

# Interacción con la Música Flamenca: Análisis y Generación

Isabel Barbancho<sup>(1)</sup>, Nicolás Ruiz<sup>(2)</sup>, Pedro Vera<sup>(2)</sup>, Francisco Cañadas<sup>(2)</sup>, Carlos Ureña<sup>(3)</sup>, Lorenzo J. Tardón<sup>(1)</sup>, Ana M. Barbancho<sup>(1)</sup>, Carles Roig<sup>(1)</sup>

ibp@ic.uma.es, nicolas@ujaen.es, pvera@ujaen.es, fcanadas@ujaen.es, curena@ugr.es, lorenzo@ic.uma.es, abp@ic.uma.es

<sup>(1)</sup>Universidad de Málaga. Andalucía Tech. ETSI Telecomunicación. Dpto. Ingeniería de Comunicaciones, 29071, Málaga

<sup>(2)</sup>Universidad de Jaén. Escuela Politécnica Superior de Linares. Dpto. Ing. Telecomunicación. 23700 Linares, Jaén.

<sup>(3)</sup>Universidad de Granada. ETSI Informática y de Telecomunicación. Dpto. Lenguajes y Sistemas Informáticos, 18071, Granada.

**Resumen**—One main concern in Europe is the preservation and promotion of cultural heritage. Music is a main component of such heritage and flamenco is a main part of Andalusian musical culture. Focusing on flamenco music, in this paper, we describe steps on the analysis and generation of flamenco music oriented towards the diffusion and promotion of this musical style. Specifically, the separation of meaningful musical audio elements from flamenco excerpts by means of the scheme that will be described provides a novel and interactive way of listening to flamenco music by enabling the user to select or emphasize audio elements at will. Also, the scheme for the interactive generation of flamenco music that will be presented allows the user to play and interact with flamenco music by means of a user friendly interface implemented for iOS devices.

## I. INTRODUCCIÓN

Dentro del patrimonio histórico-artístico, la música es uno de los pilares fundamentales. En el caso del patrimonio andaluz, el flamenco es claro representante musical de la cultura andaluza, designado en 2010 por la Unesco como Patrimonio Cultural Inmaterial de la Humanidad [1].

Hoy en día, las Tecnologías de la Información y las Comunicaciones (TIC) ofrecen nuevas posibilidades de interactuar con el patrimonio cultural. En este ámbito, este artículo presenta un esquema de análisis de música flamenca cuyo objetivo es separar las componentes rítmicas (provenientes de la percusión), melódicas (provenientes de la voz cantada) y armónicas (provenientes de los instrumentos), para poder escucharlas separadamente, y un prototipo de sistema de generación interactiva de música flamenca implementado para dispositivos iOS que ofrece al público una nueva forma de acercarse al flamenco de manera fácil y amena.

Este artículo está estructurado de la siguiente forma: la Sección II expone la técnica empleada para llevar a cabo la separación de los diferentes elementos o sonidos que constituyen la música flamenca. La Sección III presenta las características necesarias para generar de forma automática melodías de un cierto estilo flamenca. En la Sección IV se describe el generador de melodías flamencas desarrollado. A continuación, la Sección V muestra el interfaz que tendrá la aplicación iOS de generación de melodías flamencas. Finalmente, en la Sección VI, se presenta una discusión de los resultados obtenidos y algunas conclusiones del trabajo desarrollado.

## II. SEPARACIÓN DE FUENTES: ELEMENTOS MELÓDICOS, RÍTMICOS Y ARMÓNICOS

La técnica Non-negative Matrix Factorization (NMF) [2] se utiliza para realizar una representación basada en patrones activos en datos multivariable. Sin embargo, NMF no es capaz, sin aplicar restricciones no negativas, de determinar si una base procedente de la descomposición pertenece a sonidos de tipo melódico (sonido vocal generado por el cantante), rítmico (sonido generado por un instrumento de tipo percusivo) o armónico (sonido generado por un instrumento armónico, por ejemplo, la guitarra flamenca).

Así, se propone un esquema de separación de estos tres tipos de sonidos dividido en 3 etapas: segmentación, entrenamiento (training) y separación. Esta separación se realizará resolviendo la limitación de NMF anteriormente expuesta sin utilizar ningún tipo de procesamiento basado en clustering o clasificación de las bases obtenidas. Según [3], el espectrograma  $X$  de una señal mezcla (melódico+rítmico+armónico) de audio se descompone en tres espectrogramas diferentes,  $X_p$  (percusivo),  $X_h$  (armónico) y  $X_v$  (vocal) mediante la integración en el proceso de descomposición de restricciones que modelan el comportamiento típico de esos sonidos. Algunas de dichas restricciones se basan en el concepto de suavidad anisotrópica [4], [5] relacionada con la diferencia en las direcciones de la continuidad del sonido. Concretamente: los sonidos armónicos se modelan mediante sparseness en frecuencia (picos espectrales) y suavidad (smoothness) en el tiempo, los sonidos percusivos se modelan utilizando suavidad en frecuencia (ruido de banda ancha) y sparseness en el tiempo (la energía se concentra en pequeños intervalos temporales) y, finalmente, los sonidos vocales se modelan utilizando sparseness en tiempo (fluctuaciones del pitch de la voz cantada) y sparseness en frecuencia (por la presencia de sonidos armónicos).

Inicialmente, en la etapa de segmentación, el espectrograma  $X$  de la señal se etiqueta manualmente en segmentos. Los segmentos pueden ser melódicos  $X_m$  (compuestos por sonidos melódicos, rítmicos y armónicos) y no-melódicos  $X_{nm}$  (compuestos exclusivamente por sonidos rítmicos y armónicos).

Seguidamente, las bases rítmicas  $W_p$  y armónicas  $W_h$  se aprenden mediante un algoritmo NMF no supervisado [3]

usando exclusivamente los segmentos no melódicos,

$$X_{nm} \approx X_p + X_h = W_p \cdot H_p + W_h \cdot H_h \quad (1)$$

Bajo un enfoque NMF semisupervisado, se extrae el espectrograma melódico  $X_m$  a partir de los segmentos melódicos. Ahora, las bases rítmicas y armónicas obtenidas en la etapa de entrenamiento se mantienen fijas. De esta forma, el enfoque semisupervisado asociará a las bases melódicas todas aquellas nuevas bases que aparezcan en la descomposición de los segmentos melódicos y que no hayan sido aprendidas en la etapa de entrenamiento al analizar los segmentos no melódicos en los que solo existían sonidos rítmicos y armónicos:

$$X_m \approx X'_p + X'_h + X_v = W_p \cdot H'_p + W_h \cdot H'_h + W_v \cdot H_v \quad (2)$$

A partir del espectrograma melódico  $X_v$ , la señal melódica  $v(t)$  en el dominio del tiempo es sintetizada. Sin embargo, para mejorar la señal melódica obtenida se utiliza la función diferencia para cada segmento de índice  $t$ :

$$d(\tau, t) = \sum_{n=0}^{W-1} (v(t+n) - v(t+n+\tau))^2 \quad (3)$$

donde  $W$  es la longitud de la ventana suma y  $\tau$  representa el periodo correspondiente al pitch. A continuación, se aplica la función diferencia normalizada media acumulativa [6],

$$d_n(\tau, t) = \begin{cases} 1, & \tau = 0 \\ d(\tau, t) / [\frac{1}{\tau} \sum_{j=1}^{\tau} d(j, t)] & \text{otherwise} \end{cases} \quad (4)$$

$d_n(\tau, t)$  actúa como una función de coste, donde cada elemento  $(\tau, t)$  indica el coste de tener un periodo de pitch igual a  $\tau$  en cada instante  $t$ . La trayectoria de pitch melódico se estima calculando con programación dinámica el camino de menor coste a través de la matrix  $d_n(\tau, t)$ . Para ello, los puntos iniciales del camino son fijos en el eje de tiempo y el camino se limita a un avance paso-a-paso en el eje  $t$ , asumiendo que  $|\tau_{t-1} - \tau_t| \leq 1$ . Esta condición asegura un contorno continuo y suave de la trayectoria del pitch melódico.

### III. MODELADO COMPUTACIONAL DE ELEMENTOS MUSICALES DE FLAMENCO

En esta sección, damos el primer paso para la generación de melodías flamencas. Se trata de identificar qué información de las canciones flamencas necesitará el compositor automático que vamos a implementar para dispositivos iOS. El esquema de generación se basará en el sistema presentado en [7], puesto que ha demostrado ser muy eficaz para la generación de melodías replicando un determinado estilo musical.

Para un determinado palo flamenco, el compás resulta ser una característica clave. Por tanto, seleccionando un cierto compás y siguiendo la estructura de [7], se tratará de obtener, para un determinado estilo, una base de datos que contendrá patrones rítmicos y melódicos (Fig. 1).

Por tanto, se almacenará cada patrón rítmico aparecido en las melodías analizadas y todos los contornos melódicos observados para cada patrón rítmico. Es posible definir la probabilidad de observación de un determinado patrón rítmico según:

$$f(r) = \sum_{i=1}^{R_p} \frac{N_{r_i}}{N_{RT}} \delta(r - r_i) \quad (5)$$

Palo flamenco

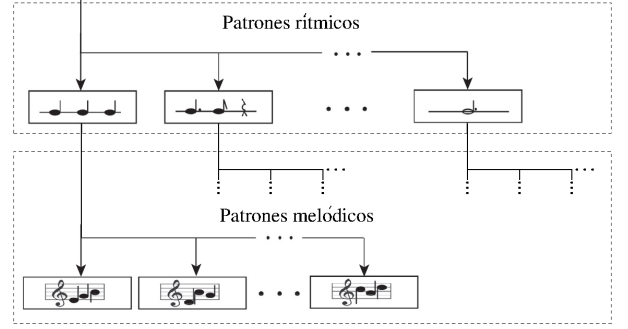


Fig. 1: Base de datos para generación musical con estilo flamenco.

donde  $p$  identifica el palo flamenco,  $R_p$  representa en número total de patrones rítmicos diferentes encontrados en las melodías de entrenamiento,  $N_{r_i}$  es el número de patrones melódicos encontrados para el patrón rítmico  $r_i$  y  $N_{RT} = \sum_{i=1}^{R_p} N_{r_i}$ .

Por otra parte, todos los patrones de contornos melódicos de cada patrón rítmico se consideran equiprobables entre sí. Así, la función de densidad de probabilidad condicional de dichos contornos melódicos será:

$$f(c|r) = \frac{1}{N_{CT}} \sum_{i=1}^{N_{CT}} \delta(c - c_i) \quad (6)$$

donde,  $N_{CT}$  representa el número de patrones melódicos diferentes encontrados para el patrón rítmico  $r$  en el palo flamenco que estemos considerando.  $c_i$  representa cada uno de los contornos melódicos válidos para el patrón rítmico  $r$ .

También, es necesario definir una progresión armónica y una estructura rítmica. Estas serán constantes, [I VII VI V I VII VI V] en modo menor y [A B C A A B A], respectivamente, para todas las composiciones.

### IV. GENERACIÓN AUTOMÁTICA DE FLAMENCO

El método automático de composición de música flamenca seleccionado [7] está basado en la reutilización de patrones rítmicos y melódicos. Existen otros trabajos anteriores en este campo, como los descritos en [8] o [9]. Sin embargo, la replicación de estilo musical no es un objetivo fundamental de estos esquemas de generación y, además, no suelen poner especial énfasis en la generación de música tonal. Estas dos características resultan distintivas en el esquema de generación musical seleccionado.

La implementación en iOS de este esquema de generación necesita los datos descritos en la Sección III que, en nuestra implementación se almacenan en una estructura XML. Los patrones rítmicos y figuras se normalizan y codifican en dicho fichero XML según se expone en [7]. Los contornos melódicos se almacenan, por comodidad, usando los números MIDI correspondientes a las notas detectadas, si bien es únicamente el contorno melódico el que debe preservarse para mantener el estilo musical deseado [10].

La Fig. 2 muestra parcialmente un fichero XML con la información que utiliza el generador de melodías. Este fichero indica el compás, ligado a estilos flamencos concretos, el número de patrones rítmicos diferentes (122) y los patrones rítmicos y melódicos. En el ejemplo, se muestra el patrón rítmico <Ritmo>0.5 0.25 0.5 0.75 0.25

```

<?xml version="1.0" encoding="utf-8"?>
<Database>
  <Compas>
    <Value>4 4</Value>
    <NumPat>122</NumPat>
    <Patron>
      <id>1</id>
      <Ritmo>0.5 0.25 0.5 0.75 0.25 0.25 0.25 0.75 0.25 0.25</Ritmo>
      <NumProg>1</NumProg>
      <Progresiones>
        <Progresion>
          <Melodia>72 71 74 72 71 69 1 71 69 67</Melodia>
        </Progresion>
      </Progresiones>
    </Patron>
  </Database>

```

Fig. 2: Detalle del fichero XML que usa el generador de melodías.

0.25 0.25 0.75 0.25 0.25</Ritmo>, al que corresponde un contorno melódico: <Melodia>72 71 74 72 71 69 1 71 69 67</Melodia>.

Con esto, el generador ya puede seleccionar patrones rítmicos para la composición. En concreto se seleccionan tres patrones rítmicos diferentes, de acuerdo con la estructura rítmica predefinida, de forma aleatoria según la función de densidad de probabilidad definida en la ec. (5). A continuación, se selecciona el contorno melódico para cada compás de forma también aleatoria, de acuerdo con la ec. (6).

Finalmente, el sistema debe adaptar las notas (contorno melódico) a los acordes de la progresión armónica. Para ello, en primer lugar definimos como notas de paso a todas aquellas que están en tiempos débiles y se aplican las reglas definidas por Schellenberg en [11] de forma que:

- La primera nota de la melodía sea la tónica.
- La primera nota de cada compás será la nota del acorde más cercana a la nota candidata según las reglas de Schellenberg.
- Las notas de paso se asignan siguiendo los cambios de pitch que indica el contorno melódico.
- En cada tiempo fuerte se asigna la nota del acorde más cercana siguiendo la dirección de cambio del contorno melódico.

Para lograr más diversidad en las melodías generadas y acercar el estilo flamenco a usuarios aficionados a otros estilos musicales, se ha implementado un mecanismo de combinación de patrones armónicos y rítmicos. Así, el conjunto de patrones rítmicos, y sus correspondientes patrones armónicos, que usará el generador de melodías contendrá una mezcla de patrones rítmicos de diferentes estilos flamencos. Adicionalmente se incorpora el pop, por ser un musical muy aceptado, para acercar las composiciones flamencas a dicho estilo, si se desea. El porcentaje de patrones rítmicos de cada estilo será seleccionable por el usuario por medio del interfaz que se describe a continuación.

## V. DESARROLLO DEL INTERFAZ GRÁFICO

Se ha diseñado un interfaz para la aplicación de generación de flamenco teniendo como objetivos principales la usabilidad y portabilidad a distintos dispositivos iOS con distintos factores de forma. Adicionalmente, se ha incorporado una animación cuyo aspecto se hace depender de los parámetros que se usan para la generación de audio y que pretende reflejar visualmente sus características. La aplicación incluye una pantalla de introducción, con el logotipo del proyecto CEI-patrimonio, una pantalla de selección de mezcla de los palos de flamenco, y otra de generación y reproducción (Fig. 3).



Fig. 3: Pantalla de selección de mezclas (*mixtures*) (izq.), y pantalla de generación y reproducción (der.).

La pantalla de selección de porcentajes de patrones rítmicos de cada palo en la mezcla permite variar dichos valores de forma intuitiva mediante varios deslizadores. Asimismo, el usuario tiene retroalimentación inmediata de la importancia relativa de cada estilo en un diagrama de sectores.

En la pantalla de generación y reproducción se generará la melodía permitiendo, durante la reproducción, cambiar el tempo de la misma, con efecto tanto en la señal de audio como en la visualización.

La animación que se visualiza durante la reproducción incluye una figura con forma fractal, cuyos parámetros varían periódicamente con el tiempo produciendo una sensación de movimiento rítmico que acompaña visualmente a la percepción de ritmo en la propia señal de audio (ver Fig. 3 der.).

La figura consiste en cuatro árboles fractales (autosemejantes), iguales, rotados 90 grados alrededor de la raíz (común, en el centro del rectángulo de la animación). El modelo del árbol fractal consiste en un tronco (un segmento *padre*) que sale de la raíz y que en el extremo opuesto se ramifica de forma recursiva en dos segmentos (*hijos*) iguales pero algo más cortos. Se puede describir como un sistema de Lindemayer (un sistema-L) [12] sencillo con cuatro parámetros: los dos ángulos que forman los dos segmentos hijos con respecto al segmento padre ( $\alpha_0$  y  $\alpha_1$ ) y las dos fracciones de longitud de los hijos respecto del padre ( $f_0$  y  $f_1$ ).

El sistema-L completo puede expresarse con una gramática libre de contexto parametrizada sencilla, usando la notación popularizada por Prusinkiewicz [13]. El símbolo inicial es  $S$  y admite dos parámetros:  $l$  (longitud del tronco raíz de cada árbol) y  $\theta$  (ángulo de dicho tronco), este símbolo se traduce en cuatro instancias de  $T$ , de esta forma:

$$\begin{aligned}
 S(l, \theta) \rightarrow & +(\theta)T(l) + (\pi/2)T(l) \\
 & +(\pi/2)T(l) + (\pi/2)T(l)
 \end{aligned} \quad (7)$$

el valor de  $\theta$  cambia linealmente con el tiempo  $t$ . Un parámetro  $k$ , que multiplica a  $t$ , se hace depender linealmente del tempo (en *bpm*), de forma que la figura completa gira a ritmo proporcional al tempo.

Cada uno de los cuatro árboles  $T$  consta de un tronco

(segmento generado por el símbolo de avance,  $F$ ), y dos copias reducidas (por los ratios  $f_0$  y  $f_1$ ) y rotadas ( $\alpha_0$  y  $\alpha_1$ ) del mismo símbolo  $T$ , que por tanto se puede describir como:

$$T(s) \rightarrow [F(s)[+(\alpha_0)T(sf_0)]+(\alpha_1)T(sf_1)] \quad (8)$$

Los cuatro parámetros del sistema-L ( $\alpha_0, \alpha_1, f_0$  y  $f_1$ ) varían de forma periódica con el tiempo  $t$  (seg.) según:

$$\alpha_0 = a_0 + (b_0 - a_0) [1 + \sin(k2\pi t/\tau_0)]/2 \quad (9)$$

$$\alpha_1 = a_1 + (b_1 - a_1) [1 + \sin(k2\pi t/\tau_1)]/2 \quad (10)$$

$$f_0 = a_2 + (b_2 - a_2) [1 + \sin(k2\pi t/\tau_2)]/2 \quad (11)$$

$$f_1 = a_3 + (b_3 - a_3) [1 + \sin(k2\pi t/\tau_3)]/2 \quad (12)$$

Cada período  $\tau_i$  se hace depender linealmente del peso de cada palo del flamenco en la composición (parámetros  $w_0, w_1, w_2, w_3$ ):

$$\tau_i = c_i - d_i w_i \quad (13)$$

donde  $0 < d_i < c_i$ . Por tanto  $\tau_i$  nunca es cero o negativo, ya que los pesos están todos ellos entre 0 y 1. De esta forma, cada combinación de mezclas de palos produce una animación distinta.

## VI. RESULTADOS Y CONCLUSIONES

En este artículo se han presentado avances significativos en la interacción con la música flamenca.

En relación con el análisis, se ha abordado la descomposición de flamenco en componentes de audio con sentido musical. Algunos ejemplos son accesibles al público<sup>1</sup>. De cada fichero de audio se proporciona la mezcla original y cada componente de tipo melódico, rítmico o armónico separado. Los resultados muestran que el elemento melódico (voz cantada) presenta una muy buena calidad. El algoritmo es capaz de extraer la voz cantada evitando la captura de interferencias rítmicas o armónicas. Este resultado es especialmente relevante puesto que el análisis de únicamente este elemento melódico simple puede permitir la extracción de patrones utilizables por el generador de música flamenca.

En cuanto al elemento rítmico, éste se compone principalmente de la parte percusiva de la guitarra, sin embargo, también se captura parte de la voz. La señal correspondiente a la parte armónica está formada fundamentalmente por la guitarra cuando ésta se comporta como un instrumento puramente armónico pero se observa interferencia de la voz cantada debido a que ésta muestra un comportamiento más similar a los sonidos armónicos.

En relación con la generación de melodías de música flamenca, se ha implementado un esquema de generación de para sistemas iOS de forma que permite la creación de melodías de estilo flamenco con compás 4/4 y la mezcla con el estilo musical pop. La aplicación desarrollada posee un interfaz intuitivo y cómodo de usar, ya que permite seleccionar de una forma ágil y visual los valores de los parámetros que determinan las composiciones. La animación que se presenta durante la reproducción traslada al mundo visual el ritmo y la variabilidad de las composiciones, acompañando de forma agradable y estéticamente coherente a la pieza generada.

En la web del grupo de investigación ATIC<sup>2</sup> es posible escuchar el resultado de algunas melodías generadas con el compositor descrito. Se puede comprobar que su estilo es, en efecto, flamenco, muy similar al de las canciones usadas para el entrenamiento (como: Viva el pasodoble o María la portuguesa). También se puede comprobar la influencia de la mezcla de patrones de música pop en la generación y como el sistema de adaptación de contorno melódico funciona correctamente, por lo que consideramos que la generación de otros palos flamencos y la mezcla de los mismos entre sí y/o con patrones de estilo pop será también correctamente realizada. Aún es preciso analizar música correspondiente a los palos flamencos seleccionados para crear las bases de datos XML que necesita el compositor.

## AGRADECIMIENTOS

Ese trabajo ha sido parcialmente financiado por “Campus de Excelencia Internacional en Patrimonio, PATRIMONIUN-10”, para programas de generación de investigación de referencia internacional en materia de Patrimonio Cultural y Natural, convocatoria realizada en el ámbito del proyecto PATRIMONIUN-10, Jaén, Marzo de 2014, y por el Ministerio de Economía y Competitividad del Gobierno de España, Proyecto TIN2013-47276-C6-2-R. Parte de este trabajo ha sido realizado en la Universidad de Málaga. Campus de Excelencia Internacional Andalucía Tech.

## REFERENCIAS

- [1] Unesco. Decisión 5.COM 6.39, 2010.
- [2] D. Lee and H. Seung, “Algorithms for Non-negative Matrix Factorization”, in *Advances in NIPS*, pp. 556-562, 2000
- [3] F. Cañadas and P. Vera and N. Ruiz and J. Carabias and P. Cabanas, “Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints, ”, in *Journal on Audio, Speech, and Music Processing*, vol. 2014, no. 26, pp. 1-26, 2014
- [4] T. Virtanen, “Monaural Sound Source Separation by Non-Negative Matrix Factorization with Temporal Continuity and Sparseness Criteria” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 3, no. 15, March 2007.
- [5] N. Ono and K. Miyamoto and J. Le Roux and H. Kameoka and S. Sagayama, “Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram, ”, in *Proceedings of EUSIPCO*, 2008
- [6] A. de Cheveigné and H. Kawahara, “YIN, a Fundamental Frequency Estimator for Speech and Music”, *Journal of the Acoustic Society of America*, vol. 111, no. 4, pp. 1971-1930, 2002.
- [7] C. Roig, L.J. Tardon, I. Barbancho, A. M. Barbancho, “Probabilistic Automatic Melody Composition Based on a Probabilistic Model of Music Style and Harmonic Rules”, *Elsevier Knowledge Based Systems*, vol. 71, pp. 419-434, November 2014.
- [8] B. van der Merwe and W. Schulze, “Music Generation with Markov Models”, *IEEE Multimedia* vol.18 , no. 3, pp. 78-85, Aug. 2011.
- [9] D. Cope, “Computer Models of Musical Creativity”, MIT Press Cambridge, 2005.
- [10] W.J. Dowling and D.S. Fujitani, “Contour, Interval and Pitch Recognition in Memory”, *Journal of Acoustic Society of America*, vol.49 , no. 2B, pp. 524-531, 1971.
- [11] E. Schellenberg, “Simplifying the Implication-Realization Model of Musical Expectancy”, *Music Perception*, vol.14 , no. 3, pp. 295-318, 1997.
- [12] A. Lindenmayer, “Mathematical models for cellular interaction in development, Parts I and II”, *Journal of Theoretical Biology*, vol. 18, pp. 280-315, 1968.
- [13] P. Prusinkiewicz, A. Lindenmayer, J. Hanan, *The algorithmic beauty of plants*, Ed. Springer-Verlag, 1990.

<sup>1</sup><https://dl.dropboxusercontent.com/u/22448214/WebUJAURSI2015/indexURSI2015.html>

<sup>2</sup>[www.atic.uma.es](http://www.atic.uma.es). Links.