

# Clasificación de Llanto del Bebé Utilizando una Red Neural de Gradiente Conjugado Escalado

J. Orozco García, Carlos A. Reyes García  
Instituto Nacional de Astrofísica Óptica y Electrónica  
Luis Enrique Erro # 1,  
Tonantzintla, Puebla, México,  
jose@cseg.inaoep.mx, kargaxxi@inaoep.mx

**Abstract.** El llanto es el único medio que un bebé tiene para comunicarse con el exterior. De acuerdo a los especialistas, en el llanto se refleja también el estado físico, patológico y/o anímico del bebé. El desarrollo de modelos dirigidos al estudio automático del llanto, permitirá proporcionar mejor atención al bebé para su cuidado. Este trabajo presenta el desarrollo de un sistema de reconocimiento automático del llanto del bebé, con ese objetivo en mente. Las características acústicas utilizadas se obtienen por medio de la técnica de predicción lineal y redes neuronales SCGB para clasificar entre los tipos de llantos de hambre, de dolor, de alguna patología y de placer. Se presentan los resultados iniciales obtenidos, con un número limitado de muestras, los cuales son muy promisorios.

**Palabras Clave:** Redes neuronales artificiales, SCGB, Coeficientes de predicción lineal. Llanto del Bebé.

## 1 Introducción

El llanto del bebé es un medio de comunicación, aunque más limitado, es similar al habla de personas adultas. Tanto los padres, como especialistas en el sector salud aprenden a distinguir entre los diferentes tipos de llantos de los bebés, utilizando la percepción individual para la diferenciación auditiva y la interpretación de los diferentes tipos de llanto. Tal diferenciación e interpretación es totalmente subjetiva, y su único soporte proviene del entrenamiento y expertez de cada persona. De acuerdo a los especialistas, la onda de llanto del bebé es portadora de información útil, tanto para determinar el estado físico y/o anímico del bebé, como para detectar posibles patologías físicas mayores, principalmente cerebrales, desde etapas muy tempranas. En trabajos anteriores sobre el análisis acústico del llanto del bebé se ha mostrado que existen diferencias significativas entre los diversos tipos de llanto, como llantos de niños sanos, llantos de dolor y patológicos entre otros, utilizando metodologías de clasificación basadas en el uso de mapas auto-organizados de Kohonen[3], redes neuronales[4] y análisis espectrográfico[5].

Los esfuerzos dirigidos a desarrollar estudios más formales y precisos, que permitan una interpretación más robusta del mensaje contenido en el llanto del bebé, son prácticamente nulos. Aun no existe una técnica de investigación concreta y efectiva sobre el análisis del llanto del bebé con propósitos clínicos y de diagnóstico. El método presentado en este artículo usa una red neuronal artificial. En esta primera etapa de prueba, coeficientes de predicción lineal (Lineal Predictive Coding, LPC) sirven como entrada a una red neuronal dada, variando los parámetros derivados de la señal de llanto del bebé, los cuales son ajustados para una clasificación correcta de diferentes tipos de llantos.

## **2 Proceso del Reconocimiento Automático del Llanto del Bebé**

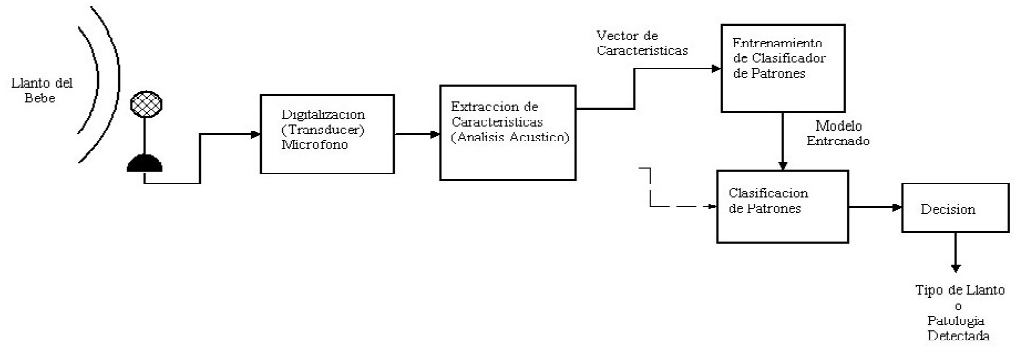
El proceso del Reconocimiento Automático del Llanto del Bebé (fig. 1) es básicamente un problema de procesamiento de patrones, similar al Reconocimiento Automático del Habla (RAH). La meta es tomar la onda de llanto del bebé como el patrón de entrada, y al final obtener el tipo de llanto o patología detectada en el bebé. Generalmente, el proceso de Reconocimiento Automático del Habla se hace en dos pasos. El primer paso es conocido como procesamiento de la señal, o extracción de características, mientras que el segundo se conoce como procesamiento o clasificación de patrones.

En el análisis acústico, la señal de llanto del bebé es analizada para extraer las características más importantes en función del tiempo. Se efectúa el depuramiento de la señal tratando de eliminar la información irrelevante e indeseable como el ruido de fondo, distorsión del canal, y características particulares de la señal. Aunque los datos son reducidos al remover componentes repetitivos, la información relevante para la clasificación de patrones es conservada de una manera óptima. Algunas de las técnicas simples más usuales para el procesamiento de las señales son: coeficientes de predicción lineal, coeficientes cepstrales, timbre, intensidad, análisis espectral, y bancos holmes entre otros. El conjunto de características obtenidas puede ser representado como un vector, y cada vector puede ser tomado como un patrón. El vector de características es comparado con el conocimiento que tiene la computadora. Por el lado de los métodos de reconocimiento de patrones, se han utilizado tradicionalmente cuatro enfoques principales: comparación de patrones, modelos estadísticos, sistemas basados en conocimientos y modelos conexionistas.

## **3 Coeficientes de Predicción Lineal**

El análisis acústicos implica la selección y aplicación de técnicas de filtrado, extracción de características, segmentación de la señal, normalización y compresión de datos. Con la aplicación de estas técnicas se intenta describir la señal en términos de sus componentes fundamentales.

El análisis de Predicción Lineal ha sido una de las técnicas en el dominio del tiempo más utilizadas en los últimos años. Esta técnica intenta predecir una



**Fig. 1.** Proceso de Reconocimiento Automático del Llanto del Bebé

muestra de señal a través de una combinación lineal de varias muestras de señal previas. Los parámetros obtenidos muestran un espectro suavizado que proporciona información mas representante de la señal. Este método de análisis de LPC, es una técnica particularmente apropiada para analizar la señal hablada, por lo cual es considerada para obtener buenos resultados en al análisis del llanto del bebé. La forma particular en la cual los datos son segmentados determina si el método de covarianza, el método de autocorrelación, o cualquiera de otros métodos, del análisis de predicción lineal es usado. El primer método que estamos usando es predicción lineal por autocorrelación, donde la función de autocorrelación, para la señal de longitud finita  $s(n)$  se define por[1]:

$$r_l = \sum_{n=0}^{N-1-l} s(n)s(n+l) \quad (l \geq 0) \quad (1)$$

donde,  $l$  es el retardo que existe entre la señal y su versión retardada,  $N$  es el número total de muestras. La función de error de predicción entre la muestra de datos actual  $s(n)$  y la muestra predecida  $\hat{s}(n)$  está definida de la forma [1]:

$$e(n) = \sum_{i=0}^M a_i s(n-i) \quad (a_0 = 1) \\ \text{para } n = 0, 1, \dots, N + M - 1 \quad (2)$$

Cuando el orden del modelo de predicción lineal incrementa, mas detalles de energía del espectro de la señal puede ser aproximada.

## 4 Redes Neuronales

Las redes neuronales son una de las metodologías más utilizadas para clasificación y reconocimiento de patrones. Recientemente su utilización en problemas de reconocimiento de habla ha ido cada vez en aumento. Entre los modelos más utilizados de redes neuronales se encuentran las redes con alimentación hacia adelante (feed-forward) y que utilizan alguna variación del método de entrenamiento de retropropagación (back-propagation). En general, una red neuronal está compuesta por un conjunto de nodos y un conjunto de ligas. Los nodos corresponden a las neuronas y las ligas representan las conexiones y el flujo de datos entre neuronas. Las conexiones son cuantificadas por medio de pesos, que son dinámicamente ajustados durante el entrenamiento. El entrenamiento requerido puede ser hecho a través de la técnica conocida como Retro-propagación (back-propagation). En el entrenamiento, se presentan a la red el patrón de entrada y el patrón de salida deseado, y se ajustan los pesos de la red para minimizar la diferencia entre la salida real y el resultado deseado. El método de aprendizaje de propagación del error hacia atrás utiliza la siguiente técnica[7]: dado un patrón de entrada, compara el resultado obtenido en las unidades de salida con la respuesta que se desea obtener. A continuación reajusta los pesos de la red de manera que, la siguiente vez que se presente el mismo patrón de entrada, la red produzca un resultado más cercano al deseado, es decir, que el error disminuya.

## 5 Red Neural de Gradiente Conjugado Escalado

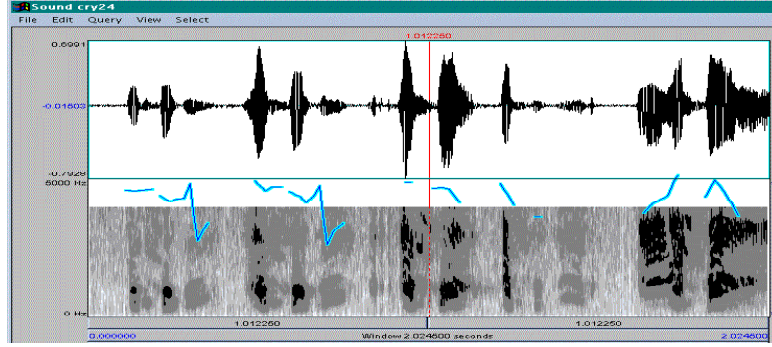
Desde un punto de vista de optimización de aprendizaje en una red neural es equivalente a minimizar la función de error global. Muchos de los algoritmos están basados en el algoritmo de gradiente descendente para optimización. Usualmente, estos algoritmos tienen un índice de convergencia pobre y dependen de parámetros los cuales tienen que ser especificados por el usuario. Los valores de estos parámetros son cruciales para el éxito del algoritmo. Un ejemplo es el algoritmo backpropagation standard el cual se comporta mal en problemas a gran escala, dependiendo de los parámetros del índice de aprendizaje dados por el usuario.

La minimización es un proceso local iterativo en el cual una aproximación a la función en una vecindad de un punto actual en el espacio de peso es minimizada. Los Métodos de Gradiente Conjugados están basados en la estrategia general de optimización.

El algoritmo SCG[9] denota una aproximación cuadrática para el error  $E$  en una vecindad de un punto  $w$  por[9]:

$$E_{\text{qw}}(y) = E(w) + E'(w)^T y + \frac{1}{2} y^T E''(w) y \quad (3)$$

en orden para determinar el mínimo para  $E_{\text{qw}}(y)$  deben encontrarse los puntos críticos para  $E_{\text{qw}}(y)$  los cuales son la solución para el sistema lineal definido por[9]:



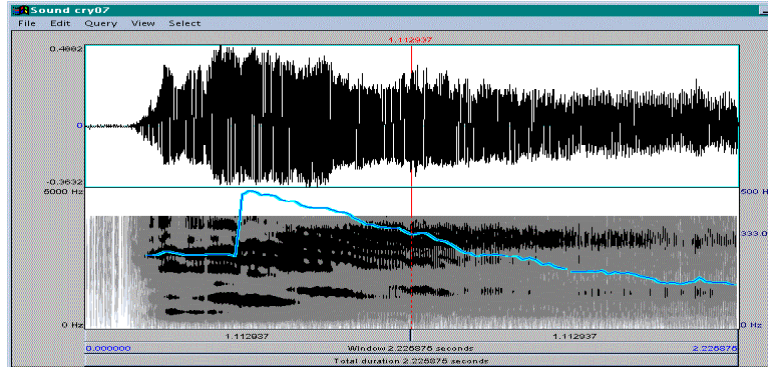
**Fig. 2.** Llanto de hambre

$$E'_{qw}(y) = E''(w)y + E'(w) = 0 \quad (4)$$

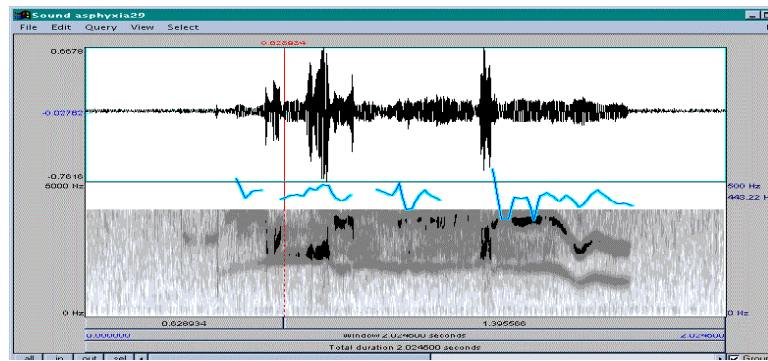
Una variación del Método de Gradiente Conjugado es el Método de Gradiente Conjugado Escalado (SCGM). El algoritmo SCG muestra una convergencia superlineal en la mayoría de los problemas. Usando un mecanismo para decidir que tan lejos ir en una dirección especificada SCG evita un tiempo de consumo en la búsqueda lineal por iteración de aprendizaje, el cual hace el algoritmo mas rápido que otros algoritmos de segundo orden.

## 6 Conjunto de Datos

Se tomaron cuatro tipos de llantos: llantos de dolor, de hambre, patológicos y de placer. 20 registros de llantos etiquetados pertenecientes a 4 categorías, se dividieron en señales de un segundo de duración con información relevante para poder analizarlas. 42 muestras ajustadas fueron obtenidas para el análisis de las cuales 8 pertenecen al llanto de hambre, 16 al de dolor, 11 patológicos y 7 al de placer. El llanto del bebé presenta diferencias significativas entre los diversos tipos de llanto. Por ejemplo el llanto del bebé por hambre presenta características como pequeños hipos, patrón de tonos creciente-decreciente rápidos, plosivos glotales, fonaciones débiles como se muestra en la figura 2. Los llantos de dolor (fig. 3) presentan características similares como: tensión más grande de la normal, tono más grande que el normal, patrón de tono decreciente y generalmente largos (2.5 sec), cambios rápidos de tono, patrón interrumpido con cambios de tensión, y un patrón no estable. El llanto con alguna patología (fig. 4) es característico por: un tono extremadamente alto, inhabilidad para mantener un tono armónico, así como intensidad débil. Finalmente, el llanto de placer (fig. 5) presenta tono relativamente débil, falta de tensión, tono parejo, fonación armónica, intensidad relativamente más baja.



**Fig. 3.** Llanto de dolor

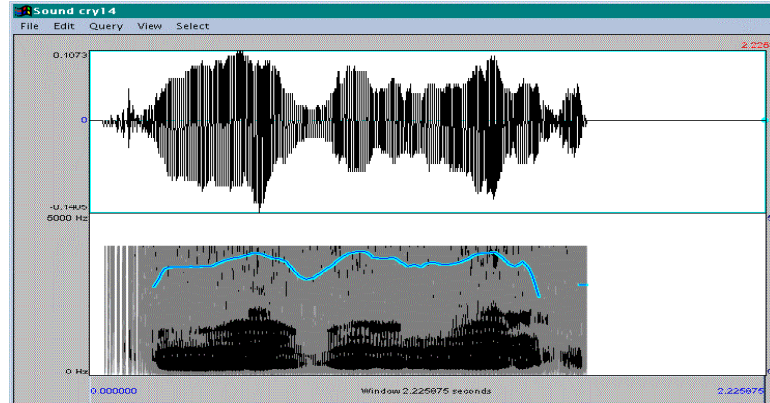


**Fig. 4.** Llanto ocasionado por asfixia

Para propósitos de este experimento, deseamos encontrar parámetros cambiando la longitud de ventana de análisis, así como el número de coeficientes, donde obtengamos mejores resultados de clasificación. La arquitectura de red entrenada y probada, tenía 4 salidas, para cada una de las clases del conjunto de datos. Es muy difícil saber el número óptimo de neuronas y capas que la red debe tener.

## 7 Representación Paramétrica

El análisis LPC ha sido una de las técnicas de análisis acústico que ha dado buenos resultados para el reconocimiento del habla. Siguiendo este criterio el análisis LPC se ha tomado como base para aplicarlo al análisis acústico de llanto. En la onda del llanto del bebé no tenemos una unidad básica para el análisis, como los formantes en el habla, por lo que enfocamos el análisis y extracción de



**Fig. 5.** Llanto de placer

características a patrones de periodo largo. Por cada muestra de llanto de un segundo de duración fue segmentada la señal, en ventanas de 100 ms, y 50 ms. Se extrajeron 16 coeficientes LPC por ventana, consecuentemente, los vectores de características contienen 160 y 320 parámetros, equivalente a 160 y 304 entradas a la red neuronal respectivamente.

## 8 Software Empleado

Praat 4.0.2 es el software utilizado para obtener los coeficientes LPC's, a partir de una señal de entrada. Praat es un sistema para hacer fonética por computadora, desarrollado por Paul Boersma y David Weenink en Summer Institute of Linguistics [2]. En orden de entrenar y probar, se tomaron los vectores de características a una red neuronal, implementada en Matlab 6.0.0.88 [6] desarrollado por *The MathWorks Inc.* 1984 - 2000. Usado para simular la red Backpropagation de gradiente escalada conjugada.

## 9 Procedimientos de Entrenamiento de Red y Experimentación

De las 42 muestras de llanto existentes con 1 segundo de duración y a 4000Hz, 8 muestras de llanto pertenecen al tipo de Hambre, 16 al tipo de Dolor, 11 clasificadas como llanto con alguna patología (síndrome de down, cri du chat, asfixia) y 7 expresiones cuando el bebé está contento. Varios conjuntos de datos fueron usados para entrenar, usando la estrategia *cross validation* [8]. Esto involucra dividir aleatoriamente los datos en un conjunto de datos de entrenamiento y un conjunto de datos de prueba. Las muestras se dividieron en 4 conjuntos (Tabla 1), incluyendo muestras de las cuatro categorías. Dos conjuntos con 10 muestras

y dos conjuntos con 11 muestras cada uno, y usando tres de estos conjuntos para entrenamiento de la red y el conjunto restante es usado como conjunto de prueba. Este proceso es repetido hasta que todos los conjuntos han sido probados como conjunto de prueba una vez.

Categoría	ID	GPO 1	GPO 2	GPO 3	GPO 4	Total
Hambre	1	2	2	1	3	8
Dolor	2	4	5	3	4	16
Patología	3	2	2	4	3	11
Placer	4	2	1	3	1	7
Total		10	10	11	11	42

Tabla 1. Conjuntos de muestras para entrenamiento y clasificación

Por el momento se utilizó la red de Retropropagacion de Gradiente Conjugado Escalado (Scaled Conjugated Gradient Backpropagation-SCGB), por ser una de las más simples y rápidas, al momento de entrenarse, generando resultados aceptables. El número de nodos de entrada de la red es 160, 5 neuronas en la capa oculta y 4 neuronas en la salida de la red, correspondientes a las categorías de llantos. Para la experimentación con 16 coeficientes, y una longitud de las ventanas utilizadas de 50 ms, la red contenía 304 nodos de entrada, 20 nodos en la capa oculta y 4 en la capa de salida de la red. El entrenamiento se detiene cuando el número máximo de 600 épocas es alcanzado, cuando la cantidad máxima de tiempo ha sido excedida, o cuando el desempeño ha sido minimizado para el objetivo.

## 10 Resultados Experimentales

La red neuronal fue usada para entrenar y clasificar el llanto de dolor, hambre, patología y placer. Los conjuntos de datos que sirven de entrada a la red, son obtenidos por la técnica LPC de la onda del llanto del bebé. Para el primer conjunto de parámetros, la red neuronal Backpropagation de gradiente conjugada escalada utilizó 160 nodos de entrada correspondiente 16 coeficientes LPC por 10 ventanas de 100 ms cada una, en la señal de duración de 1 segundo. La capa oculta consistía de 5 nodos y la capa de salida de 4 nodos correspondientes a las categorías de los tipos de llanto del infante. Para obtener mejores resultados se varió la longitud de las ventanas y el número de coeficientes de predicción lineal. Los mejores resultados obtenidos fueron dados por los parámetros de 16 coeficientes por cada una de las 19 ventanas de 50 ms, resultando 304 nodos para la capa de entrada y 20 nodos en la capa oculta. La clasificación fue calculada tomando el número de clases de salida clasificadas correctamente, dividido por el número total de muestras de prueba. Los resultados de clasificación (Tabla 2), se muestran para cada categoría de llanto, usando conjuntos de parámetros diferentes. Los valores bajo las columnas representan la exactitud en la clasificación de acuerdo al número de coeficientes LPC, en cada categoría de llanto, localizado en las filas.



	16 LPC, 100 ms	16 LPC, 50 ms
Hambre	50 %	100 %
Dolor	100 %	66.66 %
Patología	100 %	75 %
Placer	100 %	100 %
Clasificación	87.5 %	85.41 %

Tabla 2. Clasificación de tipos de llantos utilizando una red SCG

## 11 Conclusión

Una variación del Método de Gradiente Conjugado es el Método de Gradiente Conjugado Escalado (SCGM), el cual evita el tiempo de consumo en la búsqueda lineal por iteración de aprendizaje, haciéndolo más rápido con respecto a otros métodos CG. Este trabajo ha presentado los resultados obtenidos, con la técnica LPC, usando una arquitectura de red neuronal. Fue mejor la clasificación usando 16 coeficientes por ventana con duración de 50 ms. Aún cuando el número de muestras para el entrenamiento de la red neuronal son muy pocas y las categorías que se pretenden clasificar son varias, se han obtenido resultados aceptables. Se pretende variar aún el número de coeficientes, así como la longitud de las ventanas, para obtener mejores resultados.

## 12 Reconocimiento

Este trabajo es parte de un proyecto que esta siendo financiado por CONACYT (numero 37914-A).

## References

1. John D., Gray, Augustine H., *Linear prediction of speech*. New York: Springer-Verlag, 1976. pp 10-16.
2. Paul Boersma, David Weenink, Praat v. 4.0.2. *Sistema para hacer fonética por computadora. 1992 - 1998* Summer Institute of Linguistics.
3. Sergio D. Cano, Daniel I. Escobedo y Eddy Coello, *El Uso de los Mapas Auto-Organizados de Kohonen en la Clasificación de Unidades de Llanto Infantil*, Grupo de Procesamiento de Voz, Universidad de Oriente, Cuba.
4. Marco Petroni, Alfred S. Malowany, C. Celeste Johnston, Bonnie J. Stevens. *Identification of pain from infant cry vocalizations using artificial neural networks (ANNs)*, pp.729-738. The International Infant Cry Research Group. Applications and Science of Artificial Neural Networks. The International Society for Optical Engineering. Volume 2492. Part two of two. Paper #: 2492-79. 1995.
5. O. Wasz-Hockert, J. Lind, V. Vuorenkoski, T. Partanen y E. Valanne, *El Llanto en el Lactante y su Significación Diagnóstica*, Ed. Científico-Médica, Barcelona, 1970.
6. Manual "Signal Processing Toolbox", Matlab V.6.0.0.88, release 12. Desarrollada por The MathWorks Inc. 1984 - 2000.
7. Bernal, J., Bobadilla, J. y Gomez, P., *Reconocimiento de Voz y Fonética Acústica*, Ed. Alfaomega ra-ma, México, 2000. pp 284 - 288.

8. Haykin, Simon S. *Neural networks: A comprehensive foundation*. New York: Macmillan College Publishing Company, Inc., 1994. pp 179 - 181
9. Moller, *A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning*, Neural Networks 6 (4), 1993, 525-533