1. Introducción

En el primer artículo de esta serie revisamos el concepto matemático de distancia y definimos formalmente la medida de similitud melódica; además, ilustramos estos conceptos con algunos ejemplos musicales. En este segundo artículo sobre la distancia matemática y la similitud melódica vamos a describir un buen número de medidas de similitud, las cuales agruparemos por familias dependiendo de su propósito y filosofía.

"El discurso musical progresa a través de múltiples transformaciones de su material musical", decíamos en otra ocasión [Góm11]. En particular, en el ámbito melódico esa variación es fundamental. La trascendencia de la similitud melódica ha sido, y es, tal que numerosas disciplinas se han ocupado intensivamente de su estudio:

- En etnomusicología, por ejemplo, para entender la lógica musical, para evaluar los estilos y sus características, para conocer los criterios de improvisación. Véanse [BL51], [See66], [Hol10].
- En análisis musical, para construir modelos tanto teóricos como computacionales. Véanse [LJ83], [Mey73], [CIR98], [Typ07].
- En la resolución de conflictos de propiedad intelectual [Cro98].
- En tecnología musical, para aplicar los modelos obtenidos tras el correspondiente análisis. Véanse [MS90], [HSF98], [Pam06], [Hol10].
- En psicología de la música, para comprender mejor el hecho musical, para aportar conocimiento a un análisis integral de la música. Véanse [Sch99], [HE01], [MM01], [HE02], [HCR03].

Respecto a otro tema relacionado, la similitud rítmica, véanse las referencias citadas en el artículo [Góm11] de esta sección del mes de marzo de 2011.

2. Representación de melodías

Una melodía es un conjunto de notas que suenan en determinado orden y con determinadas duraciones. Esencialmente, una melodía es una combinación de alturas y ritmo. Las medidas de similitud se construyen a partir del siguiente proceso:

Figura 1: **El proceso de construcción de una medida de similitud.**

1. Obtención de la melodía. Esta puede venir descrita en formato audio, como un fichero de sonido de una grabación, o en formato simbólico, como un fichero midi o un fichero de partitura (Finale, Sibelius, etc.).
2. Representación de melodías abstractas. Según los propósitos perseguidos, la melodía se representa de varias formas. Estas representaciones que extraen ciertas características de las melodías reales se llaman representaciones abstractas. Varían en función del propósito final. Por ejemplo, en el análisis de cantes a palo seco en el flamenco, se ignora la duración de las notas, puesto que son cantes sin pulso regular y con mucho *rubato*.
3. Transformación de la melodía. La melodía abstracta sufre unas transformaciones que permiten el cálculo efectivo de la medida de similitud.
4. Diseño de la medida de similitud. Inicialmente, la mayor parte de las medidas se concentraron en un aspecto de las melodía, el cual medían con más o menos precisión. Poco a poco se han ido diseñando medidas que cuantifican varios aspectos de la melodía. El principal problema es cómo ponderar todos esos aspectos de manera coherente.

Es importante que la representación de las melodías cumpla las propiedades de invariancia que mencionamos en el primer artículo, esto es, invariancia por transposición de altura y tiempo más invariancia por cambio de tempo. Las melodías se suelen representar por una sucesión de pares (p_n,t_n), n = 1,…,N, donde p_n representa la altura y t_n las duraciones de cada nota. Esta representación obviamente no verifica las propiedades de invariancia. Se usan en su lugar dos representaciones diferentes, que sí respetan la invariancia:

- Para la altura se usa la representación por intervalo. En lugar de guardar las alturas absolutas, se anota el intervalo entre cada dos notas consecutivas $I_n = p_{n+1} - p_n$.
- Para el ritmo se usa la representación IOI. Estas siglas vienen del inglés, *inter-onset interval*, o intervalo de duraciones relativas. Se calcula de manera similar al caso anterior, poniendo $T_n = t_{n+1} - t_n$.

En el caso del ritmo también se usa otro método de representación, que consiste en expresar las duraciones en función de la duración mínima de la melodía. Esto no siempre es posible en todas las músicas, aunque sí en la mayor parte de la música occidental.

En las siguientes secciones del artículo describiremos las principales transformaciones melódicas que aparecen en las medidas de similitud.

3. Transformaciones de altura

3.1 Transformaciones de contorno

Este tipo de transformaciones se basa en el hecho de que la sucesión exacta de alturas en una melodía no es siempre lo más importante, sino la dirección melódica. Los puntos de giro de la melodía, los puntos en que cambia la dirección melódica, es un hecho que tiene relevancia en la percepción de la melodía. Si se representa la melodía como una línea poligonal, los puntos de giro se corresponden con los extremos relativos. La representación recibe el nombre de contorno melódico. En la figura de abajo tenemos de nuevo el tema principal de las variaciones K. 265 *Ah, vous dirai-je, Maman*. La línea poligonal que aparece sobre la melodía es el contorno melódico.
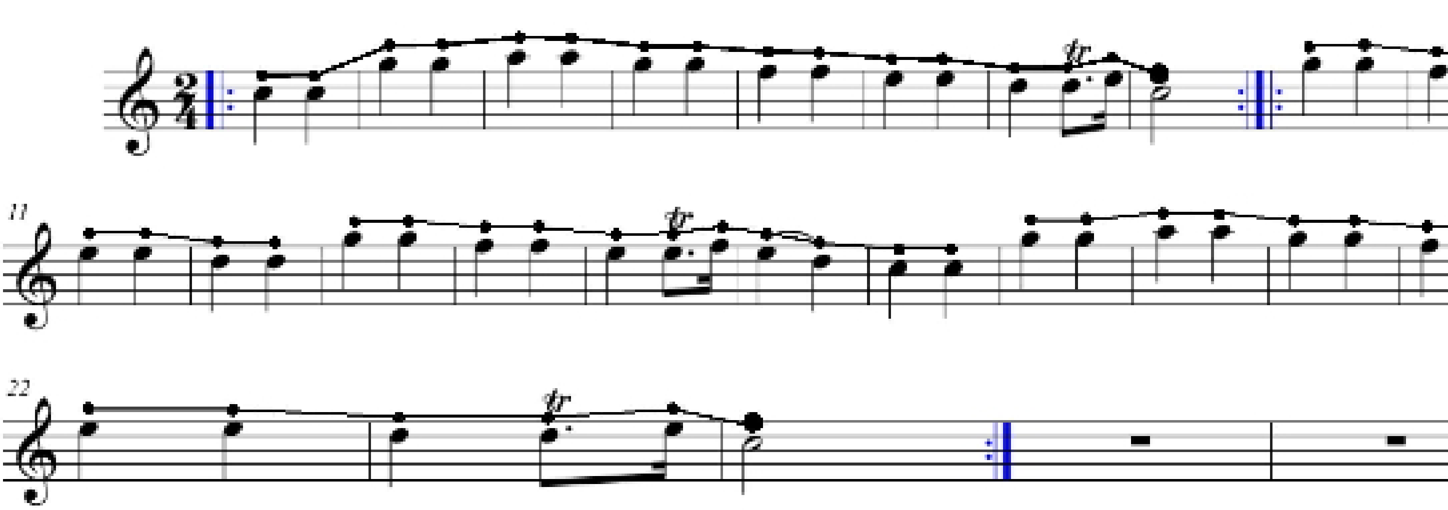
Figura 2: **El contorno melódico.**

Para más información sobre los contornos melódicos véase el trabajo de Zhou y Kankanhalli [YK03].

3.2 Transformaciones borrosas

En muchas ocasiones la percepción de un estímulo no es totalmente nítida. Este tipo de transformaciones pretenden modelizar este hecho usando la lógica borrosa. La lógica borrosa usa un continuo de valores de verdad en el intervalo [0, 1]; consúltese [Hal03]. En este modelo los intervalos se acomodan en clases según la tabla siguiente:

Clase	Intervalos	Nombre
-4	< -7	Salto descendente grande
-3	-7,-6,-5	Salto descendente
-2	-4,-3	Paso descendente grande
-1	-2,-1	Paso descendente
0	0	Unísono
1	1, 2	Paso ascendente
2	3, 4	Paso ascendente grande
3	5, 6, 7	Salto ascendente
4	> 7	Salto ascendente grande

Tabla 1: **Clasificación de los intervalos usando lógica borrosa.**

Los intervalos se cuentan por semitonos en la tabla. Para ilustrar este concepto, consideremos la melodía abstracta dada por la sucesión $\{(p_i,t_i)\}$:

{(64,1),(64,1),(71, 1),(71, 1),(73,1),(73,1),(71,1),(71,1), (69, 1),(69,1),(68,1),(68,1),(66,1),(66,3/4),(68, 1/4),(64,2)} la cual está extraída de los primeros ocho compases del tema principal de las variaciones K. 265 *Ah, vous dirai-je, Maman*; véase la figura 2. De esta representación eliminamos el ritmo, la segunda componente, y calculamos los intervalos consecutivos:

{0, 2,0,2,0,- 2,0,- 2,0,- 1,0,- 2,0,2,- 4}

Finalmente, calculamos la melodía borrosa (una vez que le aplicamos la clasificación por clases de la tabla 1):

{0, 1,0,1,0,- 1,0,- 1,0,- 1,0,- 2,0,2,- 2}

3.3 Transformaciones usando la transformada de Fourier

La transformada discreta de Fourier proporciona una buena descripción general de la forma del contorno melódico, especialmente si hay células melódicas repetidas. Tiene también la ventaja de la invariancia. Por ejemplo, la transformada discreta es invariante por cambios de escala. Sin embargo, dado que la transformada discreta de Fourier está pensada para señales periódicas (infinitas), hay ciertos efectos indeseables, el llamado efecto frontera, que hay que tener en cuenta. Véase [Sch99], páginas 303 a 306, para una discusión sobre la transformada discreta de Fourier y sus ventajas e inconvenientes.

4. Transformaciones rítmicas

4.1 Gaussificación

La gaussificación de un ritmo consiste en representarlo como la combinación de lineal de distribuciones normales. Cada distribución normal tiene su media sobre cada nota del ritmo; las desviaciones son fijas e iguales para todas las normales. Si t_n, n = 1,…,N, es la sucesión de tiempos de la melodía, la gaussificación g(t) es

$$g(t) = \frac{1}{N} \sum_{i=1}^{N} \frac{e^{-\frac{(t-t_i)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

En la figura 3 tenemos un conocido ritmo, la clave son 3/2.

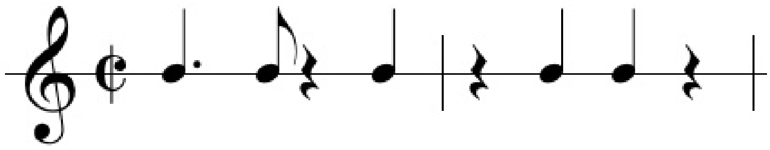

Figura 3: **El ritmo de la clave son.**

Si contamos el tiempo en semicorcheas de 0 a 15, las distribuciones normales en las notas ocurren en las posiciones 0, 3, 6, 10 y 12. Fijemos σ = 1 como desviación típica para todas esas distribuciones. Las medias de las normales serán las posiciones de las notas.

$$g_1(t) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}},\ g_2(t) = \frac{e^{-\frac{(t-3)^2}{2}}}{\sqrt{2\pi}},\ g_3(t) = \frac{e^{-\frac{(t-6)^2}{2}}}{\sqrt{2\pi}}, g_4(t) = \frac{e^{-\frac{(t-10)^2}{2}}}{\sqrt{2\pi}},\ g_5(t$$

La figura 4 muestra el gráfico de estas 5 distribuciones normales.

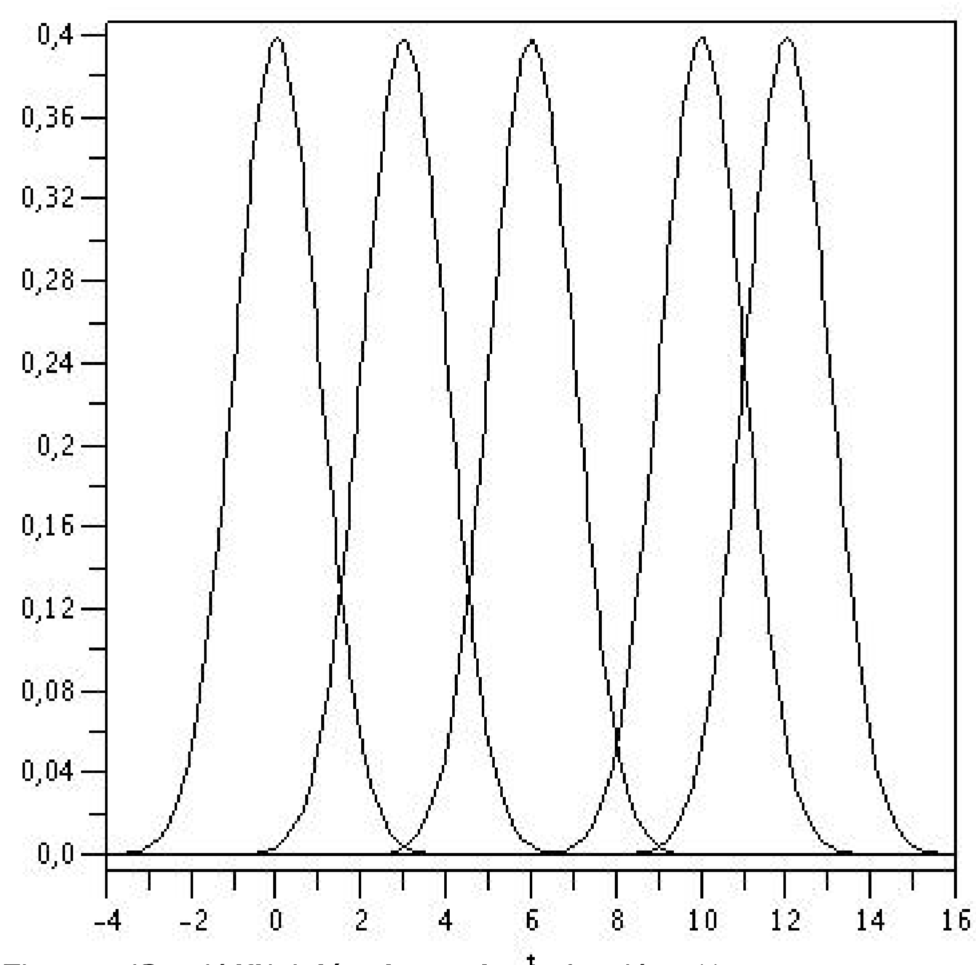

Figura 4: Gaussificación de un ritmo

MS90

$$\sigma(c_1, c_2) = 1 - \frac{d_e(c_1, c_2)}{\max(|c_1|, |c_2|)}$$

$$\sigma(C_1, C_2) = 1 - \frac{d_e(C_1, C_2)}{\max(|C_1|, |C_2|)} = 1 - \frac{3}{\max(4, 4)} = 1 - \frac{3}{4} = \frac{1}{4}$$

$$\sigma(a, b) = \frac{|a_n \cap b_n|}{\max(|a_n|, |b_n|)}$$