



Polibits

ISSN: 1870-9044

polibits@nlp.cic.ipn.mx

Instituto Politécnico Nacional

México

Solís Villarreal, José Francisco; Yáñez Márquez, Cornelio; Suárez Guerra, Sergio  
Reconocimiento automático de voz emotiva con memorias asociativas Alfa-Beta SVM

Polibits, vol. 44, 2011, pp. 19-23

Instituto Politécnico Nacional

Distrito Federal, México

Disponible en: <http://www.redalyc.org/articulo.oa?id=402640458003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# Reconocimiento automático de voz emotiva con memorias asociativas Alfa-Beta SVM

José Francisco Solís Villarreal, Cornelio Yáñez Márquez y Sergio Suárez Guerra

**Resumen**—Una de las de investigación de mayor interés y con más crecimiento en la actualidad, dentro del área de procesamiento de voz, es el reconocimiento automático de emociones, el cual consta de 2 etapas; la primera es la extracción de parámetros a partir de la señal de voz y la segunda es la elección del modelo para hacer la tarea de clasificación. La problemática que actualmente existe es que no se han identificado aún los parámetros más representativos del problema ni tampoco se ha encontrado al mejor clasificador para hacer la tarea. En este artículo se introduce un nuevo modelo asociativo de reconocimiento automático de voz emotiva basado en las máquinas asociativas Alfa-Beta SVM, cuyas entradas se han codificado como representaciones bidimensionales de la energía de las señales de voz. Los resultados experimentales muestran que este modelo es competitivo en la tarea de clasificación automática de emociones a partir de señales de voz [1].

**Palabras Clave**—Reconocimiento de voz emotiva, memorias asociativas Alfa-Beta SVM, procesamiento de voz.

## Automatic Emotional Speech Recognition with Alpha-Beta SVM Associative Memories

**Abstract**—One of the research lines of interest and more growth at present, within the area of voice processing is automatic emotion recognition. It is vitally important the study of speech signal not only to extract information about what is being said, but how is being said, this in order to be closer to the human-machine interaction. In literature the procedure of automatic emotion recognition consists of two stager, the first is the extraction of parameters from the voice signal and the second is the choice of model for the classification task, the problem that currently exists is not yet identified the most representative parameters of the problem nor has found the best classifier for the task, but have not yet been tested several models, this paper presents a two-dimensional representation of energy as data entry for Alpha-Beta associative machines SVM (Support Vector Machine) for the classification of emotions.

**Index Terms**—Emotional speech recognition, Alpha-Beta SVM associative memories, voice processing.

## I. INTRODUCCIÓN

EN la última década, el interés en el reconocimiento automático de emociones ha ido creciendo de manera notable. El objetivo de este tema es mejorar la interacción hombre-máquina, haciendo posible que los equipos de cómputo puedan rescatar información afectiva más que el contenido de lo hablado. Esto es necesario para el propósito de tener una comunicación más natural [2].

No obstante que muchos trabajos han contribuido con diferentes enfoques, el rendimiento alcanzado en esta tarea todavía no es suficientemente bueno. En reconocimiento de voz emocional, uno de los principales problemas es localizar los rasgos que se ajusten para llevar a cabo la clasificación. Los investigadores han reportado una gran variedad de parámetros, pero hasta la fecha no se han identificado cuáles son los parámetros más representativos de la información afectiva de la señal de voz que sean útiles para realizar la clasificación de estados emocionales. [3]. A continuación se describen brevemente las aportaciones reciente más importantes en esta área de investigación.

En [4], se usaron 2 bases de datos; la de Berlín [1], que contiene emociones actuadas y la de SmartKom, la cual incluye emociones espontáneas. El enfoque en este trabajo fue el uso de la detección del género previo al reconocimiento de emociones, y se alcanzó un 90% de reconocimiento para la detección del género. Con respecto a la base de datos de Berlín, los autores reportan un desempeño aproximado del 80%.

En otro trabajo publicado en [3], se usaron 102 parámetros de dos bases de datos, la de Berlín y una Polaca. La selección de rasgos fue realizada mediante árboles binarios de decisión, buscando tripletas óptimas, y midiendo la utilidad del conjunto de parámetros mediante su correlación; si la correlación es alta, se desecha el conjunto de parámetros y se selecciona otro al azar. Cada conjunto de parámetros contiene un parámetro de frecuencia, otro de energía y otro de duración. Para la base de datos de Berlín, se alcanzó un rendimiento del 72.04% de reconocimiento usando 6 emociones de este corpus.

En otro enfoque, reportado en [5], se seleccionaron 68 parámetros de la base de datos de Berlín, y se usaron redes neuronales artificiales (backpropagation) como modelo de clasificación; también se realizó una clasificación de género previa a la clasificación de emociones, obteniendo un desempeño del 79.47% de reconocimiento usando 6

Manuscrito recibido el 03 de marzo de 2011. Manuscrito aceptado para su publicación el 22 de junio de 2011.

Los autores trabajan en el Centro de Investigación en Computación del Instituto Politécnico Nacional, México, D.F. (tlilectic.mixtzin@gmail.com, cyanez@cic.ipn.mx, ssuarez@cic.ipn.mx).

emociones de la base de datos (enojado, feliz, miedo, neutral, tristeza y aburrido).

En [14] se reporta el uso de un modelo basado en los  $k$ -vecinos más cercanos con una estimación de costo del error. Los investigadores usan las 7 emociones de la base de datos de Berlín [1] y reportan un resultado del 82.44% de reconocimiento de emociones.

El resto del artículo está organizado como sigue: en la sección II se describe el modelo de las memorias asociativas Alfa-Beta SVM, cuyo algoritmo es la base del modelo asociativo de reconocimiento automático de voz emotiva, el cual es introducido en la sección III a través del diseño experimental. Mientras que en la sección IV se detallan los resultados experimentales y la discusión correspondiente, en la sección V se plantean las conclusiones del este trabajo de investigación. Finalmente, se incluyen las referencias.

## II. MEMORIAS ASOCIATIVAS ALFA-BETA SVM

El modelo de clasificación usado en este trabajo es la Memoria Asociativa Alfa-Beta SVM [8-9]. Los conceptos básicos concernientes a las memorias asociativas han sido reportados en varios trabajos [10-12]; sin embargo, en este artículo se usan la notación y conceptos introducidos en [10]. Una memoria asociativa  $M$  relaciona patrones a la manera de un sistema de entrada-salida:  $x \rightarrow M \rightarrow y$  donde  $x$  es un patrón de entrada siendo  $y$  es un patrón de salida. Por cada patrón de entrada se forma una asociación con el correspondiente patrón de salida. La  $k$ -ésima asociación está dada por  $(x^k, y^k)$ , donde  $k$  es un entero positivo. La Memoria Asociativa  $M$  está representada por una matriz cuya  $ij$ -ésima componente es  $m_{ij}$ .

La matriz  $M$  es generada a partir del conjunto fundamental, el cual es representado como:  $\{(x^\mu, y^\mu) | \mu = 1, 2, \dots, p\}$ , donde  $p$  es la cardinalidad del conjunto.

Si  $x^\mu = y^\mu \forall \mu \in \{1, 2, \dots, p\}$ , la memoria  $M$  es autoasociativa; de otro modo, es heteroasociativa. La versión distorsionada del patrón  $x^k$  a ser recuperado, está denotada como  $\tilde{x}^k$ . Se dice que la recuperación es correcta cuando se recupera  $y^k$  a partir de  $\tilde{x}^k$ .

Las memorias asociativas Alfa-Beta se basan en dos operadores binarios; el operador  $\alpha$  es usado en la fase de aprendizaje y el operador  $\beta$  es útil para la fase de recuperación. Sean los conjuntos  $A = \{0, 1\}$  y  $B = \{0, 1, 2\}$ ; entonces los operadores  $\alpha$  y  $\beta$  están definidos de manera tabular en la TABLA I y en la TABLA II, respectivamente.

Los conjuntos  $A$  y  $B$ , los operadores  $\alpha$  y  $\beta$ ,  $\wedge$  (mínimo) y  $\vee$  (máximo) forma el sistema algebraico que es la base matemática para las Memorias Asociativas Alfa-Beta.

Todos los conceptos básicos descritos anteriormente [10], son necesarios para describir el algoritmo principal de Alfa-Beta SVM [8,9]. Se tiene un problema de reconocimiento de

TABLA I  
 $\alpha: A \times A \rightarrow B$

$x$	$y$	$\alpha(x, y)$
0	0	1
0	1	0
1	0	2
1	1	1

TABLA II  
 $\beta: B \times A \rightarrow A$

$x$	$y$	$\beta(x, y)$
0	0	0
0	1	0
1	0	0
1	1	1
2	0	1
2	1	1

patrones, donde el conjunto fundamental se describe como  $\{(x^\mu, y^\mu) | \mu = 1, 2, \dots, p\}$ ,  $x^\mu \in A^n \forall \mu \in \{1, 2, \dots, p\}$ ,  $n, p \in \mathbb{Z}^+$  y  $A = \{0, 1\}$ . El algoritmo de Alfa-Beta SVM tiene dos fases, la fase de aprendizaje y la de recuperación.

### Fase de aprendizaje:

1. Del conjunto fundamental, se obtiene el vector soporte  $S$ , cuya  $i$ -ésima componente se calcula así:

$$S_i = \begin{cases} \bigwedge_{k=1}^{p/2} \beta(x_i^{2k-1}, x_i^{2k}) & \text{si } p \text{ es par} \\ \beta \left[ \bigwedge_{k=1}^{(p-1)/2} \beta(x_i^{2k-1}, x_i^{2k}), x_i^p \right] & \text{si } p \text{ es non} \end{cases}$$

2. Para cada  $\mu \in \{1, 2, \dots, p\}$  se forma el vector  $x^\mu|_S$ , a fin de obtener, a partir de estos resultados, el conjunto fundamental restringido  $\{(x^\mu|_S, x^\mu|_S) | \mu = 1, 2, \dots, p\}$ .

3. Para cada  $\mu \in \{1, 2, \dots, p\}$  se forma el vector  $\overline{x^\mu}$  (vector negado de  $x^\mu$ ). Con los  $p$  vectores negados, se obtiene el conjunto fundamental negado  $\{(\overline{x^\mu}, \overline{x^\mu}) | \mu = 1, 2, \dots, p\}$ .

4. Del conjunto fundamental negado, se calcula el vector soporte  $\hat{S}$  (como en el paso 1, usando el conjunto fundamental negado).

5. De manera similar a como se hizo en el paso 2, para cada  $\mu \in \{1, 2, \dots, p\}$  se forma el vector  $\overline{x^\mu}|_{\hat{S}}$ , a fin de obtener, a partir de estos resultados, el conjunto fundamental negado restringido  $\{(\overline{x^\mu}|_{\hat{S}}, \overline{x^\mu}|_{\hat{S}}) | \mu = 1, 2, \dots, p\}$ .

### Fase de recuperación:

Siendo  $\tilde{x} \in A^n$  un patrón de entrada cuyo patrón asociado  $x^\mu$  es desconocido:

1. Se obtiene la restricción  $\tilde{x} \mid_s$ .
2. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\tau(\tilde{x} \mid_s, x^\mu \mid_s)$ , donde la transformada  $\tau$  del vector  $x \in A^n$  con respecto al vector  $y \in A^n$  da como resultado un vector  $\tau(x, y)$  de dimensión  $n$ , cuya  $i$ -ésima componente se calcula de la siguiente manera:  $[\tau(x, y)]_i = \beta[x_i, \alpha(0, y_i)]$ .
3. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\tau(x^\mu \mid_s, \tilde{x} \mid_s)$ .
4. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\theta(\tilde{x} \mid_s, x^\mu \mid_s)$ , donde la transformada  $\theta$  del vector  $x \in A^n$  con respecto al vector  $y \in A^n$  da como resultado un escalar  $\theta(x, y)$  definido así:  $\theta(x, y) = \sigma_n[\tau(x, y)] + \sigma_n[\tau(y, x)]$ , donde el escalar  $\sigma_i(x)$  representa el número de componentes con valor 1 que contiene el vector  $x \in A^n$  en las primeras  $i$  componentes, con  $1 \leq i \leq n$ .
5. Encontrar el valor  $\psi \in \{1, 2, \dots, p\}$  para el cual se cumple esta expresión:  $\theta(\tilde{x} \mid_s, x^\psi \mid_s) = \bigwedge_{\mu=1}^p \theta(\tilde{x} \mid_s, x^\mu \mid_s)$ .
6. Se obtiene  $\bar{\tilde{x}}$  (el vector negado de  $\tilde{x} \in A^n$ ).
7. Se obtiene la restricción  $\bar{\tilde{x}} \mid_s$ .
8. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\tau(\bar{\tilde{x}} \mid_s, x^\mu \mid_s)$ .
9. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\tau(x^\mu \mid_s, \bar{\tilde{x}} \mid_s)$ .
10. Para cada  $\mu \in \{1, 2, \dots, p\}$  se obtiene  $\theta(\bar{\tilde{x}} \mid_s, x^\mu \mid_s)$ .
11. Encontrar el valor  $\varphi \in \{1, 2, \dots, p\}$  para el cual se cumple esta expresión:  $\theta(\bar{\tilde{x}} \mid_s, x^\varphi \mid_s) = \bigwedge_{\mu=1}^p \theta(\bar{\tilde{x}} \mid_s, x^\mu \mid_s)$ .
12. Si  $\theta(\tilde{x} \mid_s, x^\psi \mid_s) \leq \theta(\bar{\tilde{x}} \mid_s, x^\varphi \mid_s)$ , entonces  $\omega = \psi$ ; de otro modo  $\omega = \varphi$ .
13. Se obtiene  $(x^\omega \mid_s)^S$ , vector que es precisamente  $x^\omega$ .

No obstante ser de reciente creación, el algoritmo de las Memorias Asociativas Alfa-Beta SVM ha sido aplicado con éxito en algunas tareas de reconocimiento de patrones [8-9], arrojando resultados comparables a los de otros algoritmos de actualidad, y en algunos casos superando su desempeño.

### III. MODELO PROPUESTO Y DISEÑO EXPERIMENTAL

El modelo de reconocimiento automático de voz emotiva propuesto en este trabajo de investigación, resulta de la fusión del modelo de las Memorias Asociativas Alfa-Beta SVM y la selección, cálculo y representación de los parámetros de señales de voz. Estas señales de voz fueron tomadas de la base de datos de Berlín, la cual se describe en la siguiente sección.

#### A. Experimentos Tipo 1 (Parámetros Clásicos)

En el primer tipo de experimentos se usan los parámetros clásicos usados en el análisis de voz; se extraen 95 parámetros basados en la energía, la frecuencia fundamental, la duración de los silencios, las primeras 4 formantes y 13 MFCC's. Posteriormente a la selección de estos parámetros, el nuevo modelo incluye el uso de un método de selección de rasgos basados en el encadenamiento hacia adelante, con el fin de mejorar el desempeño del clasificador; como resultado de estos procesos, 14 parámetros fueron seleccionados usando el software WEKA [6]. Finalmente estos datos fueron usados como patrones de entrada para las Memorias Asociativas Alfa-Beta SVM.

Los 95 parámetros que se calcularon para la experimentación de tipo 1, se pueden agrupar en 8 cúmulos diferentes:

- Energía: máximo, promedio, mediana, moda, y desviación estándar.
- Amplitudes de Energía: máximo, mínimo, promedio, mediana, moda, y desviación estándar.
- Duraciones de silencios: máximo, mínimo, promedio, mediana, moda, y desviación estándar.
- Frecuencia fundamental: máximo, mínimo, promedio, mediana, moda, y desviación estándar.
- Duración de frecuencia fundamental: máximo, mínimo, promedio, mediana, moda, y desviación estándar.
- Sonoridad: máximo, mínimo, promedio, mediana, moda, y desviación estándar.
- MFCC's: Del archivo de audio, se extrae una matriz con los coeficientes MFCC's; los renglones representan 13 coeficientes MFCC's y las columnas representan el número de ventanas que tiene el archivo. Seis vectores son calculados de esta matriz; el primero contiene los valores máximos de la matriz, mientras que el segundo contienen los mínimos; En el tercero están los promedios y en el cuarto las medianas; el quinto contiene las modas y el sexto las desviaciones estándar. Posteriormente, a fin de reducir la cantidad de estos vectores, se calcula el máximo, mínimo, promedio, mediana, moda y desviación estándar para representar cada uno de los grupos de MFCC's.
- De los primeros 4 formantes: máximo, mínimo, promedio, mediana, moda, y desviación estándar.

Para el método de selección de rasgos incluido en el nuevo modelo, fue seleccionado el de Encadenamiento hacia adelante, que empieza con un conjunto vacío, se busca luego el atributo que más aporte a la clasificación y se va añadiendo hasta que al añadir el siguiente parámetro haga que caiga el desempeño de la clasificación.

Usando este método en WEKA [6], 14 parámetros fueron extraídos y se enuncian en la TABLA III.

TABLA III

PARÁMETROS SELECCIONADOS POR ENCADENAMIENTO HACIA ADELANTE

ENERGÍA PROMEDIO
AMPLITUD MÁXIMA DE LA ENERGÍA
MÍNIMO DE LA AMPLITUD DE LA ENERGÍA
AMPLITUD PROMEDIO DE LA ENERGÍA
DESVIACIÓN ESTÁNDAR DE LA AMPLITUD DE LA ENERGÍA
MODA DE LAS DURACIONES DE LA FRECUENCIA FUNDAMENTAL
PROMEDIO DE LA SONORIDAD
MFCC MÁXIMO DE LOS PROMEDIOS
MFCC MÍNIMO DE LOS MÍNIMOS
MFCC MÍNIMO DE LOS PROMEDIOS
MFCC MÍNIMO DE LAS DESVIACIONES ESTÁNDAR
MFCC PROMEDIO DE LAS DESVIACIONES ESTÁNDAR
MFCC MEDIANA DE LOS PROMEDIOS
DESVIACIÓN ESTÁNDAR DE LA SEGUNDA FORMANTE

### B. Experimentos Tipo 2

#### (Representaciones Bidimensionales)

El segundo tipo de experimentos incluye el desarrollo de otro enfoque que es original de este trabajo. Como todos los parámetros extraídos en el primer experimento son medidas estadísticas de la energía, frecuencia fundamental, duración de silencios, formantes, 13 MFCC's, tales como el promedio, moda, máximo, mínimo, desviación estándar y mediana, prácticamente son medidas de dispersión que representan en este caso uno o varios vectores de datos; sin embargo, en esta experimentación se observó que el parámetro que más información afectiva extrae de la voz, es la energía, por lo que se decidió trabajar con ese único parámetro: la energía.

Aquí surge la aportación principal del nuevo modelo asociativo de reconocimiento automático de voz emotiva: la representación bidimensional del parámetro energía, cuyo proceso se inicia con la extracción de los valores de energía que genera la envolvente (Figura 1).

Posteriormente, se rellena toda el área bajo la envolvente y se realiza una normalización en el eje de la amplitud para homogeneizar todas las instancias (Figuras 2 y 3).

Alfa-Beta SVM es una herramienta de clasificación efectiva para tareas de clasificación de datos binarios, por lo que se usó la representación bidimensional en forma de matriz como datos de entrada para el proceso de clasificación.



Fig. 1. Representación bidimensional del parámetro energía.



Fig. 2. Relleno de toda el área bajo la envolvente.



Fig. 3. Normalización en el eje de la amplitud.

## IV. RESULTADOS EXPERIMENTALES Y DISCUSIÓN

Trabajando en el campo de reconocimiento de voz emotiva, se debe elegir un corpus. Hay algunas bases de datos diseñadas para éste propósito [7], y la mayoría de las emociones comúnmente utilizadas en diferentes corpus de voz son: enojado, tristeza, felicidad, miedo, disgusto, alegría, sorpresa y aburrimiento.

La base de datos de Berlín [1] fue seleccionada por su disponibilidad, ya que la mayoría de las bases de datos reportadas en la literatura son de uso privado o requieren de una onerosa licencia de uso. La base de datos de Berlín contiene 535 instancias, etiquetadas como 127 en estado enojado, 81 de aburrido, 46 de disgustado, 69 de miedo, 71 de felicidad, 62 de tristeza y 79 de neutral. Este corpus fue grabado por 10 actores profesionales, 5 varones y 5 mujeres; contiene registros de 10 diferentes sentencias, con una frecuencia de muestreo de 16,000 datos por segundo en formato wav.

Dado que es la única base de datos disponible sin cargos, muchos investigadores la usan por lo que una de las consecuencias de este hecho es que los resultados de clasificación pueden ser fácilmente comparables y sus emociones son usualmente incluidas en otras bases de datos. La base de datos de Berlín está orientada al reconocimiento de voz emotiva actuada en idioma Alemán.

Al realizar los experimentos tipo 1, se sometieron los 14 parámetros que resultaron de la selección de rasgos de los 95 parámetros extraídos de la base de datos de Berlín, con los modelos que presenta WEKA [6]. Se muestran únicamente los modelos que produjeron los resultados más altos.

El primero de ellos es el Naive-Bayes (ver TABLA IV), que al clasificar toda la base de datos, recuperó correctamente 327 instancias (61.12%). El segundo modelo que se probó es el de Simple-Logistic, que alcanzó un precisión del 79.81%; es decir, recuperó satisfactoriamente 427 instancias.

El modelo que alcanzó el desempeño más alto de los incluidos en [6], es el Perceptrón Multicapa, que alcanzó un desempeño del (86.54%); es decir, 463 instancias correctamente clasificadas.

TABLA IV

DESEMPEÑO ALCANZADO POR LOS MODELOS EN WEKA [6] Y POR LAS MEMORIAS ASOCIATIVAS ALFA-BETA SVM

MODELO	INSTANCIAS CORRECTAS	INSTANCIAS CORRECTAS	DESEMPEÑO (%)
NAIVE BAYES	327	208	61.12
SIMPLE LOGISTIC	427	108	79.81
PERCEPTRON MULTICAPA	463	72	86.54
ALFA-BETA SVM	508	27	94.95

Es importante hacer notar que el nuevo modelo asociativo de reconocimiento automático de voz emotiva basado en las memorias asociativas Alfa-Beta SVM, el cual ha sido introducido en este artículo, recuperó 508 instancias

correctamente; es decir, alcanzó un desempeño del **94.95%**, con lo cual superó a todos los demás modelos.

Respecto de los experimentos de Tipo 2, al usar las imágenes de la energía normalizadas en el eje de la amplitud, con un relleno por debajo de la envolvente de la señal de energía junto con las máquinas asociativas Alfa-Beta SVM como modelo de clasificación, se clasificaron correctamente 506 instancias; es decir, se logró un desempeño del **94.5%**.

Este resultado está muy cercano al alcanzado por los experimentos de Tipo 1, con la aclaración relevante de que en este caso se usa **sólo un parámetro**: la energía.

## V. CONCLUSIONES

La selección de rasgos que se obtuvo de 14 parámetros ha demostrado ser buena. Esto se verifica cuando estos parámetros se usan como entrada en el nuevo modelo asociativo de reconocimiento automático de voz emotiva basado en las memorias asociativas Alfa-Beta SVM.

Al hacer lo anterior, el desempeño resulta casi del 95%, superando a los clasificadores de emociones conocidos, dado que el desempeño más alto reportado en la literatura, con esta misma base de datos, se encuentra alrededor del 80% [13, 14].

Es notable el alto desempeño alcanzado por el nuevo modelo asociativo de reconocimiento automático de voz emotiva basado en las memorias asociativas Alfa-Beta al usar como entrada un único parámetro, consistente en la representación bidimensional de la energía extraída de la señal de voz. El resultado es competitivo con el que se obtiene con 14 parámetros clásicos.

## AGRADECIMIENTOS

Los autores agradecen el apoyo de las siguientes instituciones para la realización de esta obra: Secretaría de Investigación y Posgrado, Secretaría Académica, COFAA y CIC del Instituto Politécnico Nacional, CONACyT y Sistema Nacional de Investigadores (SNI); específicamente, los proyectos SIP-20090807, SIP-20101709 y SIP-20110661.

## REFERENCIAS

- [1] *Berlin emotional speech database*, <http://www.expresive-speech.net/>
- [1] T. Vogt, E. André, and J. Wagner, *Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realization, Affect and Emotion in Human-Computer Interaction: From Theory to Applications*, Springer-Verlag, Berlin, Heidelberg, 2008.
- [2] J. Cichosz and K. Slot, "Emotion recognition in speech signal using emotion-extracting binary decision trees," in *Proceedings of Affective Computing and Intelligent Interaction*, Lisbon, Portugal, 2007.
- [3] T. Vogt, and E. André, "Improving automatic emotion recognition from speech via gender differentiation," in *Proceedings of Language Resources and Evaluation Conference*, 2006.
- [4] Z. Xiao, E. Dellandrea, W. Dou, and L. Chen, "Hierarchical classification of emotional speech," *IEEE Transactions on Multimedia*, 2007.
- [5] H. Ian, and F. Eibe, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, San Francisco, available online at <http://www.cs.waikato.ac.nz/ml/weka/>, 2005.
- [6] D. Ververidis and C. Kotropoulos, "A state of the art review on emotional speech databases," in *Proceedings of 1st Richmedia Conference*, 2003, pp. 109–119.

- [7] L. López-Leyva, C. Yáñez-Márquez, R. Flores-Carapia, and O. Camacho-Nieto, "Handwritten Digit Classification Based on Alpha-Beta Associative Model," in *Progress in Pattern Recognition, Image Analysis and Applications. LNCS 5197, Proc. 13th Iberoamerican Congress on Pattern Recognition CIARP 2008*, Havana, Cuba, 2008.
- [8] L. López-Leyva, C. Yáñez-Márquez, and I. López-Yáñez, "A new efficient model of support vector machines: ALFA-BETA SVM," in *23rd ISPE International Conference on CAD/CAM, Robotics and Factories of the Future*, Bogotá, Colombia, 2007.
- [9] C. Yáñez-Márquez, *Associative Memories Based on Order Relations and Binary Operators* (In Spanish). PhD Thesis. Center for Computing Research, Mexico, 2002.
- [10] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, Berlin Heidelberg New York, 1989.
- [11] M. H. Hassoun, *Associative Neural Memories*, Oxford University Press, New York, 1993.
- [12] T. Kohonen, "Correlation Matrix Memories," *IEEE Transactions on Computers*, 21(4), 353–359, 1972.
- [13] M. El Ayadi, M. Kamel, and F. Darray, "Survey on speech emotion recognition: Features, classification schemes and databases," *Pattern Recognition*, vol. 44, March, (2011) 572-587.
- [14] S. Zhang, L. Li, and Z. Zhao, "Spoken emotion recognition using kernel discriminant locally linear embedding," *Electronics Letters*, vol 46, 1344–1346, 2010.