad, de la clasificación generada. Estos parámetros se acercan a 1 en caso de concordancia completa entre la clasificación dada por la estrategia y la clasificación propia del corpus TONÁS¹⁴, que se tomará como *Ground Truth* (base cierta anotada manualmente por expertos de flamenco).
- Para cuantificar la eficiencia del método usado en cada procedimiento, mediremos el tiempo que consume la computadora es completar la estrategia completa.

Para el cálculo de las métricas Precision & Recall y F-measure, se establecen términos usuales en tareas de clasificación (verdadero positivo, falso positivo, falso negativo y verdadero negativo) en base a comparar la *predicción de un método de clasificación* frente a un juicio externo o *Ground Truth*. Finalmente, para establecer la clasificación de un determinado audio usaremos el método de clasificación supervisado *k-NN*¹⁵ con $k=3$, esto es, diremos que un audio corresponde a un martinete si los 3 audios más cercanos a él son martinetes.

89

6. Resultados con el corpus de tonás: precisión, efectividad y eficiencia

Evaluamos nuestra estrategia en una colección de música o corpus, compuesto por 24 fragmentos representativos de dos estilos de cantes a capela (12 deblas y 12 martinetes) seleccionados del corpus TONÁS.

12 C. J. van Rijsbergen, «A new theoretical framework for information retrieval», en *Proceedings of the 9th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, 1986, pp. 194-200.

13 C. J. van Rijsbergen, «Foundation of evaluation», *Journal of Documentation*, 30 (4) (1974), pp. 365-373.

14 <http://mtg.upf.edu/download/datasets/tonas>.

15 N. S. Altman, «An introduction to kernel and nearest-neighbor nonparametric regression», *The American Statistician*, 46 (3) (1992), pp. 175-185.

Esta colección monofónica fue elaborada por el grupo Cofla en el marco de un estudio sobre la similitud y clasificación de diferentes estilos de cantes a capela. Los archivos se transcribieron manualmente para así generar un *dataset* (conjunto de datos) fiel a la realidad.

Se han realizados distintas pruebas combinando los siguientes métodos de segmentación y cálculo de similitud que se exponen a continuación:

- **Procedimiento A:** DTW sobre la frecuencia fundamental no segmentada.
- **Procedimiento B:** Segmentación usando el algoritmo Fitting con alfa desde 2 hasta 0,5 y aplicación del método DTW sobre ella.
- **Procedimiento C:** Método aproximado DTW Sakoe-Chiba con ancho de banda: 250, 100, 25.

Mostramos en primer lugar los árboles filogenéticos que se generan a partir de la matriz de distancias de los distintos métodos que se obtienen con la herramienta SplitTree.

90

• Árboles filogenéticos:

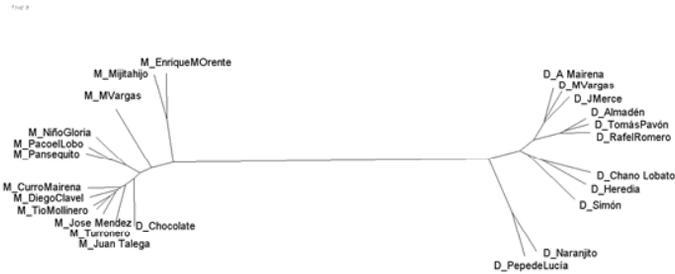


Figura 3: Árbol filogenético procedimiento A.

El árbol de la figura 3, muestra una perfecta discriminación entre los dos estilos tratados, salvo el caso de la debbla de Chocolate, que, de hecho, es razonable su localización porque el cantaor realiza una interpretación alejada del canon de la debbla.

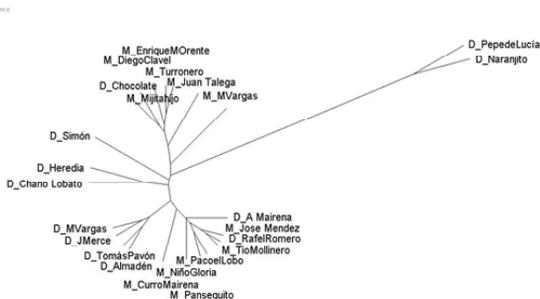


Figura 4: Árbol filogenético procedimiento B con $\alpha=2$.

La figura 4 muestra que con una tolerancia al error de 2 semitonos, el algoritmo SDTW no es capaz de establecer una discriminación entre estilos.

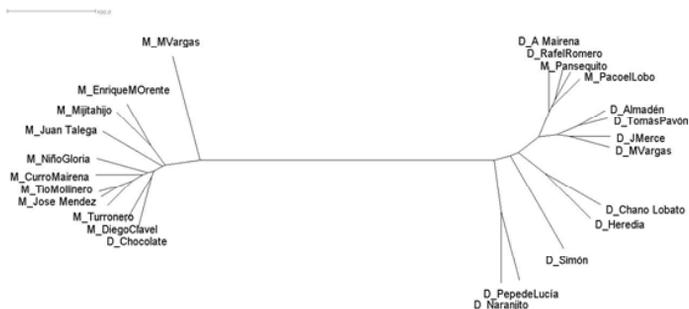


Figura 5: Árbol filogenético procedimiento B con $\alpha=1$.

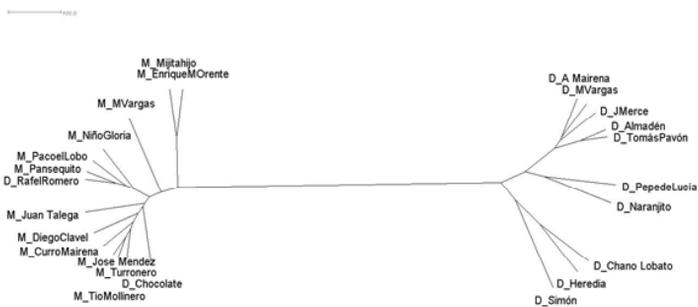


Figura 6: Árbol filogenético procedimiento B con $\alpha=0,5$.

Las figuras 5 y 6, muestran unas clasificaciones de estilos aceptables encontrando como máximo dos elementos mal clasificados.

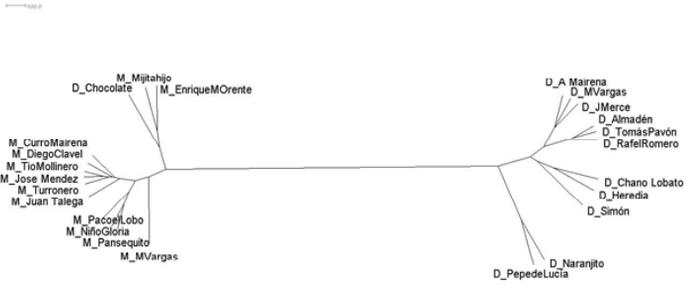


Figura 7: Árbol filogenético procedimiento C con ancho de banda=250.



Figura 8: Árbol filogenético procedimiento C con ancho de banda=100.

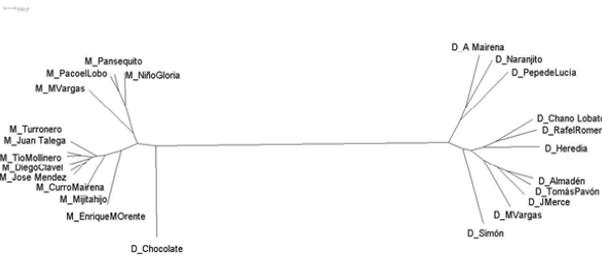


Figura 9: Árbol filogenético procedimiento C con ancho de banda=25.

Los árboles filogenéticos obtenidos mediante el procedimiento en el que aplicamos el método de aproximación de Sakoe–Chiba, la clasificación de estilos es correcta pues sólo encontramos como elemento mal clasificado la interpretación de la debla grabada por Chocolate.

- Métricas de precisión, efectividad y eficiencia:

En la siguiente tabla mostramos los parámetros que miden la calidad de las distintas estrategias:

		Mantel	Precisión	Recall	F-measure	Tiempo (s)
PRUEBA A		1	0.94	0.67	0.78	2.29×10^3
PRUEBA C	r=250	0.98	0.95	0.75	0.91	699.09
	r=100	0.93	0.95	0.95	0.95	578.11
	r=25	0.90	0.94	0.91	0.93	478.56
PRUEBA B	$\alpha=2$	0.82	0.80	0.33	0.47	13.75
	$\alpha=1$	0.91	0.93	0.54	0.68	71.26
	$\alpha=0,5$	0.95	0.94	0.62	0.75	246.06

Tabla 2: Tabla resumen.

8. Conclusiones

A la luz de los datos de la tabla 2 podemos concluir lo siguiente respecto a los distintos procedimientos seguidos:

- Procedimiento A: Este procedimiento proporciona altos valores para las distintas métricas de precisión y efectividad. En términos de eficiencia es el procedimiento que más tiempo consume en completarse y, por tanto es inoperante para trabajar con un corpus amplio de cantes flamencos.
- Procedimiento C: Los resultados de precisión y efectividad son incluso mejores que en el procedimiento A, lo cual parece contradictorio, pues aquí se calcula una aproximación. Esto puede estar motivado porque para los cantes estudiados la restricción a una banda no penaliza el cálculo del camino óptimo que calcula el método DTW. Sin embargo, esto no se generaliza para todos los cantes. Con relación a la eficiencia, al restringir las trayectorias el número de elementos con los que tiene que operar el algoritmo DTW se mejora el tiempo de ejecución con respecto al procedimiento A pero es menos eficiente que el procedimiento B.

- Procedimiento B: Es el procedimiento que proponemos para el problema de la detección eficiente de similitud melódica en el cante flamenco. Destacamos la rapidez de ejecución para el caso $\alpha=1$, para el cual los valores de precisión y efectividad son bastante aceptables.

Para finalizar, concluimos que se ha propuesto en este trabajo un nuevo algoritmo eficiente para el cálculo de similitud melódica para cantes ornamentados. Está basado en el conocido algoritmo DTW y permite realizar discriminación entre cantes del grupo de las tonás. El método supone aplicar una estrategia combinada que ejecuta una simplificación de la melodía con anterioridad a la aplicación del propio algoritmo DTW como medida de similitud. Los resultados son prometedores tanto en eficiencia como en eficacia y pueden extenderse a otros palos del flamenco donde se deben obtener incluso mejores resultados para los cantes que posean menor grado de ornamentación.