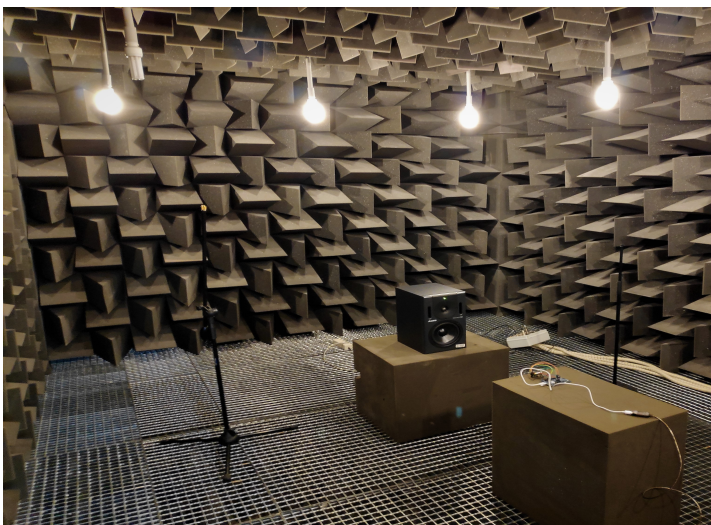
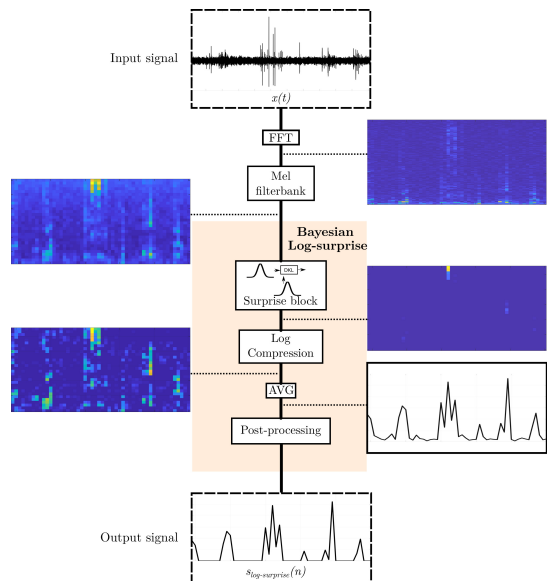


# SALIENCIA Y ATENCIÓN



## MULTIMODALIDAD, *CONTEXT-AWARENESS*, AUTO-ADAPTACIÓN Y BIO-INSPIRACIÓN (SAMURAI)

### Financiado por:

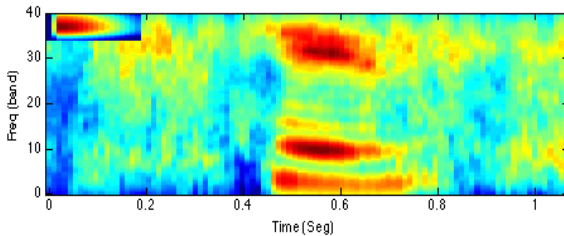
MINECO (Ministerio de Economía y Competitividad). TEC2014-53390-P (Convocatoria 2014 de Proyectos de I+D del Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia). Enero 2015 – Diciembre 2018.



# PARTICIPANTES

Ascensión Gallardo Antolín  
Carmen Peláez Moreno  
Fernando Díaz de María  
Iván González Díaz  
Francisco José Valverde Albacete  
Fernando de la Calle Silos  
Miguel Ángel Fernández Torres  
José Luis González de Suso Molinero  
Amaya Jiménez Moreno  
Javier López Labraca  
Jimmy Diestin Ludeña Choez  
Tomás Martínez Cortés  
Antonio Rodríguez Hidalgo  
Miguel Fernández Díaz  
Adrián Vázquez Romero  
Esther González Rituerto

# SAMURAI



## Entendiendo el mecanismo de la atención

El proyecto **SAMURAI** ha tenido como objetivo principal entender un mecanismo cognitivo básico para la supervivencia del ser humano: la **atención**. Se trata de un mecanismo por el cual priorizamos ciertos estímulos frente a otros: en su proceso de exploración del mundo, nuestro cerebro toma decisiones sobre qué es relevante y qué no en cada situación. La metodología que hemos utilizado para comprender este mecanismo es la de construir modelos computacionales del mismo siguiendo la filosofía de Feynman de que "lo que no puedo crear, no lo entiendo". Comprendiendo que este fenómeno es genuinamente **multimodal** hemos trabajado principalmente en dos modalidades: la **visual** y la **aural**.

### Atención visual

En el caso de la atención visual, el estado del arte se encuentra mucho más avanzado principalmente por la disponibilidad de sensores (*eye-trackers*) capaces de registrar hacia dónde dirigimos nuestra mirada (fijaciones oculares), esto es, a qué regiones prestamos más atención. Así, en este caso, habida cuenta de que con anterioridad se había estudiado ampliamente el mecanismo de saliencia (*bottom-up*), guiada por estímulos, el proyecto se ha centrado en el más general de atención (*top-down*), dirigido por la tarea u objetivo del sujeto, proponiendo un sistema jerárquico basado en el empleo de modelos de tópicos latentes que realiza un mapeo entre estímulos y características de bajo nivel y la atención dirigida a la realización de tareas (alto nivel). En particular, este mapeo se hace a través de una capa intermedia (los tópicos latentes) que representa sub-tareas de especial interés para modelar la atención en escenarios específicos, en los que un sujeto desea resolver una tarea particular.

### Atención aural

En el segundo caso (atención aural), la dificultad de obtener medidas empíricas nos ha hecho decantarnos por métodos no supervisados. Así, hemos desarrollado un sistema de atención aural no supervisado basado en métodos bayesianos, el concepto biológico de memoria ecoica o memoria sensorial auditiva y la fusión de información a diversas escalas temporales mediante la utilización de diferentes distancias o divergencias estadísticas. Hemos evaluado el funcionamiento de este sistema sobre tareas de detección de eventos acústicos y analizado su robustez frente a diversas condiciones de ruido ambiental. Por último, y aunque esta línea queda aún abierta, hemos desarrollado un sistema de detección de saliencia visual basado en la influencia que la saliencia auditiva ejerce en la percepción.



## El impacto

El proyecto ha producido un impacto destacable en varios aspectos. Por una parte, el impacto científico-tecnológico del avance del conocimiento que ha supuesto SAMURAI se ve reflejando en un número considerable de publicaciones científicas y en el desarrollo de una serie de herramientas *software* y demostradores con el objeto de dar visibilidad a los resultados más importantes del proyecto y que esperamos que también sean de utilidad a otros investigadores en el ámbito<sup>1</sup>:

- Detección precoz de melanomas: DermaKNet
- Detección de matrículas en entornos no restringidos
- Utilillaje de funciones software para atención visual

Además, se ha generado una base de datos de vídeos etiquetados con fijaciones con el objeto de facilitar la investigación en la detección de anomalías y otra base de datos de voz de pacientes de Parkinson.

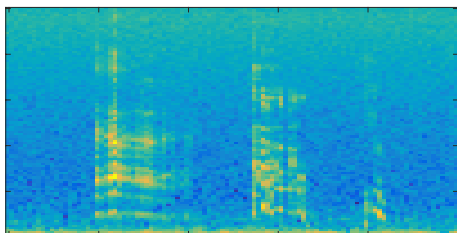
Por otra parte, el impacto socio-económico de SAMURAI puede ser muy elevado ya que hemos identificado aplicaciones en las áreas de salud, seguridad, transporte y turismo en las que ya estamos trabajando en transferir los conocimientos.

También nos hemos encontrado con impactos no previstos inicialmente. Por ejemplo:

- Se ha iniciado una colaboración científica con el CNIC para estudiar la viabilidad del empleo de técnicas de visión artificial en la caracterización de neutrófilos.
- Se ha iniciado una colaboración científica con investigadores de la Universidad de Málaga con los que se están encontrando nuevas formas de aplicar el Análisis de Conceptos Formales para reflejar la relación de independencia y de equivalencia y que podrían tener un impacto importante en técnicas de aprendizaje máquina multi-etiqueta, por ejemplo.
- Se ha iniciado una colaboración interdisciplinar con varios departamentos de la Universidad Carlos III de Madrid y con la Universidad Politécnica de Madrid así como con el Instituto de Estudios de Género de la UC3M para desarrollar un dispositivo *wearable* para alertar de situaciones de violencia de género en el que la determinación de eventos salientes que llaman la atención es un elemento importante.

---

<sup>1</sup> Estos recursos están disponibles en <http://gpm.webs.tsc.uc3m.es/resources/>



## Los sectores beneficiados

En el ámbito de la **salud**, nuestras técnicas tienen aplicación en:

- La detección precoz de melanomas. De hecho, nuestro sistema obtuvo la segunda posición global en el *Challenge* relacionado con esta temática ISIC 2017.
- La ayuda al diagnóstico basado ecografía renal.
- El estudio de los neutrófilos a partir de imágenes de microscopía intra-vital.
- La predicción del grado de severidad de Parkinson.

En el sector de **transporte**, se ha trabajado en la aplicación de las técnicas desarrolladas en el proyecto a las siguientes tareas:

- Detección de matrículas en entornos no controlados.
- Detección automática de incidentes de tráfico.
- Detección de elementos de señalización relevantes para la seguridad vial, a través de un contrato de I+D.
- Seguridad de infraestructuras de transporte.

En el sector de **seguridad**, estamos trabajando para trasladar nuestras técnicas a los siguientes campos:

- La videovigilancia, mediante el uso de técnicas de detección de saliencia temporal para la selección de la cámara de videovigilancia a observar en tiempo real o a la búsqueda de anomalías en vídeo pre-grabado.
- La protección de víctimas de violencia de género, mediante el desarrollo de un dispositivo para alertar de este tipo de situaciones.

En el sector de **turismo**, se está aplicando la tecnología desarrollada en este proyecto a:

- El desarrollo de un escaparate turístico interactivo cuyo objetivo principal es convertir los escaparates turísticos convencionales en puntos informativos interactivos accesibles desde cualquier teléfono móvil y enriquecidos con contenidos extraídos automáticamente de redes sociales.

# Para saber más



## Publicaciones Científicas<sup>2</sup>

- K. A. Abdmalak and A. Gallardo-Antolín, "Enhancement of a text-independent speaker verification system by using feature combination and parallel structure classifiers," *Neural Computing and Applications*, vol. 29, iss. 3, pp. 637-651, 2018.
- J. Ludeña-Choez, R. Quispe-Soncco, and A. Gallardo-Antolín, "Bird sound spectrogram decomposition through Non-Negative Matrix Factorization for the acoustic classification of bird species," *PLOS ONE*, vol. 12, iss. 6, pp. 1-20, 2017.
- F. de-la-Calle-Silos and R. M. Stern, "Synchrony-Based Feature Extraction for Robust Automatic Speech Recognition," *IEEE Signal Processing Letters*, vol. 24, iss. 8, pp. 1158-1162, 2017.
- M. Molina-Moreno, I. González-Díaz, and F. Díaz-de-María, "Efficient Scale-Adaptive License Plate Detection System," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1-13, 2018.
- I. González-Díaz, "DermaKNet: Incorporating the knowledge of dermatologists to Convolutional Neural Networks for skin lesion diagnosis," *IEEE Journal of Biomedical and Health Informatics*, pp. 1-1, 2018.
- J. López-Labraca, M. Á. Fernández-Torres, I. González-Díaz, F. Díaz-de-María, and Á. Pizarro, "Enriched dermoscopic-structure-based cad system for melanoma diagnosis," *Multimedia Tools and Applications*, vol. 77, iss. 10, pp. 12171-12202, 2018.
- E. Martínez-Enríquez, J. Cid-Sueiro, F. Díaz-de-María, and A. Ortega, "Optimized Update/Prediction Assignment for Lifting Transforms on Graphs," *IEEE Transactions on Signal Processing*, vol. 66, iss. 8, pp. 2098-2111, 2018.
- E. Martínez-Enríquez, J. Cid-Sueiro, F. Díaz-de-María, and A. Ortega, "Directional Transforms for Video Coding Based on Lifting on Graphs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, iss. 4, pp. 933-946, 2018.
- J. L. González-de-Suso, E. Martínez-Enríquez, and F. Díaz-de-María, "Adaptive Lagrange multiplier estimation algorithm in HEVC," *Signal Processing: Image Communication*, vol. 56, pp. 40-51, 2017.
- A. Jiménez-Moreno, E. Martínez-Enríquez, and F. Díaz-de-María, "Bayesian adaptive algorithm for fast coding unit decision in the High Efficiency Video Coding (HEVC) standard," *Signal Processing: Image Communication*, vol. 56, pp. 1-11, 2017.
- I. González-Díaz, V. Buso, and J. Benois-Pineau, "Perceptual modeling in the problem of active object recognition in visual scenes," *Pattern Recognition*, vol. 56, iss. Supplement C, pp. 129-141, 2016.
- A. Hernández-García, F. Fernández-Martínez, and F. Díaz-de-María, "Comparing visual descriptors and automatic rating strategies for video aesthetics prediction," *Signal Processing: Image Communication*, vol. 47, pp. 280-288, 2016.
- F. J. Valverde-Albacete and C. Peláez-Moreno, "Assessing Information Transmission in Data Transformations with the Channel Multivariate Entropy Triangle," *Entropy*, vol. 20, iss. 7, 2018.
- A. Rodríguez-Hidalgo, C. Peláez-Moreno, and A. Gallardo-Antolín, "Echoic log-surprise: A multi-scale scheme for acoustic saliency detection," *Expert Systems with Applications*, vol. 114, pp. 255-266, 2018.
- M. Á. Fernández-Torres, I. González-Díaz, and F. Díaz-de-María, "A probabilistic topic approach for context-aware visual attention modeling," in *2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, 2016, pp. 1-6.
- J. Ludeña-Choez and A. Gallardo-Antolín, "Acoustic Event Classification using spectral band selection and Non-Negative Matrix Factorization-based features," *Expert Systems with Applications*, vol. 46, pp. 77-86, 2016.
- A. Jiménez-Moreno, E. Martínez-Enríquez, and F. Díaz-de-María, "Complexity Control Based on a Fast Coding Unit Decision Method in the HEVC Video Coding Standard," *IEEE Transactions on Multimedia*, vol. 18, iss. 4, pp. 563-575, 2016.
- G. I. Díaz, "DermaKNet: Incorporating the knowledge of dermatologists to Convolutional Neural Networks for skin lesion diagnosis," *IEEE Journal of Biomedical and Health Informatics*, iss. 99, pp. 1-1, 2018.
- M. de-Frutos-Lopez, J. Luis González-de-Suso, S. Sanz-Rodríguez, C. Peláez-Moreno, and F. Díaz-de-María, "Two-level sliding-window VBR control algorithm for video on demand streaming," *Signal Processing-Image Communication*, vol. 36, pp. 1-13, 2015.
- F. J. Valverde-Albacete, J.M.González-Calabozo, A. Peñas, and C. Peláez-Moreno, "Supporting scientific knowledge discovery with extended, generalized Formal Concept Analysis," *Expert Systems with Applications*, vol. 44, pp. 198-216, 2016.

<sup>2</sup> En acceso abierto en <https://e-archivo.uc3m.es/handle/10016/1591>



## Congresos

E. Rituerto-González, A. Gallardo-Antolín, and C. Peláez-Moreno, "Speaker Recognition under Stress Conditions," in Proceedings of the X Jornadas en Tecnología del Habla and V Iberian SLTech Workshop (IberSPEECH 2018), 2018, pp. 15-19.

C. Jiménez-Recio, A. Zlotnik, A. Gallardo-Antolín, J. M. Montero, and J. C. Martínez-Castrillo, "Prediction of the Degree of Parkinson's Condition Using Recordings of Patients' Voices," in Proceedings of the Ninth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2017), Cham, 2018, pp. 120-129.

A. Rodríguez-Hidalgo, A. Gallardo-Antolín, and C. Peláez-Moreno, "Towards aural saliency detection with logarithmic Bayesian Surprise under different spectro-temporal representations," in Proceedings of the IX Jornadas en Tecnología del Habla and V Iberian SLTech Workshop (IberSPEECH 2016), 2016, pp. 99-108.

J. Ludeña-Choez and A. Gallardo-Antolín, "Non-negative Matrix Factorization Applications to Speech Technologies," in Proceedings of the IX Jornadas en Tecnología del Habla and V Iberian SLTech Workshop (IberSPEECH 2016), 2016, pp. 339-348.

T. Martínez-Cortés, I. González-Díaz, and F. Díaz-de-María, "Automatic Learning of Image Representations Combining Content and Metadata," in 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 1972-1976.

M. Fernández-Torres, I. González-Díaz, and F. Díaz-de-María, "A probabilistic topic approach for context-aware visual attention modeling," in 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI), 2016, pp. 1-6.

F. J. Valverde-Albacete, C. Peláez-Moreno, I. P. Cabrera, P. Cordero, and M. Ojeda-Aciego, "A Data Analysis Application of Formal Independence Analysis," in Concept Lattices and their Applications (CLA 2018), , 2018, pp. 1-12.

A. Rodríguez-Hidalgo, C. Peláez-Moreno, and A. Gallardo-Antolín, "Towards multimodal saliency detection: An enhancement of audio-visual correlation estimation," in 2017 IEEE 16th International Conference on Cognitive Informatics Cognitive Computing (ICCI\*CC), 2017, pp. 438-443.

F. J. Valverde-Albacete and C. Peláez-Moreno, "On the Relation between Semifield-Valued FCA and the Idempotent Singular Value Decomposition," in 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2018, pp. 1-8.

F. J. Valverde-Albacete and C. Peláez-Moreno, "Entropic Evaluation of Classification. A hands-on, get-dirty introduction," in International Joint Conference on Neural Networks (IJCNN'18 at WCCI 2018), 2018.

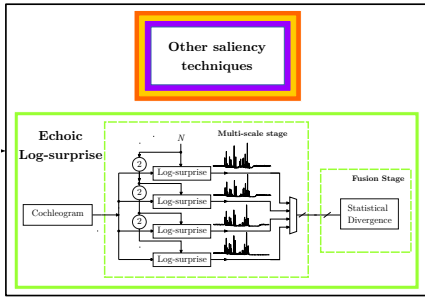
A. Zlotnik, J. M. M. Martínez, R. S. S. Hernández, and A. Gallardo-Antolín, "Random forest-based prediction of Parkinson's disease progression using acoustic, ASR and intelligibility features," in 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), 2015, pp. 503-507.

F. de-la-Calle-Silos, F. J. Valverde-Albacete, A. Gallardo-Antolín, and C. Peláez-Moreno, Preliminary experiments on the robustness of biologically motivated features for DNN-based ASR, , 2015.

V. Buso, I. González-Díaz, and J. Benois-Pineau, "OBJECT RECOGNITION WITH TOP-DOWN VISUAL ATTENTION MODELING FOR BEHAVIORAL STUDIES," in 2015 IEEE International Conference on Image Processing, , 2015, pp. 4431-4435.

F. de-la-Calle-Silos, A. Gallardo-Antolín, and C. Peláez-Moreno, "An Analysis of Deep Neural Networks in Broad Phonetic Classes for Noisy Speech Recognition," in Advances in Speech and Language Technologies for Iberian Languages, Iberspeech 2016, A. Abad, A. Ortega, A. Teixeira, C. G. Mateo, C. D. M. Hinarejos, F. Perdigao, F. Batista, and N. Mamede, Eds., , 2016, vol. 10077, pp. 87-96.

F. J. Valverde-Albacete and C. Peláez-Moreno, "The Linear Algebra in Formal Concept Analysis over Idempotent Semifields," in Formal Concept Analysis, J. Baixeries, C. Sacarea, and M. Ojeda-Aciego, Eds., 2015, vol. 9113, pp. 97-113.



## Tesis Doctorales

- J. D. Ludeña-Choez, "Contribuciones a la Aplicación de la Factorización de Matrices No Negativas a las Tecnologías del Habla," PhD Thesis, 2015.
- J. M. González-Calabozo, "Análisis exploratorio de datos de expresión genómica mediante el análisis en conceptos formales.," PhD Thesis, 2016.
- J. L. González-de-Suso-Moliner, "Contributions to the Solution of the Rate-Distortion Optimization Problem in Video Coding," PhD Thesis, 2016.
- A. Jiménez-Moreno, "Algorithms for complexity management in video coding," PhD Thesis, 2016.
- F. de la Calle-Silos, "Bio-Motivated Features and Deep Learning for Robust Speech Recognition," PhD Thesis, 2017.
- M. Á. Fernández-Torres, "Hierarchical representations for spatio-temporal visual attention modeling and understanding," PhD Thesis, 2019.