

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 573 952**

21 Número de solicitud: 201500005

51 Int. Cl.:

G06K 9/00 (2006.01)

G06T 7/20 (2006.01)

G10L 15/25 (2013.01)

12

PATENTE DE INVENCION

B1

22 Fecha de presentación:

12.12.2014

43 Fecha de publicación de la solicitud:

13.06.2016

88 Fecha de publicación diferida del informe sobre el estado de la técnica:

30.12.2016

Fecha de concesión:

17.10.2017

45 Fecha de publicación de la concesión:

24.10.2017

73 Titular/es:

UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA (100.0%)

Juan de Quesada, nº 30

35001 Las Palmas de Gran Canaria (Las Palmas) ES

72 Inventor/es:

TRAVIESO GONZÁLEZ, Carlos Manuel;
QUINTANA SANTANA, Carmelo Ruymán y
ALONSO HERNÁNDEZ, Jesús Bernardino

54 Título: **Método para la cuantificación de la carga emocional mediante el movimiento de los labios**

57 Resumen:

La presente invención se refiere a un método para la determinación de la cuantificación de la carga emocional (o Quantification of Arousal, en inglés) a través del movimiento de los labios, a partir de imágenes faciales, las cuales se obtienen de la grabación en vídeo de un diálogo del sujeto a evaluar.



Figura 2

ES 2 573 952 B1

Aviso: Se puede realizar consulta prevista por el art. 37.3.8 LP 11/1986.

DESCRIPCIÓN

Método para la cuantificación de la carga emocional mediante el movimiento de los labios.

5

La presente invención se refiere a un método para la determinación de la cuantificación de la carga emocional (o Quantification of Arousal, en inglés) a través del movimiento de los labios, a partir de imágenes faciales, las cuales se obtienen de la grabación en vídeo de un diálogo del sujeto a evaluar.

10

Antecedentes de la invención

15

Actualmente, existen numerosos estudios y desarrollos basados en la relación entre el hombre en su plano físico y sus emociones. En ellos se intentan analizar el por qué y el cómo se producen estas emociones, intentando relacionar los fenómenos o reacciones, físicos e internos del cuerpo humano, con el fin de responder a estas cuestiones, y ser capaz de distinguir dichas emociones.

20

Estos estudios no solo ayudan a comprender y diferenciar las emociones, sino que sirven de ayuda para la detección de patologías neurodegenerativas o diferentes enfermedades mentales, como el síndrome de Asperger, el autismo, etc., donde sus principales síntomas se relacionan directamente con el reconocimiento, o la propia falta de interpretación de las emociones. También ayudan a controlar situaciones de estrés, en trabajos como operaciones quirúrgicas, o cuya necesidad de concentración sea alta, un largo etcétera de posibilidades. Por ello, su análisis y detección es de suma importancia en los tiempos actuales y futuros.

25

30

Es clave distinguir entre el tipo o clase de emoción y la carga emocional que lleva consigo la emoción; ya que ambas son dos definiciones totalmente diferentes. La emoción ha sido trabajada en los últimos años, pero la carga emocional o su intensidad, es algo que se está empezando a investigar. En particular, para información extraída de la zona facial, la emoción ha sido ampliamente estudiada, pero la carga emocional o la intensidad con la que se manifiesta a partir del movimiento labial, no ha sido estudiada, y esta propuesta muestra un método para su cuantificación.

35

40

En el estudio de emociones, varias son las investigaciones, que relacionan ciertas características biométricas, con dichas emociones, con el fin de intensificar y diferenciar el estado emocional del individuo. Por tanto las modalidades usadas son varias para esta finalidad. Así, por ejemplo, el análisis de imágenes faciales, han sido utilizadas como base para detectar emociones que se manifiestan físicamente. No hay que confundir la emoción con el grado de la emoción. Ejemplos de la detección de la emoción se pueden encontrar en los siguientes artículos y patentes:

45

i) P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. Proc. of CVPR, páginas 511-518, 2001

50

ii) T. Ojala and M. Pietikäinen. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol. 24. No.7, Julio, 2002.

- iii) Timo Ahonen, Abdenour Hadid and Matti Pietikäinen, Face Recognition with Local Binary Patterns, Proc. Eighth European Conf. Computer Vision, pp. 469-481, 2004.
- 5 iv) Jo Chang-yeon, "Face Detection using LBP features", CS 229 Final Project Report, Diciembre 12, 2008.
- v) <http://www.sc.ehu.es/ccwalirx/gwdip3/imageprocessing.pdf> Página web donde se encuentra el libro: "The image processing Handbook Sixth Edition", John C. Russ, editorial CRC Press, pág. 269
- 10 vi) Antonio Manuel Rojas Latorre, "Reconocimiento biométrico de las personas basado en los labios", PFC EITE-ULPGC, Septiembre 2011.
- vii) H. Shirgahi, S. Shamsirband, H. Motameni and P. Valipour, "A New Approach for Detection by Movement of Lips Base on Image Processing and Fuzzy Decision", Septiembre 2008.
- 15 viii) Anuj Mehra, Mahender Kumawat, Rajiv Ranjan, Bipul Pandey, Sushil Ranjan, Anupam Shukla y Ritu Tiwari, "Expert System for Speaker Identification Using Lip Features with PCA", Second International Workshop on Intelligent Systems and Applications (ISA), páginas 1 -4, Wuhan (China), 2010.
- 20 ix) Nicolas EVENO, Alice CAPLIER, Pierre-Y ves COULON, "A Parametric Model for Realistic Lip Segmentation", 2002.
- 25 x) Ojala, T., Pietikäinen, M., Harwood, D.: Un estudio comparativo de las medidas de la textura con la clasificación basada en las distribuciones de características, reconocimiento de formas, páginas 51-59, 29 (1996)
- 30 xi) Ekman, P., Friesen, W., "Facial Action Coing System: A Technique for the Measurement of Facial Movements", Cosulting Psychologist Press, Palo Alto, CA, 1978.
- xii) Bashar, F., Khan, A., Ahmed, F., & Kabir, M. Robust facial expression recognition based on median ternary pattern (MTP). In Electrical Information and Communication Technology (EICT), 2013 International Conference on (pp. 1-5). IEEE, Febrero, 2014.
- 35 xiii) LoBue, V., Matthews, K., Harvey, T., & Thrasher, C., "Pick on someone your own size: The detection of threatening facial expressions posed by both child and adult models". Journal of experimental child psychology, 118, 134-142.2014.
- 40 xiv) Pantic, M., Patras, I., "Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments From Face Profile Image Sequences", IEEE Transactions on System, Man and Cybernetics-Part B: Cybernetics, vol. 36, no.2, pp. 443-449, April 2006.
- 45 xv) Huang, Yi, et al. "Human emotion recognition using the adaptive sub-layer-compensation based facial edge detection". Circuits and Systems (ISCAS), 2013 IEEE International Symposium on. IEEE, 2013.

xvi) Fu, M. H., Kuo, Y.H., Lee, K. R., "Fusing Remote Control Usage and Facial Expression for Emotion Recognition" Fourth International Conference on Innovative Computing, Information and Control, pp. 132-135, Kaohsiung, December 7-9, 2009.

5 xvii) José Gustavo Hernández Travieso, "Sistema detector de emociones basado en imágenes faciales", PFC EITE-ULPGC, Julio 2011.

xviii) WO2010133661 AI (TESSERA TECHNOLOGIES IRELAND LIMITED) "Identifying facial expressions in acquired digital images".

10

Otra modalidad usada a la hora obtener una correcta detección de la emoción, es en el estudio de señales fisiológicas, como el EEG tal como se puede comprobar en los siguientes trabajos:

15 xix) Mikhail, M., El-Ayat, K., Coan, J. A., & Allen, J. J. Using minimal number of electrodes for emotion detection using brain signals produced from a new elicitation technique. International Journal of Autonomous and Adaptive Communications Systems, 6(1), 80-97,2013.

20 xx) Petrantonakis, P. C., Hadjileontiadis, L. J., "Emotion Recognition from EEG Using High Order Crossing", IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2,pp. 186-197,2010.

25 Además, la detección morfológica de los labios, ha ayudado a detectar 3 diferentes emociones, como la alegría, tristeza o sorpresa. Igualmente es emoción, pero nunca se detecta su grado. Ejemplo de ello se puede encontrar en el siguiente artículo:

xxi) Ulrik Söderström and Haibo Li, "Emotion recognition and estimation from tracked lip features", 2007.

30

Otra modalidad biométrica utilizada para detectar y distinguir diferentes tipos de emociones, ha sido el análisis de la conductividad de la piel, temperatura corporal, electromiograma, volumen respiratorio, ritmo cardiaco y presión sanguínea:

35 xxii) Gouizi, K., Reguig, F.B., Maaoui, C., "Analysis Physiological Signals for Emotion Recognition", 7° International Workshop on Systems, Signal Processing and their Applications (WOSSPA), pp. 147-150, 2011.

40 La voz puede ser otra modalidad biométrica por la que se puede detectar la emoción, con la ayuda de ciertas características de la voz, como la calidad de voz, el tono de pronunciación, etc., que permiten identificar diferentes emociones. Las siguientes investigaciones que hacen referencia a esta línea de trabajo, son los siguientes:

45 xxiii) Kuchibhotla, Swama, *et al.* "Speech Emotion Recognition Using Regularized Discriminant Analysis". Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013. Springer International Publishing, 2014.

50 xxiv) Moataz M. H. ElAyadi, Mohamed S.Kamel, and Fakhri Karray, "Speech emotion recognition using gaussian mixture vector autoregressive models". International Journal of Engineering and Advanced Technology (IJEAT) Volume-1, Issue-5, June 2012.

xxv) Jean Vroomen, Rene Collier, Sylvie Mozziconacci, "Duration an intonation m emotional speech". Proc. Eurospeech '93. Berlin, Germany, 1993, pp 577-580.

5 Desde este enfoque, las emociones son detectadas a partir de ciertas modalidades y calculando ciertos parámetros, ya sean de la voz, cara, a nivel neuronal, etc. A modo de resumen, esto se logra con ciertas características propias o patrones cualificados en varias señales fisiológicas/biométricas. Pero otro enfoque en el estudio de la Psicología emocional, es caracterizar los parámetros fundamentales de la experiencia emocional. En este sentido, las emociones se puede catalogar por placenteras o no, y positivas o
10 negativas.

A este concepto se le denomina "valencia afectiva". Otra característica fundamental es la intensidad emocional o nivel de excitación, que en psicología se le denomina Arousal (este es el parámetro que trata de cuantificar la invención propuesta). Estas dos medidas
15 ayudan a conocer ciertos parámetros de conducta, a tener un control de las emociones, a caracterizarlas, etc. Bajo este enfoque, existen varios estudios en el que se detectan y cualifican estos parámetros de Valencia y Arousal. Como ejemplo se presentan las siguientes investigaciones, donde se cuantifican estos niveles a partir de diferentes modalidades fisiológicas y biométricas; por ejemplo, mediante el empleo de señales
20 obtenidas con el EEG y otras señales fisiológicas:

xxvi) Duru, D. G., Deniz Duru, A., Barkana, D. E., Sanli, O., & Ozkan, M. "Assessment of surgeon's stress level and alertness using EEG during laparoscopic simple nephrectomy". In Neural Engineering (NER), 2013 6th International IEEE/EMBS Conference on (pp. 452-455). IEEE. Noviembre 2013
25

xxvii) Hidalgo-Muñoz, A. R., *et al.* "EEG study on affective valence elicited by novel and familiar pictures using ERD/ERS and SVM-RFE". Medical & biological engineering & computing 52.2 (2014): 149-158.
30

xxviii) Gerber, Andrew J., *et al.* "An affective circumplex model of neural systems subserving valence, arousal, and cognitive overlay during the appraisal of emotional faces". Neuropsychologia 46.8: 2129-2139,2008.

35 El análisis de las características de la voz, en una locución o diálogo, también ha facilitado la tarea de cuantificar los niveles de Valencia y Arousal:

xxix) Arias, Juan Pablo, Carlos Busso, and Nestor Becerra Yoma. "Shape-based modeling of the fundamental frequency con tour for emotion detection in speech". Computer Speech & Language 28.1 (2014): 278-294.
40

xxx) Kuchibhotla, Swama, *et al.* "Speech Emotion Recognition Using Regularized Discriminant Analysis". Proceedings of the International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2013. Springer International Publishing, 2014.
45

xxxi) Patricia Hernández, Jesús B. Alonso, Miguel-Ángel Ferrer-Ballester, Carlos M. Travieso-González, Juan R. Orozco-Aroyave, Nonlinear Dynamics Characterization of Emotional Speech, Neurocomputing, Ed. Elsevier, (ISBN: 0925-2312), Berlin, 2012.
50

Por tanto, se puede observar que la detección del grado de emoción se realiza por modalidades totalmente diferentes a esta propuesta y bajo este punto de vista, no se ha encontrado ninguna investigación que relacione la cuantificación del nivel emocional con el movimiento de los labios.

5

Otras investigaciones y estudios diferentes a los anteriormente mostrados se centran en determinar la localización de la cara y de los labios, y determinar algunas actividades, pero no se centran en determinar el grado y cuantificación de la emoción, que es el objeto de esta propuesta; tal y como se puede observar en las siguientes patentes:

10

xxxii) xiii) WO2012128382 A1 (SHARP KABUSHIKI KAISHA) "Device and method for lip motion detection"

15

xxxiii) WO2011079458 A1 (NOKIA CORPORATION) "Method and apparatus for local binary pattern based facial feature localization"

xxxiv) WO2011074014 A1 (TATA CONSULTANCY SERVICES LTD.) "A system for lip corner detection using vision based approach"

20

xxxv) WO2011055224 A1 (INSTITUT POLYTECHNIQUE DE GRENOBLE) "Device and method for detecting and monitoring the inner and outer contours of the lips"

xxxvi) US2014050392 A1 (SAMSUNG ELECTRONICS CO., LTD.) "Method and apparatus for detecting and tracking lips"

25

xxxvii) US2013271361 A1 (SAMSUNG ELECTRONICS CO., LTD.) "Method and apparatus for detecting talking segments in a video sequence using visual cues"

xxxviii) US2013226587 A1 (HONG KONG BAPTIST UNIVERSITY) "Lip-password based speaker verification system"

30

xxxix) CN1710595 A (UNIV SHANGHAI JIAOTONG) "Mouth comer positioning method in eludes histogram of the said area is analyzed and treatment of threshold value is carried out for the area so as to obtain binary image"

35

xl) CN102129683 A (UNIV SHANGHAI) "Automatically lip image division method, involves making horizontal set function as zero, using five-point balancing process for processing lip image outline, and removing edge burr for outline to obtain last lip outline"

40

No se deberá confundir con determinar el tipo de emoción, pues en el caso referido en esta invención solo interesa el grado de la emoción y no su tipo. Muchos trabajos han sido desarrollados en la detección de la emoción, pero no en la detección del grado de la emoción a partir del movimiento de los labios. Cabe destacar que fundamentalmente se ha reconocido el grado de la emoción por la voz, pero el único trabajo que usa el movimiento de los labios para la detección del grado o carga de la emoción detectado es el referido en esta propuesta. El usar otra modalidad, como el movimiento de los labios al hablar, es una nueva vía para analizar el grado de la emoción, y sería un modo más que añadir a los sistemas actuales ya existentes basados en electroencefalograma, voz, o medidas corporales. Esta nueva forma de medirlo da mucha más riqueza a un posible estudio multimodal.

45

50

La presente invención tiene por objeto la cuantificación de la carga emocional, en contraposición a la detección del tipo de emoción. La invención, por tanto, tendría potenciales aplicaciones en la detección de patologías neurodegenerativas o diferentes enfermedades mentales, como Alzheimer, Parkinson, síndrome de Asperger, autismo, etc., donde sus principales síntomas se relacionan directamente con el reconocimiento, o la propia falta de interpretación de las emociones, con independencia de la emoción, por lo que se ha identificado como problema técnico a resolver gracias a la presente invención la cuantificación del grado de la emoción, solución no hallada en el estado de la técnica, y no la identificación de la emoción, problema que ya han resuelto otras invenciones. La cuantificación del grado de emoción también ayuda a controlar situaciones de estrés, en trabajos como operaciones quirúrgicas, o cuya necesidad de concentración sea alta, un largo etcétera de posibilidades. Por ello, su análisis y detección es de suma importancia en los tiempos actuales y futuros.

Cabe concluir tras estos antecedentes, que los estudios que se han desarrollado hasta el momento y que han tenido como parámetro característico los labios, han sido utilizados básicamente para el reconocimiento biométrico, para distinguir el tipo de emoción, incluso para reconocer sus características morfológicas. El método propuesto, a diferencia de lo observado en el estado de la técnica, relaciona el nivel de carga emocional o excitación, y el movimiento de los labios, durante el tiempo en que la persona se encuentre hablando, y presenta la ventaja de no ser invasivo, pues con una sola cámara pudiera captar y analizar el vídeo (véase la figura 1), manteniendo un grado de colaboración por parte del usuario muy bajo o nulo; a diferencia de la voz, que requiere un direccionamiento del micrófono más el ruido ambiente, o sensores fisiológicos de contacto directo al cuerpo humano.

Sumario de la invención

La presente invención se refiere a un método para la determinación de la cuantificación de la carga emocional a partir del movimiento de los labios siguiendo cinco pasos:

i) Detección de la zona bucal: Primeramente se realiza una detección facial, y se aplican varios subprocesos para obtener la región de interés (ROI), que en este caso son los labios, con el fin de distinguirlos.

ii) Detección de los labios mediante una transformación de la colorimetría: Se realiza una transformada de color automatizada, con la cual se consigue segmentar el color rojo de los labios, esto favorece la lectura del movimiento labial, diferenciándolo de cualquier otra parte de la boca, como puede ser la barba, granos, manchas, sombras, etc.

iii) Filtrado de realce y binarización de los labios: Después de esta transformada se realizan varios procesados de realce, y mejora de la imagen, para luego ser binarizada.

iv) Detección de labio superior e inferior mediante transformación en polares: Una vez obtenidos los labios perfectamente diferenciados, el proceso sigue realizando una transformación a polares, con el fin de separar ambos labios en dos regiones, esto facilita la implantación de un sistema con acumuladores, donde registra y almacena los movimientos de los labios.

v) Cuantificación del grado de la emoción mediante el estudio estadístico del movimiento labial: Estos movimientos son analizados estadísticamente y se logra evaluar el índice de

carga o intensidad emocional, a través del movimiento de los labios, producidos por los momentos en el que un paciente o persona se encuentre realizando la prueba.

5 Se han establecido 3 grados de carga emocional: neutra, baja y alta, simplificando de enorme manera los diferentes grados de emoción o Arousal. Estas cargas, pueden llegarse a relacionar con diferentes tipos de emociones, según sea el nivel con el que hayan sido detectadas.

10 **Descripción de las figuras**

La figura 1 muestra de forma esquemática la interfaz básica desarrollada.

La figura 2 muestra de forma esquemática los subprocesos que componen el método.

15 La figura 3 muestra de forma esquemática el patrón local binario.

La figura 4 muestra de forma esquemática el efecto Umbralizar - Procesado Morfológico.

20 La figura 5 muestra de forma esquemática el efecto Abrir y Cerrar - Procesado Morfológico.

La figura 6 muestra de forma esquemática el efecto erosionar - Procesado Morfológico.

25 La figura 7 muestra de forma esquemática el efecto Dilatar - Procesado Morfológico.

La figura 8 muestra de forma esquemática el efecto Suavizar - Procesado Morfológico.

30 La figura 9 muestra de forma esquemática la transformación a polares, con ambos labios separados y con una muestra de los acumuladores.

La figura 10 muestra de forma esquemática el detector de contornos.

35 La figura 11 muestra de forma esquemática un ejemplo de gráfica obtenida tres diferentes cargas emocionales, Baja, Media y Alta.

La figura 12 muestra de forma esquemática un ejemplo del termómetro - grado emoción.

Descripción detallada de una realización preferida de la invención

40 Aunque la invención se describe en términos de una realización específica preferida, será fácilmente evidente para los expertos en esta técnica que se pueden hacer diversas modificaciones, redistribuciones y reemplazos. El alcance de la invención está definido por las reivindicaciones adjuntas a la misma.

45 La invención propuesta consiste en un método que aplica varios subprocesos hasta llegar a la cuantificación del grado de la emoción. El primero, realiza una detección facial y un recorte de la zona de interés (ROI), los labios (i). A continuación, se aplica un preprocesado de imagen para contrarrestar los labios frente al resto del ROI; mediante una transformación del color (ii). Esta imagen pasa a ser tratada, en la misma etapa, por
50 diversos procesados, para realzar y mejorar los labios obtenidos (iii).

Con los labios detectados, se realiza una transformación de coordenadas, concretamente a polares, con el fin de separar ambos labios, en dos partes diferentes; para así, registrar y analizar todos sus movimientos (iv). Estos movimientos serán evaluados estadísticamente, y obtener así un índice o nivel de carga emocional (v).

5

A continuación, se describen en detalle los subprocesos enumerados previamente.

(i) En el sistema de detección facial, se emplea el método LBP, del inglés, Local Binary Pattern. Este patrón es un descriptor de micro-imágenes bastante robusto, Figura 3. Su modo de funcionamiento es bastante sencillo, se trata de un operador, que consiste en una simple matriz 3X3, donde los valores de la imagen se binarizan por el método de umbralización, tomando como referencia el valor central y evaluando cada vecino, tomando como resultado un 0 o 1, según supera o no dicho umbral. De este modo, se obtendrá una cadena de 8 bits que etiqueta a dicho píxel central. Si se procesa toda la imagen con este patrón, se logra etiquetar toda la imagen. Con estas etiquetas, se puede construir un histograma completo de la imagen, pero para identificar cada parte de un rostro, es importante tener información espacial. Por ello se recurre a dividir la imagen en R_m regiones. De este modo se extraerá información de cada región de forma independiente, obteniendo finalmente un histograma espacial, concatenando cada uno de los histogramas, obtenidos en cada región.

El potencial de este operador es que a diferencia de muchos otros este método consigue ser bastante robusto y eficiente ante cambios de luz, con lo que para imágenes en movimiento está técnica es muy eficaz.

25

Una vez obtenida la imagen facial, se realiza un recorte de la zona de la boca, empleando un método donde consigue separar la parte de la boca, segmentando el rostro, en varias proporciones equitativas. Si se divide el rostro horizontalmente y verticalmente en 3/4 partes, fácilmente se puede extraer la parte de la boca, localizada en la parte inferior de la misma.

30

(ii) Una vez obtenido el recorte de la boca, la imagen se separa en tres canales de color: rojo, verde y azul. Estos canales son procesados, para realizar una transformada de color, mediante la ecuación:

35

$$\text{Transformada_de_color} = \text{Azul} + \text{Rojo} - \text{escalar} * \text{Verde}$$

Esta transformada ayudará a diferenciar el color de los labios del resto de la boca. Su modo de funcionamiento se explica mediante el trabajo que desempeña el escalar que multiplica al canal verde. Gracias a este factor, se consigue controlar el color verde, que es el que mayormente afecta a las otras partes de la boca, como puede ser la piel. Por ello, de forma automatizada, se varía el valor de este escalar, cuyo parámetro fundamental de control es la media resultante de dicha transformada. Esta media, está estimada para lograr diferenciar de la mejor forma posible, la parte de los labios. De modo que diferentes zonas de la boca, como la barba, manchas, etc. son discriminadas para poder evaluar correctamente el movimiento labial. De forma empírica se ha establecido un rango de valores idóneos que van de -18 a -30, siendo este último el más agresivo.

(iii) A la salida de la transformada de color, se obtienen los labios, pero no siempre se consigue que estén bien definidos. Los brillos, manchas, ruido, y la resolución de las

50

imágenes, provoca que existan partes con falta de información, pequeñas discontinuidades, orificios etc., que pueden ser mejoradas mediante la utilización de procesados morfológicos, como son la dilatación, erosión o el suavizado.

5 En la erosión (véase la Figura 6), se consigue eliminar pequeños puntos, que pueden aparecer alrededor de la boca, la barba, pecas etc., pueden ser los causantes de los mismos. Por lo contrario, el proceso de dilatación ayuda a tapar pequeños huecos, y
10 orificios, que se pueden encontrar en los labios, como se puede ver en la figura 7. Para continuar mejorando los labios, se empleó un suavizado tipo "Gaussiano" (figura 8), con el cual no se perdía mucha definición, y terminaba de mejorar pequeñas imperfecciones. Una vez realizadas los diferentes procesados morfológicos iniciales, la imagen se binariza, utilizando un umbral que se establece por el método de OTSU, figura 4, que usa el valor de la media de la imagen como dicho umbral, para dicha binarización. De este modo se obtienen de color blanco los labios, procesados.

15 Una vez binarizada, se ha procedido a utilizar otros procedimientos morfológicos, como apertura y cierre, que realizan operaciones combinadas de dilatación y erosión, con el fin de afinar un poco más las posibles imperfecciones, que se puedan encontrar en los labios, sobre todo en los momentos de apertura de la boca, figura 5.

20 (iv) Aun realizando estas operaciones, siempre pueden existir pequeños defectos en la imagen, como manchas, sombras u orificios que perjudican la diferenciación de los labios. Por ello se ejecuta un detector de contorno, cuyo umbral se realiza respecto al tamaño de los diferentes contornos cerrados que se pueden encontrar en la imagen. Por
25 ello, solo se dará a la salida del detector de bordes, al más grande que corresponderá al de los labios, véase la figura 10.

Con estos procesados, se tendrán los labios de tal forma que se pueda cuantificar perfectamente la apertura y cierre de los mismos. Para ello y para facilitar esta tarea, se
30 recurre a realizar una transformación de coordenadas, de cartesianas a polares. Antes de esto es necesario localizar el centro de masas de la imagen, que corresponde al centro de los labios detectados. A la salida de la transformación, se obtiene una imagen como la de la figura 9. Donde gracias a esta transformación, se pueden separar y medir, ambos labios, tanto el superior como el inferior.

35 (v) Para la medida de la apertura y cierre de los labios, se emplean dos matrices superpuestas en las zonas de trabajo de los labios, con el fin de operar como acumuladores, donde se van rellenando según corresponda, en cada momento. Un punto importante a tener en cuenta en este momento es que para diferenciar los movimientos
40 mínimos de los labios, con respecto a la boca cerrada, es necesario realizar un calibrado inicial para determinar este umbral. Para ello se determina un método de calibrado en el que se exige a la persona a evaluar mantener la boca cerrada durante un determinado tiempo, que equivaldría a unos 140 frames, para tener una colección de valores suficientes con los que cuantificar el estado de la boca cerrada. En cambio para la boca
45 abierta, se establece un valor que se ha obtenido de forma empírica, bajo diversas pruebas.

Después del calibrado, los valores de los acumuladores obtenidos se normalizan y se van almacenando en un vector de 50 elementos, lo cual es necesario para relacionar los
50 estados anteriores con el actual a analizar, y así evidenciar los diferentes estados. Para ello, en cada frame se calcula la media y la varianza instantánea de dicho vector. Estos

valores sirven de ayuda, para evaluar los movimientos y variaciones de las aperturas y cierres de la boca, mediante la siguiente fórmula:

$$\text{nivelDeCargaEmocional} = (\alpha \times \text{varianza}) + (\beta \times \text{media})$$

5

donde α y β son variables ajustables para alcanzar el mejor nivel de la detección de la carga emocional.

10 En este punto, basta con representar los diferentes valores que se obtienen en cada frame con la fórmula anterior. Mediante los experimentos realizados con las diferentes bases de datos trabajadas, se han determinado 3 umbrales diferentes, establecidos de forma empírica, que corresponden a tres cargas emocionales diferentes: baja, neutra y alta (véanse las figuras 11 y 12). Esto se ha estimado así para simplificar de cierto modo los diversos grados de nivel de excitación emocional. Los valores son los siguientes:

15

Bajo: $5 \leq x < 20$

Neutro: $20 \leq x \leq 35$

20

Alto: $35 < x$

REIVINDICACIONES

1. Método para el cálculo de la cuantificación de la carga emocional (Arousal) mediante el estudio estadístico del movimiento de los labios, en el transcurso de una conversación o de cualquier otra forma de expresión no verbal, registrado mediante un método de captación de imágenes en movimiento que comprende:
- 5
- (i) Medida de la apertura y cierre de los labios mediante matrices superpuestas en las zonas de trabajo de los labios, con el fin de operar como acumuladores,
- 10
- (ii) Diferenciación de los movimientos mínimos de los labios, con respecto a la boca cerrada, mediante un calibrado inicial que implica el registro en frames de la imagen captada de los labios para determinar un umbral que cuantifique el estado de la boca cerrada y un valor para la boca abierta,
- 15
- (iii) Normalización y almacenamiento de los valores de los acumuladores obtenidos para relacionar los estados anteriores con el actual a analizar, y así evidenciar los diferentes estados,
- 20
- (iv) Representación de los diferentes valores que se obtienen en cada frame mediante un número de umbrales diferentes que corresponden a grados de nivel de excitación emocional diferentes.

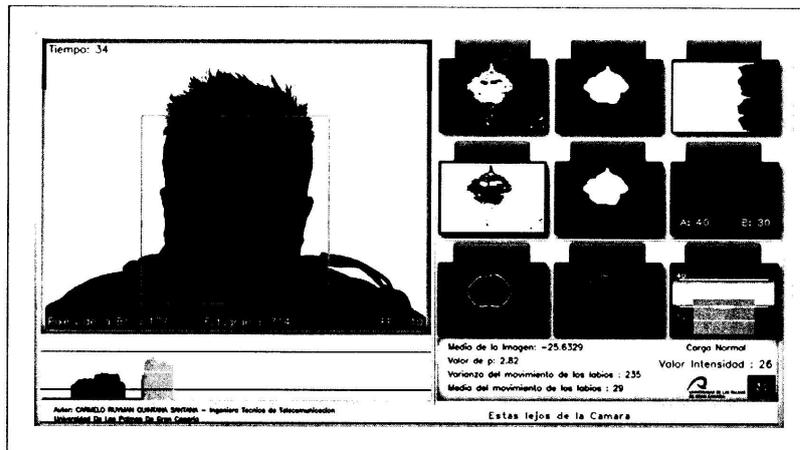


Figura 1

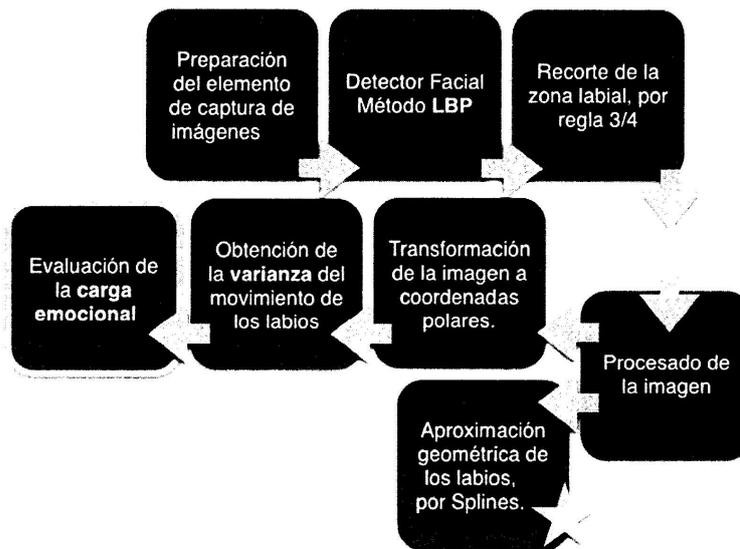


Figura 2

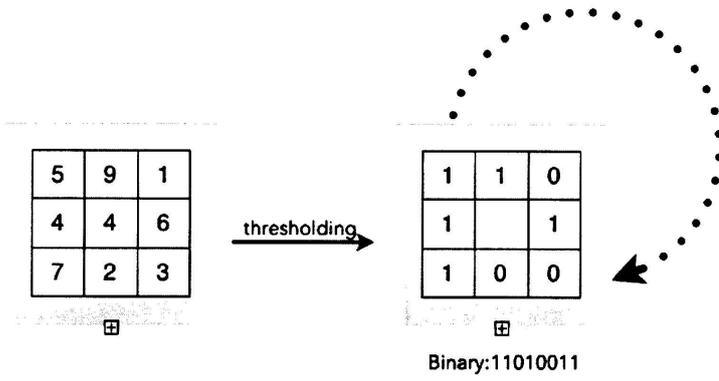


Figura 3

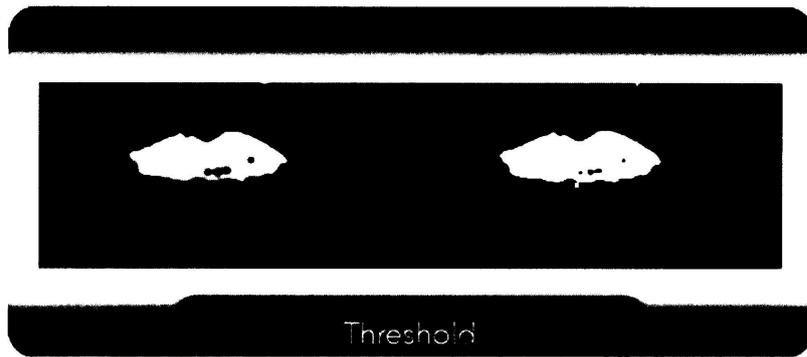


Figura 4

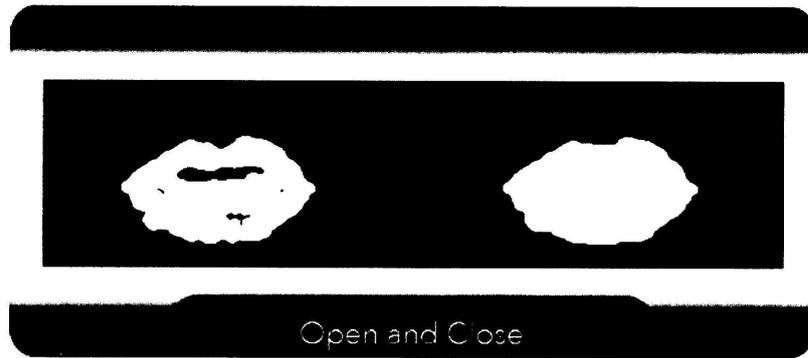


Figura 5

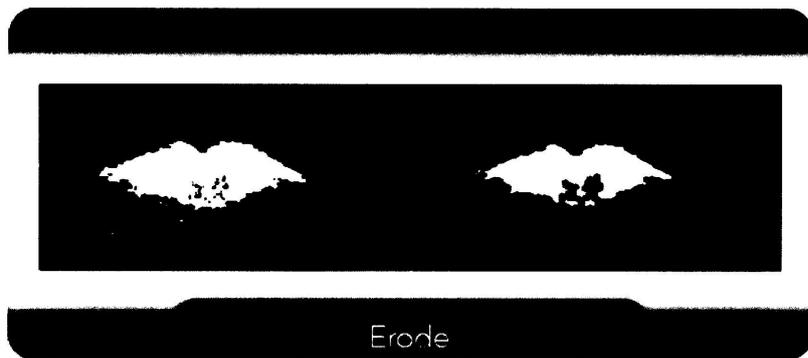


Figura 6

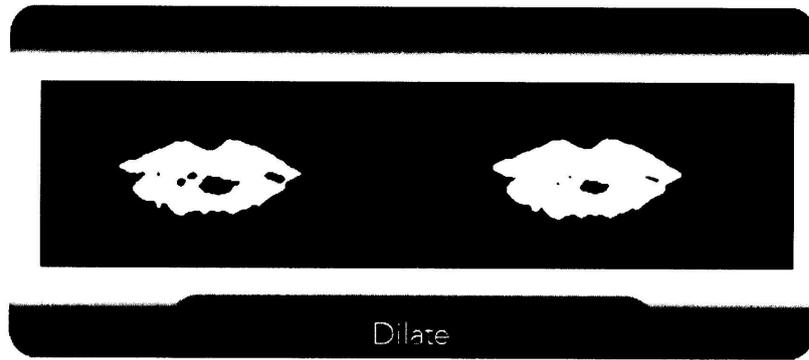


Figura 7

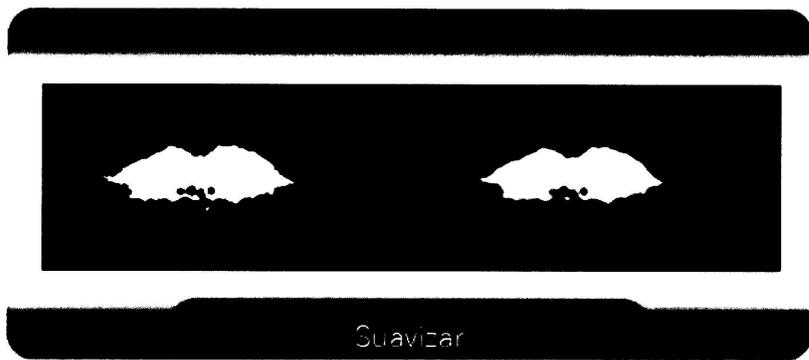


Figura 8

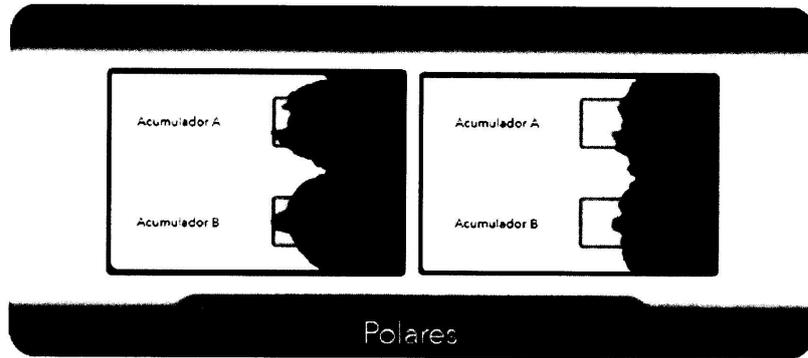


Figura 9

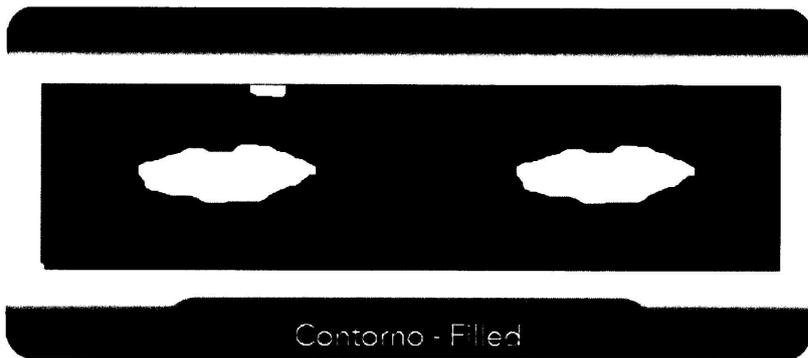


Figura 10

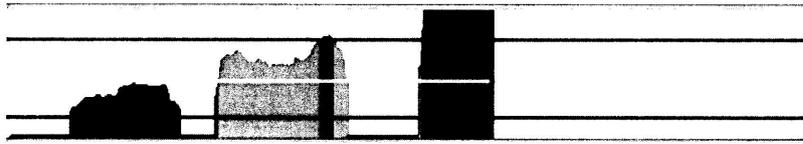


Figura 11

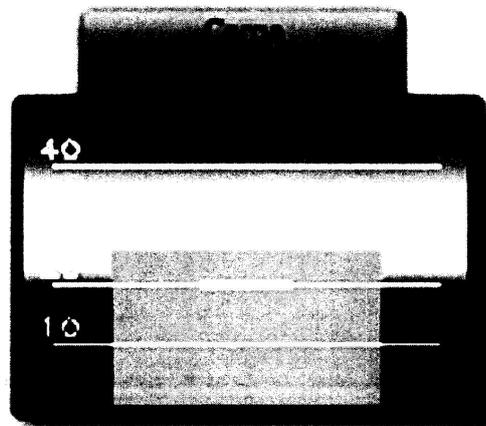


Figura 12



- ②¹ N.º solicitud: 201500005
②² Fecha de presentación de la solicitud: 12.12.2014
③² Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤¹ Int. Cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑤ ⁶ Documentos citados	Reivindicaciones afectadas
X	QUINTANA SANTANA, CARMELO RUYMAN: "Implementación de un detector de habla basado en vídeo aplicado a la evaluación de la carga emocional", Proyecto Fin de Carrera, Universidad de Las palmas de Gran Canaria, 28 de abril de 2014, capítulo 4, páginas 45-76, figuras 4.1-4.25.	1
X	QUINTANA SANTANA, CARMELO RUYMAN. "PFC Detector de Labios". Vídeo Youtube. 28 de junio de 2013. Recuperado de < https://www.youtube.com/watch?v=YfPj41nUBAo >	1
X	QUINTANA SANTANA, CARMELO RUYMAN. "Carga emocional v1". Vídeo Youtube. 16 de julio de 2013. Recuperado de < https://www.youtube.com/watch?v=FPidn4FVNQg >	1
A	SAHA, C., GHOSH, K.: "Estimation of Facial Expression Intensity from a Sequence of Binary Face Images"; Publicado en: International Conference on Image Information Processing (ICIIP 2011); Fecha de publicación: 3-5 nov. 2011; DOI: 10.1109/ICIIP.2011.6108935; URL: < http://ieeexplore.ieee.org/document/6108935/ >	1
A	KIM, J., LEE, S., NARAYANAN, S.: "An Exploratory Study of the Relations between Perceived Emotion Strength and Articulatory Kinematics"; Publicado en: INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31. 2011, páginas 2961-2964; URL: < http://www.isca-speech.org/archive/interspeech_2011/i11_2961.html >	1

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe
21.12.2016

Examinador
A. Astudillo Lizaga

Página
1/5



②① N.º solicitud: 201500005

②② Fecha de presentación de la solicitud: 12.12.2014

③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

⑤① Int. Cl.: Ver Hoja Adicional

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	LEE, S., NARAYANAN, S.: "Effects of Emotion on the Lower Lip Movements at Phrase Boundaries", Proceedings of Speech Prosody, Shanghai, CN, 2012; URL: < https://www.researchgate.net/publication/229438952_Effects_of_Emotion_on_the_Lower_Lip_Movements_at_Phrase_Boundaries >	1

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

para todas las reivindicaciones

para las reivindicaciones nº:

Fecha de realización del informe
21.12.2016

Examinador
A. Astudillo Lizaga

Página
2/5

CLASIFICACIÓN OBJETO DE LA SOLICITUD

G06K9/00 (2006.01)

G06T7/20 (2006.01)

G10L15/25 (2013.01)

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06K, G06T, G10L

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI

Fecha de Realización de la Opinión Escrita: 21.12.2016

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones	SI
	Reivindicaciones 1	NO
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones	SI
	Reivindicaciones 1	NO

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	QUINTANA SANTANA, CARMELO RUYMAN: "Implementación de un detector de habla basado en vídeo aplicado a la evaluación de la carga emocional", Proyecto Fin de Carrera, Universidad de Las palmas de Gran Canaria, 28 de abril de 2014, capítulo 4, páginas 45-76, figuras 4.1-4.25.	28 de abril de 2014

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

De los documentos encontrados para la realización de este informe, el documento D01 se considera el más próximo del estado de la técnica al objeto de la reivindicación 1.

A continuación, se reproduce literalmente la reivindicación 1 indicándose entre paréntesis las partes del documento D01 correspondientes:

Método para el cálculo de la cuantificación de la carga emocional (Arousal) mediante el estudio estadístico del movimiento de los labios (**ver capítulo 4, páginas 45-76, figuras 4.1-4.25**), en el transcurso de una conversación o de cualquier otra forma de expresión no verbal, registrado mediante un método de captación de imágenes en movimiento que comprende:

- (i) Medida de la apertura y cierre de los labios mediante matrices superpuestas en las zonas de trabajo de los labios, con el fin de operar como acumuladores.
- (ii) Diferenciación de los movimientos mínimos de los labios, con respecto a la boca cerrada, mediante un calibrado inicial que implica el registro en frames de la imagen captada de los labios, para determinar un umbral que cuantifique el estado de la boca cerrada y un valor para la boca abierta,
- (iii) Normalización y almacenamiento de los valores de los acumuladores obtenidos para relacionar los estados anteriores con el actual a analizar, y así evidenciar los diferentes estados,
- (iv) Representación de los diferentes valores que se obtienen en cada frame mediante un número de umbrales diferentes que corresponden a grados de nivel de excitación emocional diferentes.

Así pues, el documento D01 divulga idénticamente todas las etapas del método definido en la reivindicación 1 y por tanto dicha reivindicación no satisface los requisitos de novedad y actividad inventiva según se establece en los artículos 6.1 y 8.1 de la Ley de Patentes 11/1986.