
UNIVERSIDAD NACIONAL DEL LITORAL
Facultad de Ingeniería y Ciencias Hídricas

# Modelización de secuencias para el reconocimiento automático de patrones

**César Ernesto Martínez**

Tesis remitida al Comité Académico del Doctorado
como parte de los requisitos para la obtención
del grado de
DOCTOR EN INGENIERIA
Mención Inteligencia Computacional, Señales y Sistemas
de la
UNIVERSIDAD NACIONAL DEL LITORAL

**2011**

Comisión de Posgrado, Facultad de Ingeniería y Ciencias Hídricas, Ciudad Universitaria,
Paraje "El Pozo", S3000, Santa Fe, Argentina.

**Doctorado en Ingeniería**
**Mención Inteligencia Computacional, Señales y Sistemas**

Título de la obra:

# Modelización de secuencias para el reconocimiento automático de patrones

Autor: César Ernesto Martínez

Director: Dr. Hugo Leonardo Rufiner

Codirector: Dr. Diego Humberto Milone

Lugar: Santa Fe, Argentina

Palabras Claves:

Reconocimiento de patrones
Clasificación de cromosomas
Cariotipado automático
Representación cortical del habla
Reconocimiento robusto del habla

*A Teresa, quien realizó un esfuerzo muy superior al mío para que este trabajo vea la luz.*

*A Franco y Aldana.*

# Agradecimientos

A la hora de escribir estas líneas me doy cuenta de la cantidad de gente que puso de sí para que yo hoy pudiera terminar este trabajo.

Quiero agradecer, en primer lugar, a mis directores Leo y Diego, quienes tomaron este trabajo como propio y estuvieron siempre dispuestos a ayudarme y brindarme su tiempo cuando lo necesité. Aprendí y sigo aprendiendo mucho de ellos.

Agradezco especialmente a Marcelo y Leandro, con quienes comparto la oficina en el sinc(i), un poco de trabajo y muchos momentos de amistad. A Leandro DP (Dr. H), por las discusiones y su aporte A todos mis compañeros del sinc(i) y del ARSiPRe: DiegoT, Matías, Fede, Carlos, Analía y María Eugenia.

A los investigadores de la Universidad Politécnica de Valencia que me dirigieron en esa institución durante la parte inicial de este trabajo, Francisco Casacuberta Nolla y Alfons Juan Císcar. También quiero agradecer a los compañeros de laboratorio que tuve durante mis estancias, en especial a Héctor, Adrià, Raúl, Piedachu y Ramón.

A mi familia, por todo el apoyo que me brindaron en este tiempo.

Agradezco al Dr. Gunter Ritter (*Fakultät für Mathematik und Informatik, Universität Passau, Passau, Germany*), el acceso al corpus de imágenes Cpa utilizado en este trabajo.

Se utilizaron con permiso imágenes obtenidas del *Department of Patholo-*

*gy, University of Washington*, copyright 2004 by David Adler [1].

Este trabajo ha sido subvencionado parcialmente por las siguientes instituciones y subsidios:

- UNER, bajo proyectos PID 6036 y PID 6062.
- BID-UNER, bajo contrato FOMEC.
- SECyT-UNER, bajo proyectos PICT 11-12/700A y PAV 2003 127-1.
- UNL, bajo proyecto PROMAC-POS 2004.
- UNL, bajo proyecto CAID 12/G407.

*A todos, mi más sincera gratitud,*

César Ernesto Martínez
Santa Fe, Marzo de 2011

[1] Página web al 19/03/08: http://www.pathology.washington.edu/galleries/Cytogallery

Modelización de secuencias
para el reconocimiento automático de patrones

César Ernesto Martínez

Director de tesis: Dr. Hugo L. Rufiner
Co-Director de tesis: Dr. Diego H. Milone
Departamento de Informática, 2011

## Resumen

La modelización de secuencias es un problema de gran interés en el campo del reconocimiento de patrones. En el mismo se busca diseñar y construir sistemas especializados en capturar las particularidades de tramos distintivos de las secuencias y su estructura de repetición. Estos sistemas pueden ser empleados tanto para la clasificación (modelos discriminativos) como para la síntesis (modelos generativos). El rango de las aplicaciones de interés es muy amplio, incluyendo el reconocimiento automático del habla, el análisis de secuencias en bioinformática, el reconocimiento óptico de caracteres, etc. En este trabajo se avanza sobre la extracción de características y modelización de secuencias en dos dominios de aplicación: las clasificación de imágenes de cromosomas y el reconocimiento robusto de señales de habla.

En citología, una tarea laboriosa para el experto es la clasificación de cromosomas. Se proponen aquí nuevas parametrizaciones que explotan la variabilidad de las bandas de grises a lo largo de los cromosomas. Se introducen, además, nuevas formas de clasificar estos patrones basadas en redes neuronales recurrentes y modelos ocultos de Markov continuos. En la rutina clínica, el experto clasifica los cromosomas de cada célula en su conjunto. A fin de emular esta forma de trabajo, se propone un algoritmo de post-clasificación contextual. El mismo toma los resultados obtenidos previamente con cualquier clasificador de manera aislada y efectúa una re-localización de cromosomas en cada clase, mejorando el desempeño de acuerdo al número esperado de cromosomas por grupo.

En la representación de la señal de la voz, recientemente se ha acrecentado el interés en el desarrollo de los denominados *sistemas bioinspirados*. En este trabajo se proponen nuevas alternativas de parametrización que modelan las características obtenidas experimentalmente a nivel de corteza auditiva primaria. Los nuevos esquemas de extracción de características se basan en la representación rala del espectrograma auditivo cortical temprano. Los patrones generados se aplican a las tareas de clasificación robusta de fonemas y limpieza de señales de habla inmersas en ruido. Los resultados obtenidos muestran que estas técnicas logran extraer las pistas útiles para el reconocimiento y/o recuperar la información que preserva objetivamente la calidad de las señales limpiadas, demostrando esta aproximación ventajas en desempeño respecto a otros métodos reportados en la literatura.

Modelización de secuencias
para el reconocimiento automático de patrones

César Ernesto Martínez

Director de tesis: Dr. Hugo L. Rufiner
Co-Director de tesis: Dr. Diego H. Milone
Departamento de Informática, 2011

## Abstract

Sequence modeling is a problem of interest in the field of pattern recognition. It is aimed at the design and building of specialized systems that capture the particularities of distinctive segments of the sequences and their repetition structure. These systems could be used for both classification (discriminative models) as well as for synthesis (generative models). The range of applications is very wide, including automatic speech recognition, sequence analysis in bioinformatics, optical character recognition, etc. In this work we present progresses on the feature extraction and sequence modeling in two domains of applications: classification of chromosome images and robust recognition of speech signals.

In cytology, a laborious task for the expert is the classification of chromosomes. New parameterizations are here proposed to exploit the variability of the gray bands along the chromosomes. Also, new ways to classify these patterns based on recurrent neural networks and continuous hidden Markov models are introduced. In the clinical routine, the expert classifies the chromosomes of each cell as a whole. In order to emulate this methodology, a contextual classification algorithm is proposed. It takes the results obtained previously with any classifier in isolated way and carries out a relocation of chromosomes in each class, improving the performance according to the expected number of chromosomes per group.

For the speech signal representation, the interest in the development of so-called *bio-inspired systems* was recently revived. In this thesis we propose new alternatives to the speech parameterization that model the characteristics obtained experimentally at primary auditory cortex level. The new feature extraction schemes are based on the sparse representation of early cortical auditory spectrogram. The patterns obtained are applied to two tasks: robust phoneme classification and speech denoising. The obtained results show that these techniques can extract useful clues for recognition and retrieval of information that objetively preserve the quality of the denoised signals, with performance benefits over other methods previously reported.

# Prefacio

Esta tesis plantea la idea general de tratar a los objetos de manera secuencial, a fin de extraer la información relevante que será luego incorporada a los modelos para el reconocimiento. Este tratamiento es independiente de la naturaleza de los objetos a clasificar, siendo las ideas expuestas luego aplicables a diferentes ejemplos en particular.

Los clasificadores se diseñan en base a redes neuronales y modelos probabilísticos. Las primeras son parte de los sistemas *conexionistas*: procesadores de información que consisten en unidades primitivas de procesamiento conectadas por enlaces, que distribuyen la información de manera similar al mecanismo biológico que se produce entre las neuronas del cerebro. Por otro lado, los *modelos ocultos de Markov* constituyen modelos de procesos estocásticos secuenciales, donde la dinámica de los mismos es contemplada en su arquitectura, aproximando la variabilidad intrínseca de los datos mediante estadísticas internas.

El rango de las aplicaciones de interés es muy amplio, incluyendo el reconocimiento automático del habla, el análisis de secuencias en bioinformática, el reconocimiento óptico de caracteres, etc. En cuanto a las aplicaciones exploradas en este trabajo, en el campo de las imágenes se presenta la clasificación de microfotografías de cromosomas, componente esencial de los sistemas computarizados de análisis clínico. En el campo del reconocimiento del habla, se presenta la clasificación de fonemas mediante una aproximación inspirada biológicamente, alternativa a las representaciones clásicas.

En este trabajo se pretende continuar y ampliar estas líneas de investigación mediante la propuesta de nuevas características que incorporen mayor

información, por ejemplo, descriptores locales a lo largo de las secuencias; que luego sean analizadas a manera de series temporales. En cuanto a los sistemas de clasificación, se proponen nuevas aplicaciones de redes neuronales parcialmente recurrentes y modelos probabilísticos continuos para el tratamiento de las secuencias generadas.

El objetivo general de esta tesis es desarrollar nuevos métodos integrales para la clasificación de secuencias, que mejoren diferentes aspectos de toda la cadena de reconocimiento respecto a los sistemas actuales: preproceso, extracción de características y modelización. Los objetivos específicos son los siguientes:

- Diseñar una metodología de preprocesamiento y extracción de características de las secuencias de interés, que obtenga una representación adecuada de los objetos bajo análisis: cromosomas presentes en imágenes de microscopía óptica y fonemas extraídos de la señal acústica de habla.

- Proponer e implementar sistemas de clasificación automática de imágenes de cromosomas basados en técnicas del reconocimiento de patrones previamente no exploradas en la tarea, como los modelos ocultos de Markov continuos y las redes neuronales parcialmente recurrentes.

- Desarrollar nuevos algoritmos de clasificación de fonemas basado en una representación alternativa de la señal de habla obtenida mediante un modelo auditivo. Proponer y evaluar métodos de limpieza de ruido en señales de habla y aplicarlos en el esquema previamente mencionado.

- Evaluar comparativamente las capacidades y desempeño de diferentes técnicas propuestas en el área del reconocimiento de patrones, en el contexto de las áreas de aplicación propuestas.

La tesis se encuentra organizada como se explica a continuación.

El Capítulo 1 introduce al lector en los conceptos teóricos sobre las aproximaciones del Reconocimiento de Formas en los cuales se sustenta el desarrollo de este trabajo: las redes neuronales tipo perceptrón multicapa, las redes de Elman y los modelos ocultos de Markov continuos.

Los aportes en los campos de aplicación se ordenan en dos partes, la primera de las cuales está dedicada a los sistemas de reconocimiento de imágenes de cromosomas. El Capítulo 2 presenta la metodología aplicada en la etapa de preprocesamiento, empleada para normalizar las imágenes y

solucionar pequeños defectos de la adquisición y segmentación de cromosomas. Posteriormente se presentan los métodos de extracción de características, tanto los propuestos en la literatura (también implementados aquí para comparación) como nuevas alternativas que buscan capturar de mejor manera la variabilidad de grises en las bandas de los cromosomas. Se expone, además, la formulación de un algoritmo propuesto para realizar la clasificación restringida al contexto celular. El Capítulo 3 presenta el desarrollo experimental y el análisis de resultados obtenidos mediante la aproximación de modelos conexionistas y modelos ocultos de Markov.

La segunda parte expone los métodos planteados en el campo del reconocimiento del habla. El Capítulo 4 presenta la idea general y variantes de la aproximación propuesta para la representación de la señal, basada en el procesamiento que se obtiene al modelar los patrones de disparo a nivel de la corteza auditiva primaria. A fin de demostrar la viabilidad de la técnica se muestran experimentos con señales artificiales. Por otro lado, se avanza también sobre un aspecto previamente no reportado: la limpieza de ruido sobre las representaciones corticales, en la búsqueda de agregar robustez en presencia de ruido aditivo. El Capítulo 5 presenta el marco experimental y los resultados obtenidos en la aplicación de la técnica propuesta a la tarea de clasificación automática de fonemas.

Finalmente, el Capítulo 6 presenta las conclusiones derivadas de todo el trabajo, algunas líneas de investigación y desarrollo que podrían continuar lo aquí expuesto y la lista de publicaciones resultantes.

El trabajo se enmarca dentro de las actividades de investigación que lleva adelante el Centro de I+D en Señales, Sistemas e Inteligencia Computacional, sinc(i), de la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral. La parte inicial de este trabajo se llevó a cabo en una estancia de investigación en el grupo de *Reconocimiento de Formas y Tecnología del Lenguaje* asociado al Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, España.

# Índice general

# Índice de figuras

# Índice de tablas

CAPÍTULO 1

# Introducción al reconocimiento de patrones

En este capítulo se presentan los conceptos fundamentales y formulaciones matemáticas de las diferentes aproximaciones del reconocimiento de formas estudiadas en esta tesis.

En primer lugar se realiza el desarrollo teórico referente a los modelos conexionistas, así denominados por la interconección física de unidades mínimas de procesamiento, junto al planteo de algoritmos para el manejo de la información entre ellas y que, en su conjunto, pueden lograr buenos desempeños en tareas muy complejas.

A continuación se expone el correspondiente tratamiento de una aproximación estructural del reconocimiento de formas, los modelos ocultos de Markov. En particular se trata la formulación de los modelos continuos, que no fueran previamente explorados para la tarea de clasificación de cromosomas.

El capítulo se completa con una revisión de los métodos de estimación del error en clasificación aplicados, los cuales sirven en la etapa de experimentación para evaluar y comparar el desempeño de los clasificadores construidos.

## 1.1 Generalidades

En la naturaleza, gran cantidad de la información que procesan nuestros sentidos para captar el medio que nos rodea se encuentra presente como muestras de los objetos que percibimos. En este contexto, podemos definir una *forma* (o *patrón*)[1] como un objeto de interés que es identificable del resto (fondo en una imagen, ruido ambiente en una grabación de audio, etc.). Muchas veces, estas formas pueden ser difusas o no estar bien definidas, lo cual hace más compleja la identificación. En otros contextos, los patrones pueden no necesariamente ser objetos visibles o tangibles: en ciencias económicas, por ejemplo, se habla del Producto Bruto Interno[2] como un patrón de análisis de la capacidad productiva de una economía nacional [1].

El Reconocimiento de Formas (RF), también conocido como Reconocimiento de Patrones, constituye una rama del conocimiento de la Inteligencia Artificial que intenta modelar los procesos de percepción y razonamiento humanos para crear sistemas informáticos que los imiten. En particular, interesa cómo somos capaces de distinguir y aislar los patrones en un ambiente particular, los reunimos en grupos de acuerdo a características comunes y les asignamos un nombre identificatorio a cada grupo [2]. Esta disciplina estuvo en constante evolución en los últimos 50 años, continuando su auge con el avance tecnológico actual, que hace posible construir sistemas cada vez más potentes, tanto en sus capacidades de procesamiento como de generalización del conocimiento adquirido [3].

Desde su concepción, numerosas aplicaciones exitosas del RF confirman su eficiencia y aceptación en la comunidad científica, y para ejemplificar la diversidad de tareas a las que fue aplicado se pueden nombrar: el reconocimiento del habla, permitiendo la traducción a texto escrito del lenguaje hablado [4, 5]; la inspección visual automática para detección de fallas en la fabricación de piezas industriales [6]; la identificación biométrica o verificación de identidad mediante rasgos físicos [7, 8]; la detección de patrones de fraude en cheques o uso de tarjetas de crédito [9, 10]; y muchas otras aplicaciones en astronomía, medicina, ingeniería ambiental, etc. [11]

[1] En adelante, se usarán estos términos indistintamente. El lector no debe confundir este término con la definición de *descripción*, *vector de características* o *feature* (en inglés), como después se introduce en el texto.

[2] Definido como el valor monetario total de la producción corriente de bienes y servicios de un país durante un período. A. Digier, *Economía para no economistas*. Valletta Ediciones S.R.L., 1999.

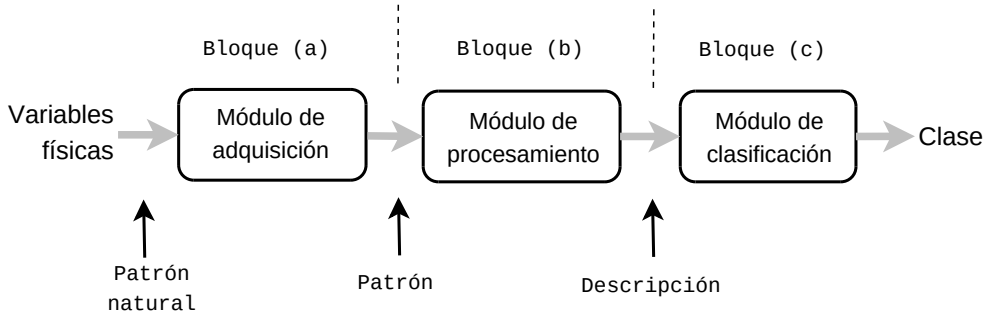



Figura 1.1: Diagrama en bloques conceptual de un sistema de reconocimiento de formas.

### 1.1.1 Paradigma de trabajo

En forma general, un sistema completo de captura y reconocimiento de patrones puede ser dividido para su estudio en tres grandes bloques, representados gráficamente en el diagrama de la Figura 1.1:

(a) **Adquisición de los datos analógicos**. Una muestra del mundo físico a analizar (*patrón natural*) es ingresada al sistema mediante un dispositivo de captura, compuesto básicamente por un transductor y filtros que transforman la variable física del patrón natural en una señal digital. A la salida de esta fase se tiene el patrón (o forma) listo para su tratamiento posterior.

(b) **Procesamiento de los datos digitales**. El patrón ingresado es acondicionado digitalmente para solucionar problemas de la transducción y luego es analizado para obtener una *descripción*: representación alternativa que conforma el conjunto de características.

(c) **Clasificación**. Finalmente se lleva a cabo el proceso de decisión sobre la *clase* o categoría a asignar al patrón analizado según el modelo o técnica de clasificación que sea implementada.

En adelante consideraremos a un sistema de RF compuesto por los bloques (b) y (c) anteriores y estudiaremos su funcionamiento, el cual puede ser dividido en dos fases bien diferenciadas: la fase de aprendizaje y la fase de reconocimiento, como se muestra en la Figura 1.2.

En el esquema mencionado es visible en líneas gruesas la secuencia de bloques del sistema de RF cuando ya se encuentra en pleno funcionamiento,

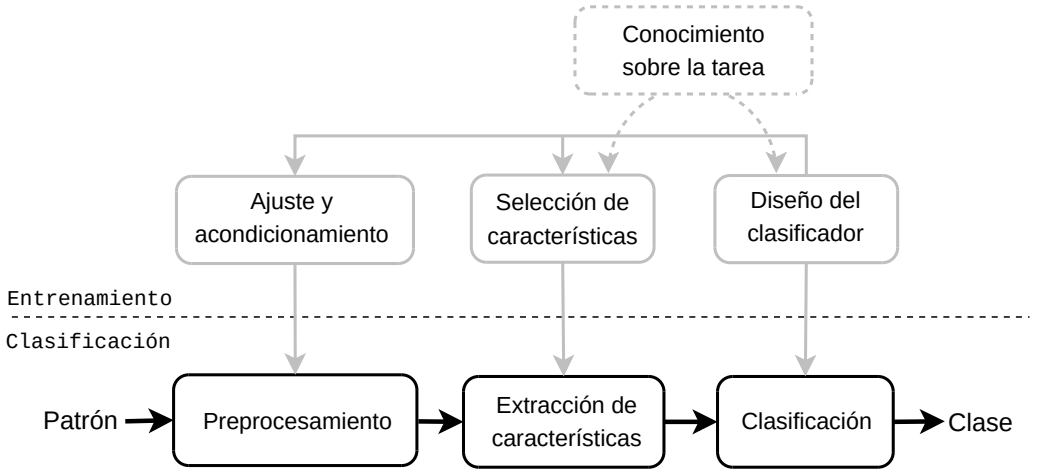



Figura 1.2: Diagrama funcional del proceso de reconocimiento de formas.

correspondiente a la fase de clasificación. El primer paso es el *preprocesamiento* del patrón de entrada, donde se aplican operaciones de realce, restauración y otros para mejorar la calidad del patrón y facilitar la tarea de las etapas siguientes. A continuación, el patrón preprocesado es analizado y una representación alternativa es obtenida a partir de la *extracción de características*. En este bloque se realizan mediciones sobre rasgos distintivos del patrón, obteniéndose como salida una descripción que puede ser numérica o simbólica. En último término, la descripción obtenida es procesada por el *clasificador*, que le asigna una etiqueta de clase de acuerdo al conocimiento que posea sobre las distintas clases y sus aspectos relevantes para la toma de la decisión.

Mientras que para una gran mayoría de aplicaciones la fase de clasificación es de relativamente alta velocidad de ejecución, en todas ellas se tiene una fase de entrenamiento costosa tanto en tiempo insumido como en recursos computacionales necesarios. Este alto costo tiene su fundamento en todos los ajustes y pruebas que se deben realizar sobre cada bloque, considerando su desempeño individual y conjunto con los otros bloques, a fin de lograr el mejor desempeño posible en clasificación.

Durante el entrenamiento se deben ajustar técnicas y parámetros del preprocesamiento, el cual influirá en la capacidad del bloque de extracción de características para realizar la descripción del patrón. Si se trabaja con patrones de imágenes, por ejemplo, el bloque de preprocesamiento puede incluir una mejora de contraste y una reducción de ruido gaussiano; en señales de habla se puede incluir un filtro de preénfasis; en electrocardiograma se puede

aplicar un filtro de ruido de línea; operaciones que en todos los casos no modifican el dominio de representación del patrón original.

Como se mencionó anteriormente, el bloque de extracción de características es el que permite obtener una representación alternativa del patrón, en un espacio cuya dimensión es generalmente mucho menor que la dimensión original. Para lograr que la descripción resultante sea favorable a los efectos de la clasificación, en el sentido de evitar redundancia de información que confunda las clases, en la fase de entrenamiento debe aplicarse un ciclo de ajuste (en la práctica, iterativo) para definir qué características se van a medir, cuáles de ellas son discriminativas de las clases de interés y cómo finalmente se genera la descripción en base a las características seleccionadas. Esta fase, al igual que la siguiente, es de experimentación numérica exhaustiva. Ejemplos de esta etapa pueden encontrarse en el reconocimiento de habla, donde tramos de señal de audio de 1024 muestras son representados por vectores de 16 valores conteniendo pistas frecuenciales; en reconocimiento facial suele normalizarse la imagen del rostro a un cuadrado de 20x20 píxeles (dimensión del patrón: 400) que luego se representa mediante 15 características (eigenfaces).

La última etapa del entrenamiento consiste en la conformación del clasificador, donde se diseñan y prueban modelos que sean capaces de efectuar una asignación satisfactoria de etiquetas de clase a las descripciones de entrada. El término "satisfactoria" es referido a que la capacidad del clasificador de tener éxito en su tarea debe ser evaluada mediante mediciones objetivas o subjetivas que permitan al operador introducir cambios en el clasificador, optimizando el desempeño en consecuencia.

Finalmente, es de notar que en la definición de un sistema de RF es necesario siempre tener algún conocimiento *a priori* sobre la tarea a resolver, el cual permitirá encaminar la búsqueda de soluciones en los dos últimos bloques del sistema. La vía de retroalimentación desde el diseño del clasificador hacia los dos primeros bloques de entrenamiento simbolizan el hecho de que los subsistemas no son independientes entre sí, sino que muchas veces es necesario realizar cambios en algún bloque inducidos por ajustes en otro; por ej., cuando por un cambio en la técnica de clasificación se deben modificar las características a extraer. Esta vía también simboliza el hecho de que para atacar tareas complejas es aconsejable tener además una visión global del sistema.

### 1.1.2 Métodos del reconocimiento de patrones

Actualmente existe un amplio abanico de técnicas para construir un clasificador, con variada naturaleza de la constitución de los patrones, modelos de clases, algoritmos de aprendizaje, y otros aspectos específicos. En particular, el tipo de modelo a aplicar estará en concordancia con la representación elegida para los patrones, la cual puede ser actualmente dividida en dos grandes grupos:

- Numérica no estructurada: los patrones se especifican mediante un conjunto de valores numéricos sin una relación morfológica particular. Un ejemplos de este tipo son patrones constituídos por mediciones de longitud y ancho de sépalo y pétalo para clasificación de plantas de iris [3].

- Estructurada: los patrones son objetos que se representan, a su vez, por un conjunto de (sub)patrones constituyentes denominados *primitivas*, que guardan diferentes tipos de interrelaciones entre sí (grafos, cadenas, etc.) y que construyen jerárquicamente objetos de mayor complejidad. Un ejemplo de este tipo de patrones lo encontramos en el reconocimiento de secuencias de ADN, compuesta por cadenas de genes que a su vez están formadas por subcadenas de aminoácidos [12].

A su vez, estos grupos de representación de los patrones dan lugar –respectivamente– a los dos conjuntos principales en los que se pueden agrupar los métodos de clasificación [2]:

- Métodos estadísticos o geométricos: basados en la teoría estadística de la decisión, trata a los patrones como vectores numéricos que constituyen puntos en un espacio $n$-dimensional, que se agrupan en zonas determinadas del espacio. Los clasificadores, entonces, buscan identificar estos grupos mediante patrones prototipo y medidas de distancia entre ellos, generando fronteras entre grupos que permitan la asignación de clases.

- Métodos estructurales o sintácticos: basados en la teoría de lenguajes formales, trata a los patrones como cadenas de símbolos. En este caso, los clasificadores constituyen autómatas o gramáticas que aplican reglas sintácticas para determinar si los patrones pertenecen al lenguaje aceptado por el sistema.

En el RF estadístico podemos considerar diferentes aproximaciones en el aprendizaje según se disponga (o no) de muestras etiquetadas durante el mismo, que sirvan como ejemplos en el ajuste del clasificador. Esta dicotomía da lugar a la siguiente división de métodos:

- Aprendizaje supervisado: se basan en la búsqueda de regiones de decisión entre conjuntos de puntos, por ej: una recta en un espacio 2D para un problema linealmente separable. Dependiendo de la tarea, es posible que se posea algún conocimiento preliminar (o no) acerca de la estructura estadística de las clases, lo cual subdivide a los métodos en:
  - Aproximaciones paramétricas: las fronteras de decisión se definen de acuerdo a las distribuciones de probabilidad asumidas para las clases. La teoría de la decisión de Bayes es la que fundamenta el método aplicado en este caso, el clasificador Gaussiano, con el cual es posible estimar los parámetros desconocidos de las funciones de densidad de probabilidad (*fdp*).
  - Aproximacion no paramétricas: los patrones etiquetados (prototipos) sirven de guía a los métodos de generación de las fronteras de decisión. Aquí no se asume ninguna *fdp* en particular para las clases, sino que el entrenamiento divide el espacio de representación de acuerdo a los patrones disponibles como ejemplo. Algunos métodos disponibles bajo esta aproximación son los árboles de decisión, aprendizaje adaptativo de la *fdp* (algoritmo de cuantificación vectorial LVQ, del inglés *learning vector quantization*), clasificación basada en distancia (*k-vecinos*), y otros.
- Aprendizaje no supervisado: en ocasiones puede no estar disponible el conocimiento de un experto para realizar el etiquetado de los patrones, por lo cual estos métodos se basan en realizar agrupamientos de los mismos de acuerdo diferentes criterios de homogeneidad entre grupos. Surgen así, entre otros, algoritmos con número de clases fijado a priori (K-medias, ISODATA), número de clases desconocido (algoritmo adaptativo), métodos basados en grafos, etc. [13]

### 1.1.3 El proceso de clasificación: conceptos y notación

En el contexto del RF estadístico podemos considerar a un patrón como una variable aleatoria $n$-dimensional

$$\mathbf{y} = [y_1, y_2, \ldots, y_n]^T,\ y_i \in \mathbb{R},\ i = 1, 2, \ldots, n,$$

la cual representa un punto en el espacio de patrones $P \in \mathbb{R}^n$.

La información relevante para el clasificador es obtenida durante la extracción de características, donde se obtiene un nuevo conjunto de variables:

$$\mathbf{x} = [x_1, x_2, \ldots, x_d]^T,\ x_j \in \mathbb{R},\ j = 1, 2, \ldots, d,$$

en el espacio de características $E \in \mathbb{R}^d$ . Ya sea por la propia selección de características o por la aplicación de alguna técnica de reducción de dimensiones, como por ejemplo PCA (Análisis de Componentes Principales, del inglés *Principal Component Analysis*), se tiene usualmente que $d \leq n$ [14].

El proceso de clasificación, entonces, consiste en la asignación de cada patrón $\mathbf{x}$ a una de las $c$ clases informacionales o categorías disponibles en el conjunto de etiquetas de clase $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$. A menudo, dependiendo de la dificultad de la tarea, se puede considerar tambien un conjunto extendido $\Omega^* = \{\omega_1, \omega_2, \ldots, \omega_c, \omega_0\}$, con $\omega_0$ siendo la *clase rechazo* a la que es asignado el patrón si no satisface los criterios de asignación para las otras clases.

La teoría de la decisión define la *probabilidad a priori* $P(\omega_i)$ de una clase $\omega_i$ como la probabilidad de que una muestra arbitraria pertenezca a $\omega_i$, y se puede calcular como la proporción de muestras de esa clase respecto del total. Asimismo, se define la *densidad condicional* $P(\mathbf{x}|\omega_i)$ a la distribución característica de las muestras de $\omega_i$ en $E$.

Con estas definiciones, y suponiendo que se dispone de una muestra $\mathbf{x}$ en particular, es posible conocer en cierta medida cuál es la probabilidad de que la muestra pertenezca a una clase mediante la *regla de Bayes* [2]:

$$\begin{aligned} P(\omega_i|\mathbf{x}) &= \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})} \\ &= \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} P(\mathbf{x}|\omega_j)P(\omega_j)}, \end{aligned} \tag{1.1}$$

donde $P(\omega_i|\mathbf{x})$ se denomina *probabilidad a posteriori* y $P(\mathbf{x})$ es la *probabilidad incondicional*.

Conceptualmente, la probabilidad *a posteriori* puede interpretarse como la probabilidad de observar una etiqueta sabiendo cuál es la muestra bajo consideración. Justamente esta interpretación es la que da lugar a la *regla de clasificación de Bayes*, que establece que cada muestra evaluada debe ser asignada a la clase con mayor probabilidad *a posteriori*:

$$\hat{\omega} = \underset{\omega_i : 1 \leq i \leq c}{\operatorname{argmax}}\ P(\omega_i|\mathbf{x}). \tag{1.2}$$

Un clasificador estadístico genérico es una máquina formada por $c$ (funciones) discriminantes:

$$g_i : E \to \mathbb{R},\ 1 \leq i \leq c, \tag{1.3}$$

tal que dado un patrón $\mathbf{x} \in E$,

$$\mathbf{x} \text{ se asigna a la clase } \omega_i \text{ si } g_i(\mathbf{x}) > g_j(\mathbf{x})\ \forall j \neq i.$$

Teniendo en cuenta la regla dada por (1.2), el *clasificador de Bayes* se define, entonces, como aquél que utiliza las probabilidades *a posteriori* como funciones discriminantes en (1.3).

La probabilidad puntual de error de este clasificador está dada por:

$$P(\text{error}|\mathbf{x}) = 1 - \max_{1 \leq i \leq c} P(\omega_i|\mathbf{x}), \tag{1.4}$$

mientras que la probabilidad media de error del clasificador se calcula como:

$$P(\text{error}) = \int_E P(\text{error}|\mathbf{x})P(\mathbf{x})d\mathbf{x}. \tag{1.5}$$

El error del clasificador de Bayes constituye una cota mínima de error alcanzable por cualquier clasificador, y es solamente calculable de manera exacta si son conocidas $P(\omega_i|\mathbf{x})$ y $P(\omega_i)$. Sin embargo, para la gran mayoría de las aplicaciones que trabajan sobre conjuntos de datos reales y muy numerosos, esta última condición no se cumple ya que es muy difícil que se conozca exactamente la forma de ambas distribuciones. Es aquí, entonces, donde tienen su origen y fundamento todas las técnicas de RF propuestas en los últimos años y que fueran introducidas en la Sección 1.1.2, que buscan estimar las densidades de probabilidad de cada clase, aproximar fronteras entre regiones de decisión, realizar agrupamientos de datos, etc. [2]

## 1.2 Redes neuronales artificiales

### 1.2.1 Generalidades

La *psicología congnitiva* constituye un campo de la psicología que estudia la mente a través de la conducta, esto es, explica cómo la mente desarrolla actividades de adquisición, almacenamiento y procesamiento de la información

disponible a fin de generar una conducta en consecuencia. A mediados del siglo XX, la forma de pensar estos procesos dio lugar a la *psicología cognitiva clásica*, la cual sostiene básicamente que la conducta es el resultado de la aplicación de un conjunto de reglas a las propiedades semánticas de los objetos representados. Esta idea, trasladada tecnológicamente a los intentos por construir sistemas artificiales que reprodujeran el comportamiento descripto, dio lugar a la aparición de la *Inteligencia Artificial (IA) clásica*.

El paradigma de la IA clásica trata al cerebro como el *hardware* de una computadora, y al procesamiento de la información llevada a cabo por la mente como el *software*. El programa informático instrumenta el conjunto de reglas y su aplicación de manera secuencial sobre la información de entrada, aplicando elementos de la lógica matemática para realizar sus cómputos y resolver los problemas. Así, en los años siguientes se avanzó en el desarrollo de algoritmos, codificación en lenguajes de programación, optimización de códigos, etc., sobre la base del desarrollo tecnológico del *ordenador de Von Neumann* [15].

Sin embargo, y a pesar del desarrollo alcanzado, hacia mediados de los años '80 se planteaba la pregunta: *¿En dónde radican las diferencias entre la mente y una computadora, que hacen a las personas mucho más hábiles para realizar las tareas cognitivas?* El cerebro está constituído por una enorme cantidad de neuronas, células especializadas del sistema nervioso que llevan a cabo el tratamiento de los estímulos recibidos del entorno. El equivalente en los circuitos electrónicos son las puertas de silicio, las cuales logran velocidades de funcionamiento varios órdenes de magnitud mayores que la capacidad natural de trabajo de las neuronas. La respuesta a la pregunta, entonces, era "la diferencia está en el software", la cual satisfacía sólo parcialmente a un grupo de pensadores.

El *conexionismo* constituye una corriente surgida en el campo de la psicología cognitiva como alternativa al pensamiento clásico expuesto anteriormente. El alto rendimiento del cerebro en tareas cognitivas de alta complejidad, tales como el reconocimiento de formas visuales o auditivas, fue siempre una motivación importante para el modelado de su funcionamiento. La teoría del conexionismo se inspira en las semejanzas biológicas de estructura y funcionamiento del cerebro: plantea que la mente está compuesta por un gran número de unidades de procesamiento (las neuronas) conectadas entre sí de maneras muy complejas (las sinapsis), y los procesos mentales surgen como consecuencia de interacciones de excitación o inhibición entre estas unidades, principalmente *en paralelo* y no de manera secuencial. Esta nueva forma de explicar la mente y la conducta surge a mediados de los años '80 con el trabajo de D. Rumelhart y J. McClelland [16], la que fuera inicialmente conocida

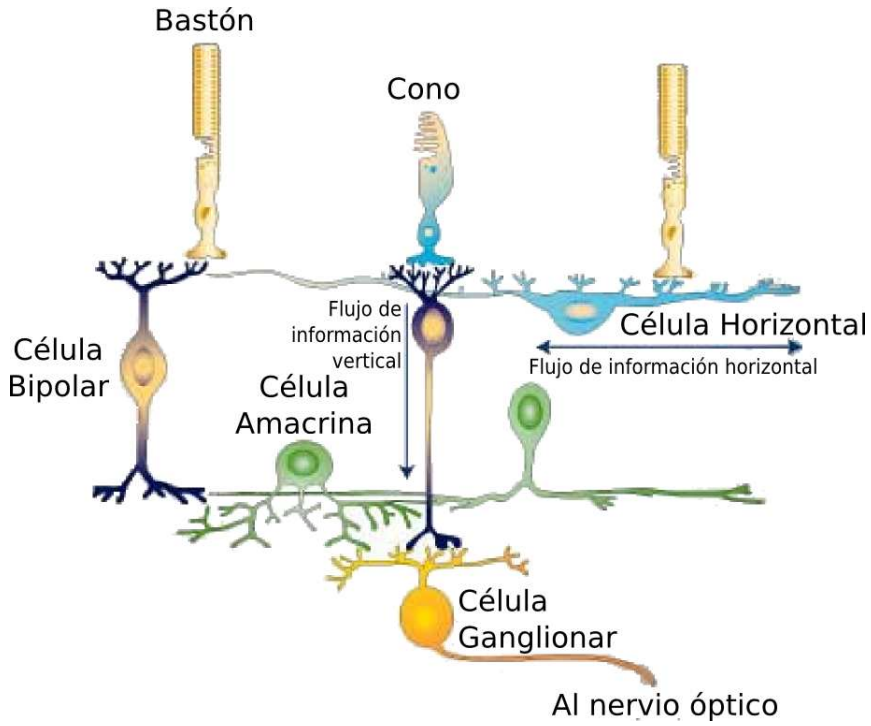



Figura 1.3: Diagrama de la estructura celular de la retina humana. Adaptada con permiso de David Colella, The MITRE Corporation, USA.

como *Procesamiento Paralelo Distribuído*. Actualmente, el conexionismo es un tema de investigación de gran auge en un variado número de disciplinas, entre las cuales podemos citar a la biología, la lingüística y la inteligencia artificial.

Los sistemas conexionistas consisten en un gran conjunto de unidades de procesamiento conectadas por enlaces, los cuales distribuyen los patrones de activación generados entre las unidades, permitiendo el procesamiento en paralelo. Las estructuras principales logradas mediante el conexionismo son las *Redes Neuronales* (RN), nombre recibido al ser las unidades de procesamiento usualmente denominadas *neuronas* o *nodos*. Sin entrar en detalles fisiológicos, y sólo con el ánimo de ejemplificar la analogía entre las redes neuronales biológicas y las modeladas mediante los sistemas conexionistas, en la Figura 1.3 se muestra un esquema de la conformación de la retina, capa interna del globo ocular donde se aloja el tejido fotorreceptor. En el esquema son visibles diferentes tipos de neuronas, con algunas sinapsis hacia adelante y otras laterales. Estas interconexiones permiten a las células lograr procesamientos complejos en el primer nivel de formación de las imágenes en el sistema visual humano.

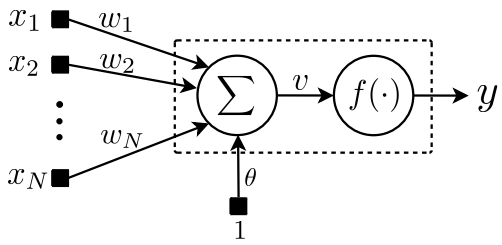



Figura 1.4: Arquitectura de un Perceptrón.

A continuación se revisa el funcionamiento básico de una única neurona artificial, y posteriormente se extiende el estudio a los modelos conexionistas implementados en este trabajo.

### 1.2.2 El Perceptrón como unidad elemental de procesamiento

El Perceptrón, introducido por Rosenblatt en 1958, constituye la red neuronal más simple, constituída solamente por una neurona. Esta neurona recibe estímulos de su entorno por medio de un conjunto de sensores simbolizados por las entradas, realiza sobre ellas un procesamiento elemental de integración y obtiene un valor de salida que representa la actividad de la neurona. La arquitectura del Perceptrón se ilustra esquemáticamente en la Figura 1.4.

Las entradas al Perceptrón se definen mediante un vector dado por

$$\mathbf{x} = [x_1, x_2, \ldots, x_N]^T.$$

Las conexiones sinápticas desde las fuentes de estímulos externos se representan mediante enlaces denominados *pesos sinápticos*, los cuales ponderan de manera diferente cada entrada de acuerdo a su importancia en el procesamiento de los estímulos. El conjunto de pesos está dado por

$$\mathbf{w} = [w_1, w2, \ldots, w_N]^T.$$

Existe una entrada adicional cuyo peso $\theta$, denominado *umbral* o sesgo, fija un valor mínimo que debe ser superado por la integración de las entradas para excitar o inhibir la actividad neuronal. Este peso puede ser agregado al vector $w$ como un primer elemento adicional $w_0 = \theta$, considerando una entrada adicional $x_0 = 1$.

La entrada neta del Perceptrón, o *potencial sináptico*, se calcula como

$$v = \sum_{i=0}^{N} w_i x_i. \tag{1.6}$$

Finalmente, la salida se obtiene por medio de una función de activación que se aplica sobre la suma ponderada de los estímulos de entrada para determinar la actividad de la neurona, según[3]

$$y = f(v). \tag{1.7}$$

En esta ecuación no se detalla específicamente la forma de $f(\cdot)$, ya que existen definiciones de diferentes funciones de activación no lineales que limitan el rango de salida del sistema. El Perceptrón *bipolar* es aquél que aplica la función

$$f(v) = \begin{cases} 1 & \text{si } v \geq 0 \\ -1 & \text{si } v < 0. \end{cases}$$

De igual manera, se define el Perceptrón *binario* mediante la expresión

$$f(v) = \begin{cases} 1 & \text{si } v \geq 0 \\ 0 & \text{si } v < 0. \end{cases}$$

Una función usualmente aplicada es la función sigmoidea, ya que provee una salida prácticamente biestable como el Perceptrón binario pero continua (fácilmente derivable). Esta función presenta, además, grandes semejanzas con las características de transferencia no lineal de las neuronas biológicas. La función sigmoidea (a veces denominada *logística*) se define como

$$f(v) = \frac{1}{1 + e^{-v}}. \tag{1.8}$$

En la Figura 1.5 se muestran las gráficas de las funciones de activación descriptas. Cualquiera sea la función aplicada, cuando la salida se encuentra en nivel alto (cercano o igual a 1) se dice que el Perceptrón está activado, y cuando la salida es baja (cercana o igual al valor mínimo de $f$) se dice que el Perceptrón se encuentra desactivado. De esta manera, es posible ver al Perceptrón como un clasificador simple de dos clases.

[3]Notar que en la bibliografía suele emplearse, alternativamente a la terminología aquí descripta, el término *función de activación* a la salida lineal calculada sobre la entrada neta y *función de salida* a la $f(\cdot)$ de (1.7). Un ejemplo de este uso puede verse en [17].

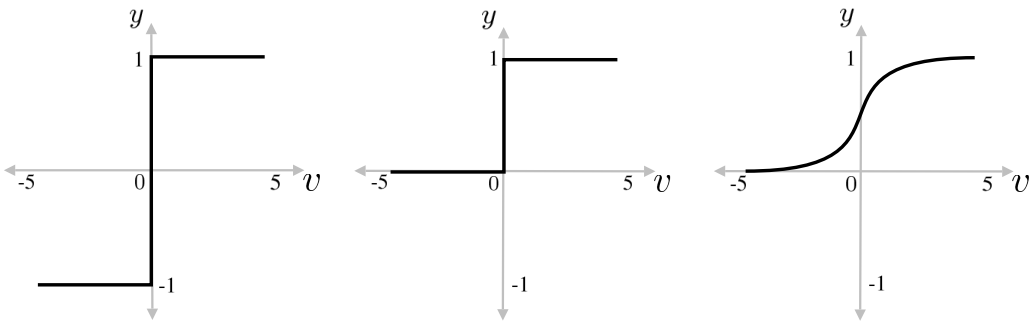



Figura 1.5: Funciones de activación: (a) bipolar, (b) binaria, (c) sigmoidea (ecuación logística).

En una RN el conocimiento es almacenado en los pesos de los enlaces. El desempeño frente a un problema de clasificación es dependiente de los valores que tomen los pesos y también de la interconexión de nodos, esto es, la topología que se defina para la red.

El aprendizaje de una RN consiste en incorporar al modelo el conocimiento necesario para cumplir con la tarea, entendiendo conocimiento como la información que permite a una persona o máquina realizar una interpretación de su entorno [18]. El proceso es llevado a cabo por medio de un algoritmo diseñado para modificar los pesos de cada enlace, conforme los patrones de entrenamiento son presentados al modelo. Cuando se satisface algún criterio de finalización (o de convergencia) preestablecido, el algoritmo es detenido y se dice que la RN está entrenada y lista para su funcionamiento en clasificación.

Según el tipo de RN que se trate, el algoritmo de aprendizaje presenta diferentes formulaciones. En el caso particular del Perceptrón con función de activación sigmoidea, el esquema más aplicado recibe la denominación de *regla delta*[4], y se basa en ajustar los pesos para reducir la diferencia entre el valor dado como correcto en el etiquetado del patrón (salida deseada) y la salida calculada por el Perceptrón. El error medido en el paso $t$ del entrenamiento es

$$e(t) = d(t) - y(t), \tag{1.9}$$

donde $d(t) \in \{0, 1\}$ es el valor de salida deseada del patrón e $y(t)$ es la salida del Perceptrón ante el patrón de entrada $\mathbf{x}(t)$. El ajuste óptimo ocurre

[4]Modificación del algoritmo de mínimos cuadrados medios, también denominado regla Adaline o regla Widrow-Hoff en honor a sus creadores [19], propuesto para nodos sin función de activación no lineal.

al maximizar la correlación entre ambas salidas, equivalente a minimizar el criterio del error cuadrático instantáneo

$$J(\mathbf{w}) = \frac{1}{2}\big(d(t) - y(t)\big)^2, \tag{1.10}$$

cuyo gradiente es

$$\nabla J(\mathbf{w}) = -\big(d(t) - y(t)\big)y'(t)\mathbf{x}(t). \tag{1.11}$$

De esta forma, la regla de actualización de los pesos queda dada por [20]

$$w_i(t+1) = w_i(t) + \eta\big[d(t) - y(t)\big]y'(t)x_i(t), \quad 0 \leq i \leq N \tag{1.12}$$

donde los pesos $w_i(0)$ son inicializados de manera aleatoria, la derivada de $y(t)$ toma la forma simple $y'(t) = y(t)[1-y(t)]$, y $\eta$ es una constante positiva denominada *coeficiente de aprendizaje* que regula la velocidad de cambio de los pesos.

Puede observarse que los pesos serán modificados solamente si $e(t) \neq 0$, esto es, si ocurre un error en la clasificación del patrón. El aprendizaje se obtiene como una minimización del término de error en (1.12) por gradiente descendiente a través de la superficie de error en el espacio de los pesos.

El Perceptrón es capaz de lograr un ajuste de los pesos con error cero solamente cuando las entradas pertenezcan a dos clases diferentes, y además las clases puedan ser separadas geométricamente por un hiperplano en el espacio $N$-dimensional de los patrones. En estas condiciones llevan a que el Perceptrón solamente pueda resolver que sean linealmente separables.

Para superar estas limitaciones se han desarrollado numerosas arquitecturas neuronales. A continuación se exponen las características y algoritmos de entrenamiento de los dos tipos de modelos conexionistas explorados en esta tesis.

### 1.2.3 Perceptrón multicapa

El Perceptrón multicapa (PMC) constituye un modelo conexionista que generaliza al Perceptrón. Basado en múltiples Perceptrones como nodos de procesamiento, el PMC se encuentra constituído por una capa de entrada, una capa de (nodos de) salida y una o más capas (denominadas *ocultas*) entre las mencionadas.

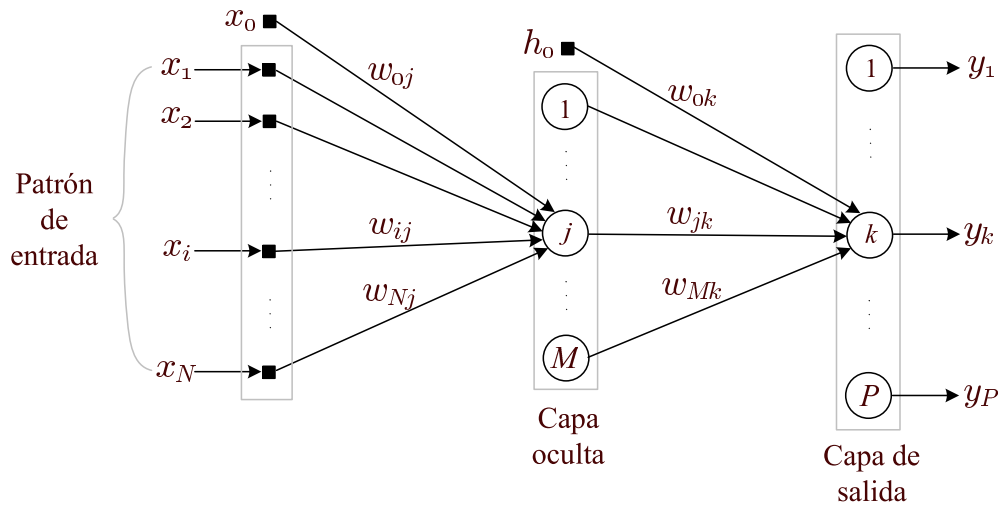



Figura 1.6: Ilustración de la arquitectura de un PMC de una capa oculta. Por conveniencia gráfica, solamente se dibujan los enlaces entre algunas entradas y el $j-$ésimo nodo de la capa oculta, pero existen enlaces de las $N+1$ entradas a los $M$ nodos ocultos (el mismo razonamiento se aplica a los enlaces entre la capa oculta y la capa de salida).

El PMC es entrenado mediante aprendizaje supervisado, en el cual los patrones de entrenamiento incluyen la salida deseada formando los pares

$$\{\mathbf{x}^q, d^q\}, \quad 1 \leq q \leq Q$$

siendo $Q$ siendo la cantidad de patrones del conjunto de entrenamiento.

La modificación de los pesos está basada en regla de corrección del error, como se mencionara previamente para el Perceptrón simple. En particular, una implementación bien establecida de este proceso de aprendizaje es utilizada en este trabajo: el algoritmo de retropropagación del error.

La estructura de un PMC de una capa oculta se define como muestra esquemáticamente la Figura 1.6, con una capa de entrada de $N$ unidades, una capa oculta de $M$ nodos y la capa de salida de $P$ nodos. Además, las unidades denotadas como $x_0$ y $h_0$ corresponden a la unidad de umbral para las capas de entrada y oculta, respectivamente, ambas con valor igual a 1. Es de notar la diferencia funcional entre los perceptrones simples (en la Figura 1.6 graficados con círculos) y las unidades de entrada (graficadas con cuadrados), donde éstas últimas solamente copian el valor correspondiente de la característica del patrón.

Cada capa está completamente conectada con su sucesora a través de los

pesos

$$\mathbf{W}^x = \{w_{ij}\},\ \text{con}\ \begin{cases} 0 \leq i \leq N \\ 1 \leq j \leq M \end{cases}$$
$$\mathbf{W}^h = \{w_{jk}\},\ \text{con}\ \begin{cases} 0 \leq j \leq M \\ 1 \leq k \leq P \end{cases}$$

como las matrices de pesos para los enlaces entrada-oculta y oculta-salida, respectivamente.

El algoritmo de retropropagación comienza con una fase de propagación hacia adelante, donde se obtienen las salidas de cada Perceptrón simple que conforma la red. Al igual que en el modelo mencionado, la salida de cada nodo se calcula como la combinación lineal de sus entradas y pesos, resultando

$$H_j = \sum_{i=0}^{N} w_{ij} x_i, \quad 1 \leq j \leq M \tag{1.13}$$
$$Y_k = \sum_{j=0}^{M} w_{jk} h_j, \quad 1 \leq k \leq P \tag{1.14}$$

siendo $H_j$ e $Y_k$ las salidas para las capas oculta y de salida, respectivamente.

En el paso final hacia adelante se obtienen las activaciones de cada nodo como

$$h_j = f\left(H_j\right), \quad 1 \leq i \leq N \tag{1.15}$$
$$y_k = f\left(Y_k\right), \quad 1 \leq j \leq M \tag{1.16}$$

estando $f(\cdot)$ definida como en (1.8).

El aprendizaje del PMC puede ser visto como un proceso de optimización sobre la función criterio error cuadrático, que busca la minimización sobre el conjunto de entrenamiento de la función

$$E(\mathbf{w}) = \frac{1}{2} \sum_{k=1}^{M} (d_k - y_k)^2, \tag{1.17}$$

donde $\mathbf{w}$ representa el conjunto completo de pesos de la red y $d_k$ es la $k$-ésima salida deseada del $q$-ésimo patrón de entrenamiento.

La regla delta para el PMC establece que en la capa de salida los pesos se

actualizan según

$$\begin{aligned}\Delta w_{jk} &= w_{jk}(t+1) - w_{jk}(t) \\ &= -\eta \frac{\partial E}{\partial w_{jk}} \\ &= \eta(d_k - y_k) f'(Y_k) h_j, \text{ con } \left\{ \begin{array}{l} 0 \leq j \leq M \\ 1 \leq k \leq P \end{array} \right. \end{aligned} \tag{1.18}$$

siendo $\eta$ el coeficiente de aprendizaje.

Con respecto a la actualización de los pesos que llegan a la capa oculta, debemos notar que no se disponen de las salidas deseadas para los nodos de la capa oculta. De todos modos, es posible inferir una regla de aprendizaje basada en modificar estos pesos para tratar de disminuir el error en la capa de salida, mediante la propagación hacia atrás o *retropropagación* de las cantidades $(d_k - y_k)$. Así, la actualización de los pesos se realiza mediante el gradiente descendiente de la función criterio error cuadrático, pero esta vez calculado respecto a los pesos $\mathbf{W^x}$ como

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}}, \text{ con } \left\{ \begin{array}{l} 0 \leq i \leq N \\ 1 \leq j \leq M \end{array} \right. \tag{1.19}$$

Mediante la regla de la cadena es posible calcular la derivada parcial en (1.19) según

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial h_j} \frac{\partial h_j}{\partial H_j} \frac{\partial H_j}{\partial w_{ij}}, \tag{1.20}$$

donde los últimos dos factores son calculados como

$$\frac{\partial H_j}{\partial w_{ij}} = x_i, \tag{1.21}$$

$$\frac{\partial h_j}{\partial H_j} = f'(H_j), \tag{1.22}$$

y finalmente

$$\begin{aligned}\frac{\partial E}{\partial h_j} &= \frac{\partial}{\partial h_j} \left\{ \frac{1}{2} \sum_{k=1}^{M} \big[d_k - f(Y_k)\big]^2 \right\} \\ &= -\sum_{k=1}^{M} \big[d_k - f(Y_k)\big] \frac{\partial f(Y_k)}{\partial h_j} \\ &= -\sum_{k=1}^{M} \big[d_k - y_k\big] f'(Y_k) w_{jk}. \end{aligned} \tag{1.23}$$

Reemplazando en (1.19) se llega a la regla de actualización de pesos buscada:

$$\Delta w_{ij} = \eta \left[ \sum_{k=1}^{M} (d_k - y_k) f'(Y_k) w_{jk} \right] f'(H_j) x_i. \tag{1.24}$$

La secuencia descripta, que obtiene el ajuste de pesos para un patrón en particular, se realiza iterativamente para los $Q$ patrones de entrenamiento, actualizándose los pesos luego de la presentación de cada patrón. La convergencia del aprendizaje a una solución adecuada se comprueba al finalizar la presentación del conjunto completo de entrenamiento, verificando el valor de alguna función error prefijada o mediante alguna técnica de estimación del error.

Una modificación introducida en las reglas de actualización de los pesos, con el propósito de acelerar el proceso de convergencia, consiste en la adición de un término de momento en la búsqueda por gradiente descendiente. Este término agrega una cantidad proporcional al último cambio de cada peso $w_i$ (genérico), de acuerdo al *coeficiente de momento* $\alpha$

$$\Delta w_i(t) = -\eta \frac{\partial E}{\partial w_i(t)} + \alpha \Delta w_i(t-1). \tag{1.25}$$

El conjunto de nodos organizados en capas le posibilitan al PMC resolver problemas que no sean linealmente separables. En el caso del modelo de una capa oculta, cada nodo oculto agrega un hiperplano de separación en el espacio de los patrones, mientras que cada nodo de salida genera las regiones de decisión aplicando una operación lógica AND sobre los hiperplanos [21].

La capa de entrada del PMC recibe todo el patrón de entrada a la vez, lo cual induce la extracción de características globales de los objetos bajo análisis, cuyos patrones deben tener la misma longitud para todos los objetos. Esta restricción debe ser salvada mediante el agregado de ceros o la interpolación de los patrones, en caso de que los mismos tengan longitudes diferentes.

### 1.2.4 Redes neuronales recurrentes

La formulación del PMC tuvo un gran impacto en el desarrollo de los modelos conexionistas. Sin embargo, este tipo de redes neuronales responde de manera satisfactoria siempre que una representación espacial completa de los objetos pueda ser obtenida para alimentar la red. En un gran número de

aplicaciones, el tiempo está intrínsecamente emparentado con muchos comportamientos (tales como el lenguaje) y las representaciones de los objetos, donde los patrones se construyen a partir de una serie temporal de eventos que usualmente tienen duraciones desiguales.

La cuestión de cómo representar el tiempo en los modelos conexionistas es de gran importancia, puesto que puede surgir como un problema especial de los modelos de procesamiento paralelo, ya que la naturaleza de los cómputos en paralelo parece ser contraria a la naturaleza serial de los eventos temporales. En este sentido, el PMC es claramente un mal candidato para aplicaciones de procesamiento temporal puesto que realiza un mapeo estático entrada-salida, trabajando con patrones de la misma longitud.

En este trabajo se adaptó el concepto unidimensional del procesamiento de señales temporales al campo bidimensional de las imágenes, donde se puede pensar que las imágenes corresponden a una serie temporal de patrones de bandas oscuras y claras alternantes a lo largo del eje longitudinal medio. De acuerdo a este razonamiento, el concepto de tiempo está dado por la secuencia de vectores que forman las bandas, y las diferencias de duración entre patrones están dadas por las diferencias de longitud entre clases.

Una aproximación que provee a los modelos conexionistas con una especie de memoria, donde los eventos pasados pueden ser usados para mejorar la decisión de la red neuronal sobre el evento actual, es la propuesta general de las *redes neuronales recurrentes* [22, 23, 24]. En estos modelos existen enlaces desde los nodos de las capas ocultas o de salida hacia las capas previas, los cuales actúan como la memoria buscada.

Entre las diferentes redes recurrentes propuestas en la bibliografía, la red de Elman (RE) es una red neuronal parcialmente recurrente que presenta algunas ventajas debido a su representación interna del tiempo. Las conexiones son principalmente hacia adelante, como en un PMC, de manera que cuando se procesa un patrón la red recibe un vector de características (segmento del patrón) en su entrada y las propaga como activaciones a la capa oculta. Las activaciones obtenidas en esta capa son, asimismo, propagadas a la capa de salida y también hacia un conjunto de nodos que forman la denominada *capa de contexto*. Luego, cuando el siguiente segmento es presentado a la red, los valores de activación de los nodos de contexto son también propagados a la capa oculta. De esta manera, la capa oculta actúa evaluando una especie de entrada aumentada, considerando al vector actual y una historia de las activaciones más recientes de la red provistas por la capa de contexto [25].

A continuación se revisa el proceso de aprendizaje, el cual es una adaptación del algoritmo de retropropagación del error introducido para el PMC.

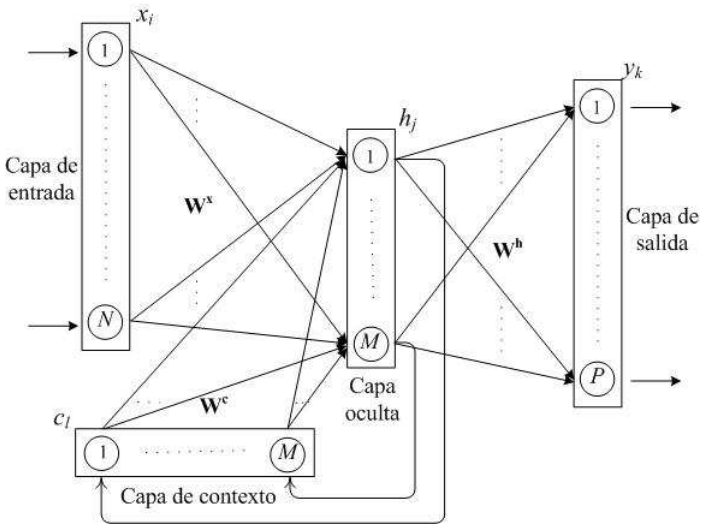



Figura 1.7: Ilustración de la arquitectura de una red de Elman. La notación $x_i$ se utiliza para referirse a las entradas de la red; $h_i$ y $c_l$ a las activaciones de los nodos de la capa oculta y de contexto, respectivamente; mientras que $y_k$ denota las salidas de la red.

En particular, cuando las conexiones de realimentación toman un valor, el algoritmo de retropropagación puede ser usado para el entrenamiento de las conexiones como si se tratara de una red anteroalimentada convencional.

La estructura de una RE de una capa oculta se define con una capa de entrada de $N$ unidades, una capa oculta de $M$ nodos, una capa de contexto de la misma dimensión que la capa oculta, y la capa de salida de $P$ nodos. Las conexiones desde la salida de las unidades ocultas a la capa de contexto corresponden a pesos fijos (no entrenables). Además, las activaciones de las unidades ocultas son copiadas de manera uno-a-uno a las unidades de contexto. La Figura 1.7 muestra un esquema de esta arquitectura.

Se definen

$$\mathbf{W^x} = \{w_{ij}\}$$
$$\mathbf{W^h} = \{w_{jk}\}$$
$$\mathbf{W^c} = \{w_{lj}\}$$

como las matrices de pesos para los enlaces entre las capas entrada-oculta, oculta-salida y contexto-oculta, respectivamente; con $1 \leq i \leq N$, $1 \leq j, l \leq M$, y $1 \leq k \leq P$. Los pesos de las conexiones oculta-contexto son fijados a un valor de 1, a fin de obtener en la entrada de la capa de contexto una copia de las activaciones de la capa oculta.

La salida lineal de cada nodo se define como la combinación lineal de sus entradas y pesos, siendo para las capas oculta y de salida, respectivamente

$$H_j(t) = \sum_{i=1}^{N} w_{ij} x_i(t) + \sum_{l=1}^{M} w_{lj}\ c_l(t), \quad 1 \leq j \leq M \tag{1.26}$$

$$Y_k(t) = \sum_{j=1}^{M} w_{jk} h_j(t), \qquad 1 \leq k \leq P \tag{1.27}$$

Las activaciones de cada nodo son obtenidas en la pasada hacia adelante, como se describe en las siguientes ecuaciones para la capa oculta, de contexto y de salida, respectivamente

$$h_j(t+1) = f\left(H_j(t)\right), \quad 1 \leq j \leq M, \tag{1.28}$$

$$c_l(t+1) = h_j(t), \qquad \forall\ l = j, \tag{1.29}$$

$$y_k(t+1) = f\left(Y_k(t)\right), \quad 1 \leq k \leq P, \tag{1.30}$$

donde $f(\cdot)$ está dada por (1.8). En tiempo $t = 0$, cuando el primer vector de un patrón nuevo es presentado a la red, las activaciones de las unidades de las capas oculta y de contexto son fijadas a 0.

Una vez que la salida de la red es obtenida, el segundo paso en el proceso de entrenamiento consiste en la propagación hacia atrás. Aquí, la diferencia entre los valores de salida actual y la salida deseada se toma como error de la red, y su acumulación sobre todos los vectores de un patrón es la función que el algoritmo de aprendizaje trata de minimizar.

La función criterio de error en las unidades de salida, para el patrón completo, se define como

$$E = \frac{1}{2} \sum_{t=1}^{R} \sum_{k=1}^{P} \Big[ d_k(t) - y_k(t) \Big]^2, \tag{1.31}$$

donde $R$ es la longitud del patrón (número de vectores de características) y $d_k(t)$ es el valor deseado en el tiempo $t$ para la $k$-ésima unidad de salida.

Aplicando el método de corrección de error general de retropropagación con término de momento y la regla de derivación de la cadena, los pesos son

actualizados mediante

$$\Delta w_{jk} = \sum_{t=1}^{R} \Big[\eta\ h_j(t)\ \delta_k(t) + \mu\ \Delta w_{jk}(t)\Big], \tag{1.32}$$

$$\Delta w_{ij} = \sum_{t=1}^{R} \Big[\eta\ x_i(t)\ f_j'\big(H_j(t)\big) \sum_{k=1}^{p} \Big(\delta_k(t) w_{jk}\Big) + \mu\ \Delta w_{ij}(t)\Big], \tag{1.33}$$

$$\Delta w_{lj} = \sum_{t=1}^{R} \Big[\eta\ c_l(t)\ f_j'\big(H_j(t)\big) \sum_{k=1}^{p} \Big(\delta_k(t) w_{jk}\Big) + \mu\ \Delta w_{lj}(t)\Big], \tag{1.34}$$

con

$$\delta_k(t) = f_k'\Big(Y_k(t)\Big)\ \Big(d_k(t) - y_k(t)\Big). \tag{1.35}$$

Aquí, $\eta$ es el parámetro de aprendizaje, el cual especifica el ancho de paso del gradiente descendiente; $\mu$ es el término de momento, que permite agregar una cantidad proporcional al cambio de pesos previo; y $f'(\cdot)$ representa la derivada de la función de activación definida como en (1.8).

Una vez que un patrón ha sido procesado completamente por la red, las activaciones de la capa de contexto son nuevamente fijadas a 0. De esta manera, cuando un nuevo patrón es presentado a la red, ésta no guarda la historia de activaciones de su predecesor.

El proceso de aprendizaje es iterativamente repetido hasta alcanzar un nivel satisfactorio en un ajuste por métodos de estimación del error.

## 1.3 Modelos ocultos de Markov

### 1.3.1 Generalidades

Para una gran diversidad de fenómenos, es posible estudiar la expresión de los mismos (señales) como representaciones que evolucionan de cierta manera. El habla corresponde a la manifestación del proceso de generación de una señal acústica que lleva el mensaje que el hablante intenta comunicar, la cual evoluciona a lo largo del tiempo [26]. Otras señales que se pueden mencionar –provenientes de fenómenos naturales o no– que varían en el tiempo son las mediciones anuales de precipitaciones o de temperatura máxima diaria en

una ciudad (en aplicaciones de meteorología), la cotización de una moneda o la variación de índices bursátiles (en aplicaciones de economía), etc. En las imágenes, puede estudiarse la evolución del brillo a lo largo de las filas y/o columnas, donde los saltos entre píxeles emulan las transiciones entre instantes de tiempo. Se encuentran ejemplos de este tratamiento en la detección de cáncer en imágenes de ultrasonido, etc.

Se pueden construir modelos matemáticos determinísticos o estadísticos para explicar estos fenómenos, analizarlos, realizar predicciones, etc. En el modelado estadístico de señales, es posible representar la evolución en alguna dimensión de los patrones mediante un autómata estocástico con una serie de estados conectados entre sí, que van generando los segmentos del patrón. Aquí, la probabilidad del pasaje de un estado a otro depende de todo el camino recorrido hasta el estado actual.

Las *cadenas de Markov* son procesos estocásticos con un número finito de estados que simplifican el tratamiento estadístico, ya que la transición a un estado posterior solamente depende del estado actual y es condicionalmente independiente de los estados previos (propiedad de Markov). En cada estado, el modelo genera un valor de salida (u observación) de manera determinística, por lo que la construcción de este modelo consiste en la estimación de las probabilidades de transición entre estados, dado el grafo de conexión entre ellos.

Los modelos ocultos de Markov (MOM) (en inglés *HMM, Hidden Markov Models*) son máquinas estocásticas de estados finitos, al igual que las cadenas de Markov, pero cuyos estados pueden emitir cualquier símbolo de salida según una distribución de probabilidad propia de cada estado. Así, dada una secuencia de símbolos de salida, ahora no se conoce de manera determinística cuál fue la secuencia de estados transitados ya que, por ej., solamente un estado podría haber generado toda la secuencia. Este comportamiento le da el nombre de *oculto* al modelo. Se tienen, entonces, dos procesos estocásticos involucrados: el de transición de estados (no observable) y el de generación de símbolos (observable) [27].

Los MOM pueden aproximar secuencias de símbolos de un alfabeto discreto, como los caracteres del código ASCII o las cuatro bases nitrogenadas que conforman una cadena de ácido nucleico, recibiendo la denominación de modelos *discretos*. En otras aplicaciones, los valores observados pertencen a un rango continuo (infintos símbolos observables) y por medio de cuantización vectorial pueden ser discretizados, solución que puede ser suficiente en algunos casos. En otras situaciones, es preferible que el modelo represente la naturaleza continua de las señales, por lo que se extienden las probabilidades de observación discretas a continuas (generalmente mediante mezcla

de densidades Gaussianas), y se tiene el MOM *continuo*. Una variante de este modelo corresponde a los MOM *semi-continuos*, donde las funciones de densidad de probabilidad que representan las observaciones son también continuas, pero cada estado no dispone de sus propias distribuciones sino que existe un único conjunto de distribuciones que es compartido por todos los estados.

En esta tesis se propone la utilización de MOM continuos en la experimentación con imágenes de cromosomas, por lo que el resto de la sección se dedica a los aspectos particulares de este tipo de modelos.

### 1.3.2 Definición de un MOM

Un MOM continuo es una máquina formalmente definida (en notación compacta) como

$$\lambda = (Q, A, B),$$

donde cada uno de las componentes se explica a continuación:

- $Q$: conjunto de estados del modelo, con $q_t \in \{1, \ldots, |Q|\}$ siendo el estado $q$ en el tiempo $t$. Una convención usualmente empleada en estos modelos, también incorporada aquí, es considerar dos estados no emisores de observaciones: el estado de entrada 1 y el estado de salida $|Q|$ (convención que permite, en sistemas complejos, crear modelos de mayor nivel concatenando estos modelos iniciales).
- $X$: espacio de representación de las observaciones (vectores de características), de dimensión igual que los patrones ($X \in \mathbb{R}^d$).
- $A = \{a_{ij}\}$: distribución de probabilidad de transición entre estados, con

$$a_{ij} = P(q_{t+1} = j | q_t = i) \tag{1.36}$$

  siendo la probabilidad de transitar desde el estado $i$ al estado $j$ en el instante de tiempo posterior. El truncado en la dependencia probabilística de la propiedad markoviana está siendo formalizada aquí como

$$P(q_{t+1} = j | q_t = i, q_{t-2} = k, \ldots) = P(q_{t+1} = j | q_t = i). \tag{1.37}$$

  Para una normalización apropiada, debe cumplirse que

$$a_{ij} \geq 0, \quad \forall i, j \in Q$$
$$\sum_{j=1}^{|Q|} a_{ij} = 1, \quad \forall i \in Q$$

- $B = \{b_j(\mathbf{x}_t)\}$: distribución de probabilidad de observaciones.

  Considerando a $\mathbf{x}_t$ el patrón $\mathbf{x} \in \mathbb{R}^d$ que emite el modelo en el estado $j$ en el tiempo $t$, se tiene que:

$$\begin{aligned} b_j(\mathbf{x}_t) &= P(\mathbf{x}_t | q_t = j) \\ &= \sum_{m=1}^{M} c_{jm} \mathcal{M}(\mathbf{x}_t), \quad \forall j \in Q \end{aligned} \tag{1.38}$$

  representa una mezcla de $M$ funciones que aproximan la densidad continua, cada una pesada por un coeficiente $c_{jm}$ para la $m$-ésima mezcla del estado $j$.

  Una función usualmente elegida para el modelado es la Gaussiana multivariada, en cuyo caso la distribución $\mathcal{M}$ toma la forma:

$$\mathcal{N}(\mathbf{x}_t, \mu_{jm}, \mathbf{U}_{jm}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{U}_{jm}|}} e^{-\frac{1}{2}(\mathbf{x}_t - \mu_{jm})' \mathbf{U}_{jm}^{-1} (\mathbf{x}_t - \mu_{jm})} \tag{1.39}$$

  donde:

  $\mu_{jm} \in \mathbb{R}^d$: vectores de media.
  $\mathbf{U}_{jm} \in \mathbb{R}^d \times \mathbb{R}^d$: matriz de covarianzas.

  Para cumplir con las restricciones de una probabilidad, se deben satisfacer:

$$\begin{aligned} & c_{jm} \geq 0, \quad \forall j \in Q,\ 1 \leq m \leq M \\ & \sum_{m=1}^{M} c_{jm} = 1, \quad \forall j \in Q \\ & \int_{\mathbf{x}} b_j(\mathbf{x}) d\mathbf{x} = 1, \quad \forall j \in Q \end{aligned}$$

La Figura 1.8 muestra un ejemplo de un MOM tipo *izquierda-derecha*, en donde sólo se permiten transiciones hacia adelante con posibles saltos entre estados (para modelar secuencias cortas) y bucles hacia el mismo estado (para modelar secuencias más largas). Aquí, el estado 2 genera 3 vectores de observación, el estado 4 genera 2 vectores mientras que el estado 3 sólo genera 1 vector de observación. Los estados no emisores 1 y 5 se utilizan para la construcción de modelos compuestos mediante concatenación, siendo $a_{12} = a_{45} = 1$.

Una vez establecido el modelo general, si se conocen sus parámetros y se tiene una secuencia de observaciones $\mathbf{x}$, es posible saber objetivamente cuál es

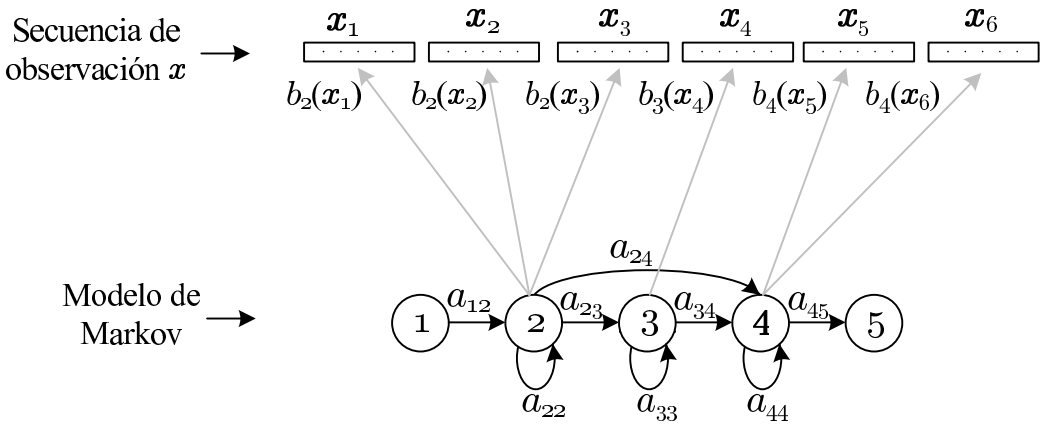



Figura 1.8: Arquitectura de un MOM *izquierda-derecha* con el estado inicial y final no emisores. El MOM transcurre por 6 instantes de tiempo en los cuales genera las observaciones.

el grado de ajuste o representación del modelo a $\mathbf{x}$. Este valor es denominado *verosimilitud* del modelo a la secuencia, la cual calcula la probabilidad de que $\mathbf{x}$ haya sido generada por el modelo $\mathcal{M}$. Para la secuencia mostrada en la Figura 1.8, este valor se calcula como la productoria de las probabilidades de transición y emisión de observaciones en cada estado visitado, según:

$$P(\mathbf{x}|\mathcal{M}) = a_{12}b_2(\mathbf{x}_1)a_{22}b_2(\mathbf{x}_2)a_{22}b_2(\mathbf{x}_3)a_{23}b_3(\mathbf{x}_4)a_{34}b_4(\mathbf{x}_5)a_{44}b_4(\mathbf{x}_6)a_{45} \tag{1.40}$$

La verosimilitud de (1.40) puede ser calculada de manera exacta porque es conocida la secuencia de estados visitada. Sin embargo, en un MOM esta secuencia es justamente oculta al análisis de un observador, por lo que deben evaluarse todas las posibles secuencias de estados que generen las observaciones $\mathbf{x}_1$ a $\mathbf{x}_6$, de acuerdo a:

$$P(\mathbf{x}|\mathcal{M}) = \sum_{v \in V} a_{v(1)v(2)} \left( \prod_{t=2}^{T-1} b_{v(t)}(\mathbf{x}_t) a_{v(t-1)v(t)} \right) a_{v(T-1)v(T)} \tag{1.41}$$

donde $V = \{v(t)/(v(1) = q_1, \ldots, v(T) = q_{|Q|})\}$ es el conjunto de todas las secuencias de estados que en $T$ instantes de tiempo (número de vectores de la secuencia) recorren el MOM desde el estado 1 al $|Q|$.

En el cálculo de la verosimilitud discutida previamente, se supone que los MOM tienen sus parámetros ajustados (medias y varianzas conocidas, en el caso de densidades Gaussianas para las probabilidades de observación). Este conocimiento se adquiere, con algún grado de certeza, mediante el proceso de entrenamiento de los modelos, el cual se detalla a continuación.

### 1.3.3 Algoritmos de entrenamiento

El aprendizaje en un MOM consiste en el proceso de estimar sus parámetros, utilizando para ello los vectores de características que conforman el conjunto de entrenamiento (patrones con sus clases conocidas). Para llegar a la formulación del algoritmo de entrenamiento de un MOM, se introducen seguidamente los conceptos y expresiones de unos coeficientes útiles en el desarrollo posterior: la probabilidad *hacia adelante* y la probabilidad *hacia atrás*[5].

Como se vio en la ecuación (1.40), cada vez que se transita a un estado se emite un vector de observaciones. En un instante de tiempo dado $t$ en la generación de las observaciones, se tiene una cadena parcialmente generada hasta un estado $j$. La probabilidad *hacia adelante* se define como la verosimilitud de la observación parcial desde el inicio de la cadena hasta ese momento, como [28]:

$$\alpha_j(t) = P([\mathbf{x}_1, \ldots, \mathbf{x}_t], v(t) = j|\mathcal{M}) \tag{1.42}$$

Mediante una recursión de (1.42) sobre los instantes de tiempo anteriores, esta probabilidad puede calcularse según:

$$\alpha_j(t) = \begin{cases} a_{1j} b_j(\mathbf{x}_1), & t = 1 \\ \left[ \sum_{i=2}^{N-1} \alpha_i(t-1) a_{ij} \right] b_j(\mathbf{x}_t), & 1 < t \leq T \end{cases} \tag{1.43}$$

con:

$\alpha_1(1) = 1$,
$1 < j < N$, siendo $N = |Q|$.

Finalmente, es de notar que la verosimilitud puede obtenerse como la condición final de recursión:

$$P(\mathbf{x}|\mathcal{M}) = \alpha_N(T) = \sum_{i=2}^{N-1} \alpha_i(T) a_{iN} \tag{1.44}$$

La probabilidad *hacia atrás*, por su parte, se define a partir de la cadena parcial evaluada desde el instante $T$ hacia atrás hasta el instante $t+1$, como

[5]También usualmente denominados por sus términos en inglés, probabilidades *forward* y *backward*, respectivamente.

la probabilidad condicional de que el MOM genere las observaciones para estos instantes sabiendo que en el instante $t$ se encuentra en el estado $j$:

$$\beta_j(t) = P(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_T | v(t) = j, \mathcal{M}) \tag{1.45}$$

Una recursión similar a la probabilidad hacia adelante puede plantearse en este caso, resultando:

$$\beta_i(t) = \begin{cases} a_{iN}, & t = T \\ \sum_{j=2}^{N-1} a_{ij} b_j(\mathbf{x}_{t+1}) \beta_j(t+1), & 1 \leq t < T \end{cases} \tag{1.46}$$

con $1 < i < N$.

Un resultado que puede inferirse de las dos definiciones anteriores es el cálculo de la probabilidad de que el modelo se encuentre en un estado $j$ para el instante $t$, como:

$$P(\mathbf{x}, v(t) = j | \mathcal{M}) = \alpha_j(t) \beta_j(t) \tag{1.47}$$

En algunas aplicaciones, por ejemplo en reconocimiento automático del habla, se construyen modelos pequeños para unidades mínimas (como los fonemas), que luego son concatenados para la creación de modelos a niveles superiores como palabras, frases, etc [27]. Para la tarea de clasificación de cromosomas como se plantea en esta tesis, se construye un único modelo para cada clase con una determinada cantidad de estados cada uno (dependiendo del largo promedio de los cromosomas), sin la necesidad de construir sub-modelos que luego deban ser concatenados.

El algoritmo de Baum-Welch (también conocido por su nombre en inglés, algoritmo *forward-backward*) provee una manera de realizar el aprendizaje de los modelos correspondientes a cada clase de MOM aislados o concatenados, estimando los parámetros que lo definen.

Tomando en consideración el conjunto de patrones (observaciones) correspondientes a la misma clase $\mathcal{X} = \{\mathbf{x}^r / \mathbf{x}^r = (x_1^r, \ldots, x_{T_r}^r) \in X, 1 \leq r \leq R\}$, las probabilidades de transición se reestiman como:

$$\hat{a}_{ij} = \frac{\sum_{1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r - 1} \alpha_i^r(t) a_{ij} b_j(\mathbf{x}_{t+1}^r) \beta_j^r(t+1)}{\sum_{1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r} \alpha_i^r(t) \beta_j^r(t+1)}, \tag{1.48}$$

donde $1 < i, j < N$ y $P_r = P(\mathbf{x}^r | \lambda)$ siendo la probabilidad total del $r$-ésimo patrón.

Las transiciones que parten del estado inicial son estimadas según:

$$\hat{a}_{1j} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{P_r}\alpha_j^r(1)\beta_j^r(1), \tag{1.49}$$

con $1 < j < N$.

Del mismo modo, las transiciones que llegan al estado final son estimadas según:

$$\hat{a}_{iN} = \frac{\sum_{r=1}^{R}\frac{1}{P_r}\alpha_i^r(T_r)\beta_i^r(T_r)}{\sum_{r=1}^{R}\frac{1}{P_r}\sum_{t=1}^{T_r}\alpha_i^r(t)\beta_i^r(t)}, \tag{1.50}$$

con $1 < i < N$.

Restan definir los parámetros de las gaussianas que conforman las mezclas de cada estado del MOM: medias, varianzas y coeficientes de la mezcla. Se define como una variable latente a la probabilidad de ocupación del estado $j$ en el tiempo $t$ para el $r$-ésimo patrón, que en el caso de que la mezcla se reduzca a una sola densidad Gaussiana por estado toma la forma:

$$L_j^r(t) = \frac{1}{P_r}\alpha_j(t)\beta_j(t). \tag{1.51}$$

Si se tienen mezclas de varias densidades gaussianas por estados, la probabilidad de ocupación se calcula para cada componente $m$ de la mezcla, como:

$$L_{jm}^r(t) = \frac{1}{P_r}\mathbf{U}_j^r(t)c_{jm}b_{jm}(\mathbf{x}_t^r)\beta_j^r(t), \tag{1.52}$$

donde

$$\mathbf{U}_j^r(t) = \begin{cases} a_{1j} & t = 1 \\ \sum_{i=2}^{N-1}\alpha_i^r(t-1)a_{ij} & t > 1 \end{cases} \tag{1.53}$$

Finalmente, las fórmulas reestimación de los parámetros de las densidades

gaussianas se expresan como:

$$\hat{\mu}_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^r(t)\mathbf{x}_t^r}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^r(t)}, \tag{1.54}$$

$$\hat{\mathbf{U}}_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^r(t)(\mathbf{x}_t^r - \hat{\mu}_{jm})(\mathbf{x}_t^r - \hat{\mu}_{jm})'}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^r(t)}, \tag{1.55}$$

$$c_{jm} = \frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{jm}^r(t)}{\sum_{r=1}^{R}\sum_{t=1}^{T_r} L_{j}^r(t)}. \tag{1.56}$$

### 1.3.4 Algoritmo de evaluación

La utilidad de los MOM reside en que se puede contar con un conjunto de modelos $\mathcal{M}_i$, cada uno especializado en secuencias de la clase $\omega_i \in \Omega = \{\omega_1, \ldots, \omega_c\}$. Recordando que la clasificación de una secuencia $\mathbf{x}$ puede obtenerse mediante el cálculo de:

$$\operatorname*{argmax}_i P(\omega_i|\mathbf{x})$$

con

$$P(\omega_i|\mathbf{x}) = \frac{P(\mathbf{x}|\omega_i)P(\omega_i)}{P(\mathbf{x})},$$

la idea que surge para clasificación es, entonces, calcular la probabilidad de que cada modelo haya generado la secuencia de observaciones $P(\mathbf{x}|\omega_i) \equiv P(\mathbf{x}|\mathcal{M}_i)$. Luego, el modelo que mejor represente la secuencia –en el sentido de la verosimilitud– será aquél que otorgue la etiqueta de clase al patrón.

La evaluación de la verosimilitud puede obtenerse, como se vio anteriormente, mediante el algoritmo *forward* según (1.44). Sin embargo, en vez de realizar el cálculo total evaluando la recursión completa sobre todos los caminos posibles, en la práctica es conveniente calcular la verosimilitud solamente con la secuencia más probable de estados. Esta es la idea central del algoritmo de Viterbi, que reemplaza la sumatoria en el cálculo de los coeficientes $\alpha_j(t)$ por una operación máximo.

Siendo $\nu_j(t)$ la máxima verosimilitud de la secuencia $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$ observada hasta el estado $j$ en el tiempo $t$, el algoritmo de Viterbi plantea la siguiente recursión:

$$\nu_j(t) = \begin{cases} a_{1j} b_j(\mathbf{x}_1), & t = 1 \\ \\ \max\limits_i \Big[\nu_i(t-1) a_{ij}\Big] b_j(\mathbf{x}_t), & 1 < t \leq T \end{cases} \tag{1.57}$$

con:

$\nu_1(1) = 1$,
$1 < j < N$, siendo $N = |Q|$.

Por lo tanto, la máxima verosimilitud (probabilidad de Viterbi) de la secuencia está dada por:

$$\hat{P}(\mathbf{x}|\mathcal{M}) = \nu_N(T) = \max_i \Big[\nu_i(T) a_{iN}\Big] \tag{1.58}$$

## 1.4 Diseño de experimentos y estimación del error

En cualquier problema de clasificación interesa conocer la exactitud con que el clasificador está llevando a cabo su tarea, mediante la cuenta de la cantidad de veces que acierta (o se equivoca) en el etiquetado de todos los puntos del conjunto de patrones a clasificar. Es posible realizar esta comprobación siempre que se tenga un esquema de clasificación supervisada, ya que en estos casos se cuenta con la clase deseada de cada patrón, la cual puede compararse con la clase otorgada por el clasificador.

La cota de error mínima teórica, como se comentó en el Capítulo 1, corresponde a un clasificador de Bayes, ya que es el que maximiza la probabilidad de acierto sobre las regiones de decisión. Sin embargo, este valor no es calculable de manera práctica, por lo que para establecer la bondad en la clasificación se procede a realizar una estimación del error calculando el porcentaje de patrones incorrectamente etiquetados.

Así, el conjunto $\chi$ conteniendo un total de $N_p = |\chi|$ patrones es empleado tanto para el *entrenamiento* (ajuste de los parámetros) de un clasificador $\Gamma$ como para la *prueba* del mismo (estimación del error). Existen diferentes formas de dividir el conjunto de patrones en conjuntos de entrenamiento y de prueba, las cuales dan lugar a los *métodos de diseño de experimentos*. Luego,

la estimación del error se realiza clasificando todos los patrones de prueba e introduciendo alguna medida basada en la siguiente función base:

$$\zeta(\mathbf{x}_i, \omega_i) = \begin{cases} 1, & \Gamma(\mathbf{x}_i) \neq \omega_i \text{ (error)} \\ 0, & \Gamma(\mathbf{x}_i) = \omega_i \text{ (acierto)} \end{cases} \tag{1.59}$$

Un método utilizado en este trabajo, tanto con los modelos conexionistas como con los modelos ocultos de Markov, es el método de *estimación por conjunto de prueba*. Este esquema es aplicado en una etapa inicial, donde se realiza una gran cantidad de experimentos exhaustivos con modificación de la arquitectura, variación en el ancho del segmento considerado en los patrones, agregado y/o modificación de características, etc., lo que implica una gran cantidad de ciclos de entrenamiento/prueba hasta encontrar un clasificador óptimo en cada caso.

El conjunto inicial de patrones se divide en dos grupos: un conjunto de entrenamiento $\chi^e$ con $N_e = |\chi^e|$ prototipos y un conjunto de prueba $\chi^p$ con $N_p = |\chi^p|$ patrones. La asignación de los patrones a cada conjunto debe hacerse siempre asegurando que:

1. ningún patrón de prueba sea utilizado para el entrenamiento, esto es, $\chi^e \cup \chi^p = \chi$ (con $N_e + N_p = N$) y $\chi^e \cap \chi^p = \emptyset$. De otro modo, se sesgarían en favor del clasificador los resultados obtenidos en la prueba ya que el aprendizaje ajusta los parámetros sobre los mismos patrones de prueba. Una elección usual de cantidades en el particionado es: $N_e = \frac{4}{5}N$ y $N_p = \frac{1}{5}N$;

2. la cantidad de patrones por clase en cada grupo esté balanceada, de otro modo alguna clase podría tener pocos datos de entrenamiento, lo que llevaría a un aprendizaje insuficiente del modelo correspondiente; y

3. ambos conjuntos posean una buena variabilidad de los datos, en el caso de que éstos hayan sido adquiridos con variación en el equipamiento, o en condiciones de trabajo diferentes, o en instancias de tiempo muy separadas (por ej. a lo largo de años), etc.

Finalmente, luego de clasificar los patrones de $\chi^p$ se obtiene la estimación del error $\epsilon(\Gamma)$ del clasificador mediante:

$$\epsilon(\Gamma) = 100 \cdot \frac{1}{N_p} \sum_{(\mathbf{x}_i, \omega_i) \in \chi^p} \zeta(\mathbf{x}_i, \omega_i), \quad [\text{en } \%] \tag{1.60}$$

El método de *estimación del error mediante validación cruzada* fue empleado en este trabajo en la fase final de experimentación, una vez que todos los detalles de arquitectura de los modelos y la información contenida en los patrones fueron ajustados.

Operativamente, el método divide a $\chi$ en $V$ conjuntos con las siguientes propiedades:

$$\begin{cases} \bigcup_{i=1}^{V} \chi_i = \chi \\ \\ \bigcap_{i=1}^{V} \chi_i = \emptyset \\ \\ |\chi_i| \approx \frac{|\chi|}{V},\ 1 \leq i \leq V \end{cases}$$

Se construyen luego $V$ clasificadores, cada uno entrenado con patrones de $\chi - \chi_i,\ 1 \leq i \leq V$, y se calcula el error mediante conjunto de prueba de cada uno según:

$$\epsilon_i(\Gamma_i) = 100 \cdot \frac{1}{|\chi_i|} \sum_{(\mathbf{x}_j, \omega_j) \in \chi_i} \zeta(\mathbf{x}_j, \omega_j), \quad [\text{en } \%] \tag{1.61}$$

Finalmente, el error por validación cruzada se obtiene como el promedio:

$$\epsilon(\Gamma) = \frac{1}{V} \sum_{i=1}^{V} \epsilon_i(\Gamma_i), \quad [\text{en } \%] \tag{1.62}$$

Este método de estimación tiene el objetivo de proporcionar una medida de error promedio del clasificador, el cual es útil como una predicción del desempeño futuro del mismo sobre datos nuevos a adquirir, que no se hallen presentes en el conjunto $\chi$ empleado para la construcción de los modelos [2].

## 1.5 Comentarios de cierre del capítulo

En este capítulo se expusieron los conceptos fundamentales de la teoría de reconocimiento de patrones y las aproximaciones abordadas en esta tesis: las redes neuronales y los modelos ocultos de Markov.

A continuación el documento presenta, en dos partes, las investigaciones realizadas sobre nuevas aproximaciones propuestas basadas en estas técnicas

a los campos de reconocimiento automático de imágenes de cromosomas y señales de habla.

# Parte I

# Reconocimiento de cromosomas

CAPÍTULO 2

# Procesamiento de imágenes de cromosomas

En este capítulo se expone el tratamiento aplicado a las imágenes de cromosomas con el fin de obtener los patrones de entrenamiento y prueba para los experimentos de clasificación.

En primer lugar se introducen los procesos aplicados a fin de obtener el esqueleto de los cromosomas, sobre el cual se realizan las mediciones que caracterizan a los mismos. Esta etapa está basada en el trabajo previo realizado por H. García Peris [29], co-autor de una publicación resultante de esta tesis.

Seguidamente se detallan las características extraídas de las imágenes, las cuales se centraron en el análisis detallado de las bandas de grises a lo largo del cromosoma. Los patrones calculados se plantean como una alternativa a las características globales clásicamente reportadas en la literatura.

## 2.1 Consideraciones preliminares

En citogenética, la tarea de clasificación de cromosomas consiste en asignar cada objeto a una de las 24 clases definidas como estándar clínico [30]. Durante los últimos años se plantearon diferentes aproximaciones del reconocimiento de patrones a un sistema automático de clasificación (o cariotipado), entre las cuales se pueden mencionar las redes neuronales [31, 32], modelos estadísticos [33], la comparación de secuencias mediante *Dynamic Time Warping* ampliamente difundida en reconocimiento de habla [34] y otras técnicas del reconocimiento de patrones [35]. A pesar de la amplia exploración reportada, todavía no se cuenta con una técnica que se tome como estándar para la tarea y que logre obtener desempeños satisfactorios en condiciones de trabajo real, justificando la continuidad en las investigaciones.

Las imágenes de cromosomas presentes en las microfotografías pueden ser analizadas como secuencias de características al recorrer su eje longitudinal, capturando la variabilidad de bandas claras y oscuras. El cálculo de las características a partir de estas imágenes es un tópico abierto [36, 37], con actual interés también sobre otras tareas del reconocimiento de patrones [38].

A manera de introducción del esquema general del método propuesto, se expone en la Figura 2.1 un diagrama en bloques que resume los pasos consecutivos llevados a cabo desde la adquisición de las imágenes hasta la conformación de los patrones para clasificación.

El método consta de tres etapas principales:

- **Preparación de las imágenes**: comprende el acondicionamiento de las imágenes del corpus y la solución de defectos de adquisición.

- **Desdoblado de los cromosomas**: consiste en la obtención de una imagen "recta", eliminando la curvatura típica que pueden exhibir en el momento de la fotografía.

- **Extracción de características**: mediciones sobre la imagen desdoblada que conformarán los patrones de entrenamiento y prueba de los clasificadores.

A continuación se desarrolla cada etapa con más detalle, explicando las funciones y particularidades de cada bloque.

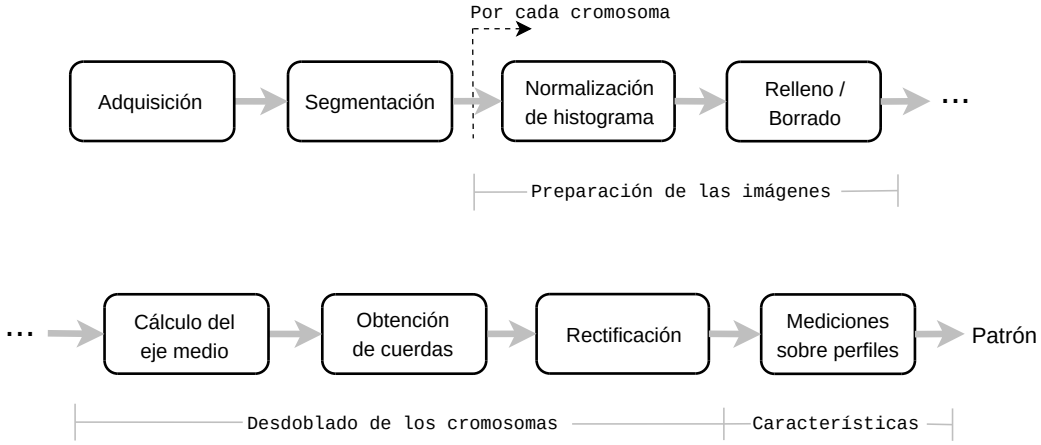



Figura 2.1: Diagrama en bloques del método de obtención de los patrones a partir de las microfotografías celulares.

## 2.2 Corpus de imágenes Cpa

La base de datos utilizada en los experimentos es la más numerosa de su tipo, una corrección del corpus *Copenhagen* completo [39]. Este corpus está compuesto por las imágenes de cromosomas pertenecientes a células humanas cariotipadas en metafase. Las imágenes se encuentran almacenadas en un archivo por célula. Cada archivo contiene las imágenes segmentadas de cada cromosoma orientadas verticalmente con el brazo corto ($q$) en la parte superior y el brazo largo ($p$) debajo, como es convención en citogenética.

El corpus consta de 2804 células, 1344 de las cuales son femeninas y 1460 son masculinas. En su gran mayoría, las células son normales: constan de 46 cromosomas correspondientes a los 22 pares de autosomas (clases 1 a 22), más el par de cromosomas sexuales ($XX$ para células femeninas o $XY$ para masculinas). Asimismo, se encuentra también un grupo de células con aberraciones de número, producto de constelaciones anormales o artefactos en la preparación o adquisición de las imágenes. Existen 26 células con un cromosoma faltante, que dan lugar a afecciones genéticas como el síndrome de Turner (donde falta un cromosoma sexual, también denominada "monosomía X"), otros casos de faltantes pueden ser debidos a problemas de adquisición (cualquier cromosoma). En 37 células hay un cromosoma extra, dando lugar a trisomías de autosomas como el síndrome de Down (triple 21), síndrome de

Edward (triple 18), o formaciones patológicas del par sexual como el síndrome de Klinefelter (indidivuo masculino con un cromosoma $X$ extra) y otras.

El corpus presenta una corrección sobre el conjunto Copenhagen original *Cpr*, que consistió en el re-etiquetado por un experto humano de un subconjunto de 200 cromosomas que aparecieron como *outliers* debido a etiquetados erróneos, y a la corrección de 100 polaridades incorrectamente establecidas en el corpus original. El nuevo conjunto etiquetado de imágenes se denominó *Cpa* [40].

## 2.3 Preparación de las imágenes

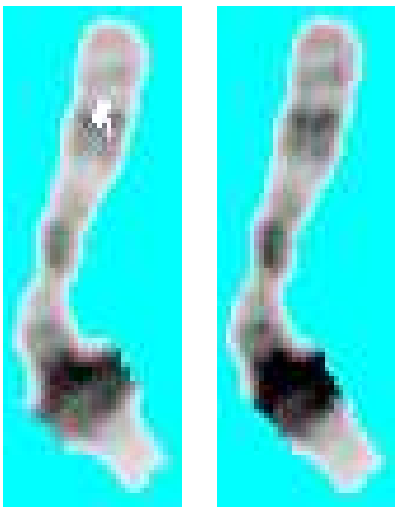
Los procesamientos presentados esta sección se encuentran dirigidos a solucionar imperfecciones encontradas en las imágenes del corpus *Cpa* empleado en los experimentos. Sin embargo, esta condición no invalida la generalidad del método de preprocesamiento ya que constituyen pasos ampliamente utilizados en el tratamiento de imágenes, cuyas formulaciones algorítmicas son independientes de los objetos de aplicación.

En la adquisición de las microfotografías, las condiciones de iluminación pueden variar entre células. Como primer paso en el preprocesamiento se aplica, entonces, una normalización de intensidades de las imágenes a fin de homogeneizar los histogramas[1] ,

Como se explica en detalle más adelante, los vectores de características se extraen recorriendo el cromosoma a lo largo de su eje longitudinal, que en el caso ideal debería ser una curva sin cortes y con solamente dos puntos terminales localizados en los extremos. Los pasos siguientes del preprocesamiento tienen como objetivo reducir y/o eliminar algunas imperfecciones que pueden poseer las imágenes del corpus, las cuales podrían dar lugar a ejes teminados en rulos, con bifurcaciones en los terminales y otras imperfecciones que implicarían la obtención de vectores de características erróneos.

Algunas imágenes presentan defectos en la adquisición que se manifiestan como pequeños huecos en el interior del cromosoma. Estos huecos pueden generar problemas posteriores en el cálculo del eje longitudinal, por lo cual se aplica un algoritmo de relleno iterativo que interpola grises en las cadenas

[1]Diversas operaciones espaciales sobre las imágenes, como la normalización de brillo mencionada, fueron implementadas utilizando las rutinas de manipulación provistas por el paquete Netpbm (URL al 01/08/07: `http://netpbm.sourceforge.net/`).



(a) (b)

Figura 2.2: (a) Cromosoma de clase 1 de la célula *kmkka*, con un hueco en su tercio superior y una protuberancia a ambos lados en el tercio inferior. (b) Imagen normalizada y filtrada, con ambos defectos arreglados. El fondo de las imágenes fue coloreado en cian para facilitar el análisis del efecto del filtrado sobre el borde.

de contorno de los huecos.

Asimismo, otros defectos encontrados en algunas imágenes consisten en protuberancias o golfos sobre el contorno, los cuales llevarían en etapas siguientes a producir ejes con rulos (y sólo un terminal) o ejes con más de un terminal. Para suavizar el borde, y basados en las operaciones fundamentales de erosión y dilatación, se aplican consecutivamente los algoritmos morfológicos de apertura y cierre [41].

El proceso completo de filtrado consiste, entonces, en el rellenado de huecos seguido de una apertura y un cierre. La Figura 2.2(a) muestra un cromosoma de clase 1, donde son visibles un hueco (defecto en la adquisición) y una protuberancia (defecto en la segmentación). El efecto del preprocesado completo se muestra en la Figura 2.2(b). Es posible observar, por un lado, un relleno del hueco de manera satisfactoria, ya que los grises interpolados reconstruyen de manera adecuada la zona perdida (que incluye además una transición de bandas clara a oscura). Por otro lado, se observa también la reducción de la protuberancia lateral por las operaciones morfológicas, las cuales mantienen el contorno en el resto del cromosoma debido a la forma especificada por el elemento estructurante.

## 2.4 Desdoblado de los cromosomas

### 2.4.1 Obtención del eje medio longitudinal

El eje medio del cromosoma es una línea que lo recorre de manera longitudinal de extremo a extremo, y que representa estructuralmente la forma del cuerpo. Los patrones que se extraen de las imágenes se construyen de manera de capturar características de los cromosomas sobre bandas sucesivas (cortes) perpendiculares al eje longitudinal del cromosoma. Así, una vez normalizadas y filtradas las imágenes, el paso siguiente consiste en la obtención de dicho eje.

Para encontrar el eje medio se aplica una combinación de técnicas basadas en la obtención del esqueleto morfológico. En particular, se aplica el algoritmo de González-Woods [41] y el algoritmo de Hilditch [42] de manera sucesiva para aprovechar las bondades de cada método.

El primer proceso obtiene una primera aproximación al esqueleto iterando dos pasadas de un algoritmo de adelgazado de imágenes binarias sobre los puntos del contorno. En la primer pasada, todos los puntos que satisfacen un conjunto de requerimientos son marcados como *puntos del contorno*: aquéllos que no forman parte del esqueleto y pueden ser borrados. En la segunda pasada (una vez procesado todo el cromosoma), los puntos marcados son borrados. Las condiciones evaluadas durante el marcado previenen de aplicar una erosión excesiva o de obtener finalmente puntos desconectados.

El esqueleto obtenido tiene todos sus puntos interiores al borde externo del cromosoma, de modo que un segundo paso consiste en extender los extremos del esqueleto hasta cubrir la longitud total del cromosoma. Esta extensión se aplica sobre ambos extremos de manera iterativa (de a un punto por vez) siguiendo en cada uno de ellos la dirección tangente al terminal [29].

Hasta aquí, el esqueleto generado es 4-adyacente (considera pixeles vecinos solamente en las direcciones N, S, E, O), por lo que el tercer paso consiste en limpiar algunos puntos extra obtenidos en las partes curvas (por la restricción mencionada) aplicando el algoritmo de Hilditch. Este proceso obtiene un esqueleto 8-adyacente (pixeles vecinos en todas las direcciones) barriendo el cromosoma de izquierda a derecha y de arriba hacia abajo, revisando una serie de condiciones para marcar los puntos del contorno y borrar aquéllos que no pertenezcan al esqueleto buscado.

### 2.4.2 Obtención de las cuerdas y rectificación

Denominamos *cuerdas* a las líneas transversales al eje en cada punto de discretización, obtenidas mediante el cálculo de la recta perpendicular a la tangente en cada punto del esqueleto. Las cuerdas, luego, corresponden al segmento comprendido entre los límites del cromosoma (borde exterior izquierdo y derecho), y sobre éstas se realizarán las mediciones que conformarán los patrones de características.

La rectificación, último paso del preprocesamiento, consiste en el muestreo del cromosoma sobre las cuerdas, reacomodando éstas en filas horizontales, una sobre otra. De esta manera, el cromosoma queda orientado de forma vertical, estirado a lo largo de su eje medio. Como los puntos de la cuerda no se corresponden unívocamente con localizaciones de pixeles en la imagen, se aplica interpolación bilineal entre los 4 vecinos para asignar el valor de gris a cada punto.

La Figura 2.3(a) muestra un cromosoma de clase 1, luego de la normalización. En la Figura 2.3(b) se superponen el eje medio longitudinal y las cuerdas, mientras que en la Figura 2.3(c) se puede observar el cromosoma obtenido mediante el desdoblado.

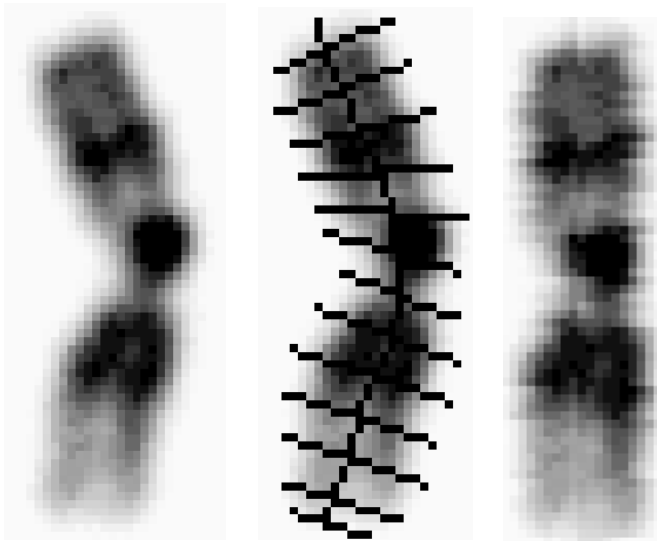
## 2.5 Cálculo de perfiles

Piper y Granum [43] establecieron un conjunto de características que luego fuera adoptado por las investigaciones subsiguientes de diversos grupos como un conjunto de referencia para la tarea. De acuerdo a la cantidad de información necesaria para producir la medida, las características fueron agrupadas en cuatro niveles:

1. Mediciones sobre la imagen: area total, perímetro y otras.

2. Requerimiento de eje del cromosoma: longitud, perfiles de densidad, gradiente y forma.

3. Requerimiento de eje y polaridad: características de forma *global*, que se obtienen multiplicando el perfil de forma por diferentes pesos entre -1 y 1.



(a) (b) (c)

Figura 2.3: Ilustración del preprocesamiento y desdoblado de los cromosomas: (a) cromosoma de clase 1 de la célula *a4078-89*, con histograma normalizado, (b) Superposición del eje medio longitudinal y las cuerdas (sólo una de cada 5 cuerdas fue dibujada, por conveniencia visual), (c) cromosoma desdoblado.

4. Requerimiento de eje, polaridad y localización del centrómero: mediciones de área, longitud y densidad relativas al índice centromérico.

Se entiende por *perfiles* a las representaciones unidimensionales de alguna propiedad del cromosoma, obtenidas a lo largo del eje del mismo. Estas representaciones son usualmente gráficos de la variación del patrón de grises de las bandas y/o otras medidas relacionadas. Por motivos de comparación con estas características "estándar" de nivel 2, en este trabajo se implementaron los perfiles de densidad, gradiente y forma, cuyas formulaciones se introducen a continuación.

### 2.5.1 Perfil de densidad

El perfil de densidad provee una primera aproximación al patrón de bandas del cromosoma. El valor puntual de la densidad está definido como el valor

de gris promedio para la cuerda:

$$d(i) = \frac{\sum_{j=1}^{N_i} c(i,j)}{N_i},\ 1 \leq i \leq M \tag{2.1}$$

donde

$M$: longitud del cromosoma (número de cuerdas),

$N_i$: número de píxeles en la $i$-ésima cuerda del cromosoma,

$c(i,j)$: valor de gris del cromosoma en el pixel $(i,j)$.

Es de notar que la densidad aparece en la bibliografía de la temática con diversas definiciones. Al igual que en este trabajo, algunos autores [44] consideran la suma de grises normalizada respecto al ancho de la cuerda, para reducir el efecto de este valor en la sumatoria; mientras que otros autores [45, 46] consideran la densidad solamente como la suma de grises sobre la cuerda.

### 2.5.2 Perfil de gradiente

El perfil de gradiente corresponde a las diferencias absolutas del perfil de densidad, calculado según:

$$g(i) = |d(i) - d(i-1)|\,,\ 1 < i \leq M \tag{2.2}$$

donde

$M$: ídem (2.1).

Esta operación es similar a calcular la derivada del perfil de densidad, por lo cual al perfil obtenido se le aplica un suavizado mediante un filtro de medias de 3 puntos.

### 2.5.3 Perfil de forma

El perfil de forma consiste en una normalización de la distribución de grises a lo largo de la cuerda, considerando la distancia de los puntos al eje

medio, y se calcula según:

$$f(i) = \frac{\sum_{j=1}^{N_i} c(i,j) \left| j - \frac{M}{2} \right|^2}{\sum_{j=1}^{N_i} c(i,j)}, \ 1 \leq i \leq M \tag{2.3}$$

donde

$M,\ N_i,\ c(i,j)$: ídem (2.1).

La información importante contenida en el perfil de forma, al considerar los anchos de cada cuerda, es el mínimo global de la función: en este punto se encuentra el centrómero.

La Figura 2.4 muestra la imagen de un cromosoma desdoblado y alineado con sus perfiles de densidad y forma, donde en ambos casos se presentan las gráficas con la inversión usual de magnitudes en el eje de ordenadas: el máximo valor de gris (blanco) corresponde al valor de ordenada 0. Se puede observar claramente la corresponencia entre los picos del perfil de densidad con las bandas oscuras del cromosoma, y los valles del perfil con las bandas claras. La correspondencia entre el mínimo global del perfil de forma con la localización del centrómero se marca con la flecha vertical punteada.

## 2.6 Características locales: muestreo de grises

En numerosas células, la microfotografía fue obtenida en un momento de la reproducción celular en el cual son visualmente diferenciables las dos cromátidas[2] en proceso de separación. Así, el eje medio longitudinal queda situado sobre una línea central de grises claros, que corresponde a la separación entre cromátidas. En estos cromosomas, la densidad media de la cuerda podría estar integrando en un solo valor numérico una información valiosa para el clasificador. En el conjunto original de [43], todos los perfiles propuestos son funciones univariadas. En esta tesis se propusieron nuevos conjuntos de características *locales* a cada cuerda, dado que los clasificadores a utilizar se especializan en modelar las distribuciones de probabilidad de las secuencias de entrada por segmentos.

[2]Unidades longitudinales que luego se separan por el centrómero, estando formada cada una por un brazo $q$ y uno $p$.

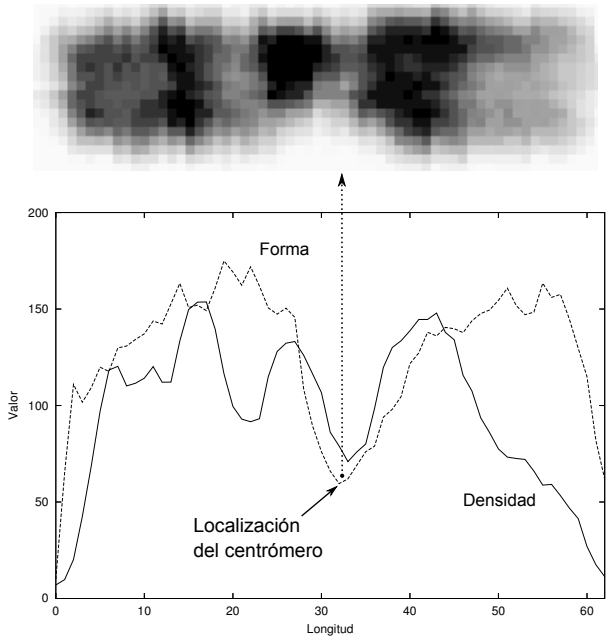



Figura 2.4: Cromosoma desdoblado de clase 1 de la célula *a4078-89* dispuesto de manera horizontal para la alineación con sus perfiles de densidad (línea continua) y forma (línea a trazos).

Los perfiles propuestos consisten, entonces, en un muestreo de niveles de gris sobre la cuerda, obteniendo grises representativos de más de un punto por cuerda. El muestreo se realiza sobre pixeles equiespaciados entre el inicio y el fin de cada cuerda, conformando un vector de características local. La mayor cantidad de información "perpendicular" que aporta este perfil permitiría capturar la variabilidad de tonalidades presentes en los cromosomas que tienen sus cromátidas diferenciadas,

Sobre cada píxel, la medida implementada consiste en la salida de un filtro de promediado bidimensional dado por:

$$p(x,y) = \frac{\sum_{s=-a}^{a} \sum_{t=-b}^{b} m_{s+a,t+b}\ c(x+s, y+t)}{\sum_{s=-a}^{a} \sum_{t=-b}^{b} m_{s+a,t+b}}, \tag{2.4}$$

donde

$a = b = 2$,

$m_{s,t}$: coeficientes de la máscara de filtrado con los siguientes valores:

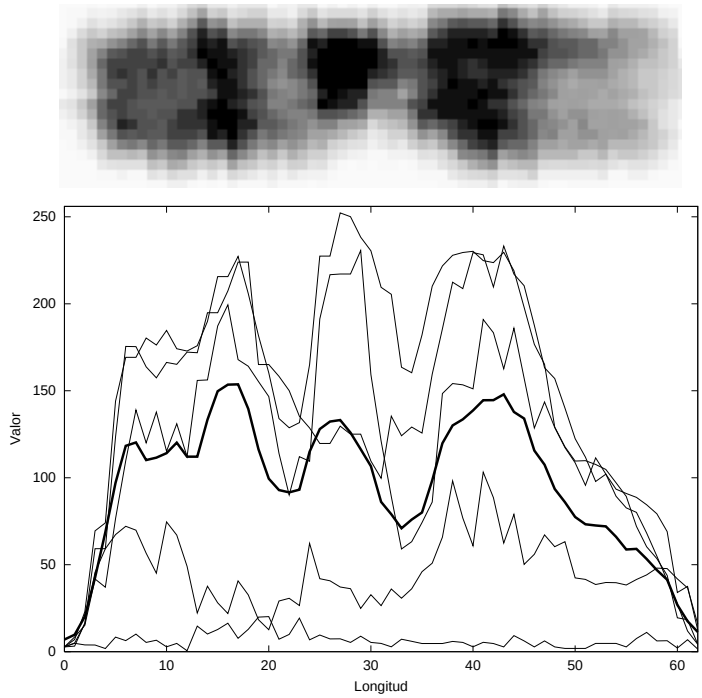



Figura 2.5: Cromosoma desdoblado de clase 1 de la célula *a4078-89* dispuesto de manera horizontal para la alineación con sus perfiles de densidad (línea gruesa) y muestreo de 5 puntos (líneas finas).

$$\mathbf{m} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 2 & 16 & 2 & 1 \\ 1 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Los pesos decrecientes desde el centro reducen el efecto de desenfoque introducido en el proceso de filtrado [41].

A fin de ilustrar la variabilidad de grises sobre las líneas paralelas al eje, información que aporta el perfil de muestreo y se está perdiendo en el promedio realizado por el perfil de densidad, se muestra en la Figura 2.5 un cromosoma de clase 1 desdoblado y alineado con sus perfiles de densidad (línea gruesa) y el perfil de 5 puntos (líneas finas).

En aplicaciones que trabajan sobre señales temporales, como por ejemplo en reconocimiento automático del habla, es usual la incorporación de pistas sobre los cambios dinámicos mediante el agregado de la derivada primera y

segunda al vector de características, lo cual mejora significativamente los resultados [47]. De igual manera, en aplicaciones de video se calculan derivadas espacio-temporales de la intensidad, las cuales proveen información relacionada al movimiento local que ayudan en la detección de objetos en la escena [48].

Tomando estas ideas, y de manera adicional a los grises muestreados sobre las cuerdas, en una segunda instancia de generación de patrones se proponen vectores extendidos con el agregado de las derivadas del perfil de muestreo. Estas magnitudes proveen información de cambio de la imagen y se calculan en ambas direcciones: derivada horizontal $\Delta_h$ sobre las cuerdas y derivada vertical $\Delta_v$ a lo largo de los perfiles de muestreo, de acuerdo a la definición de los coeficientes *delta* empleados en el campo del reconocimiento automático del habla [28]:

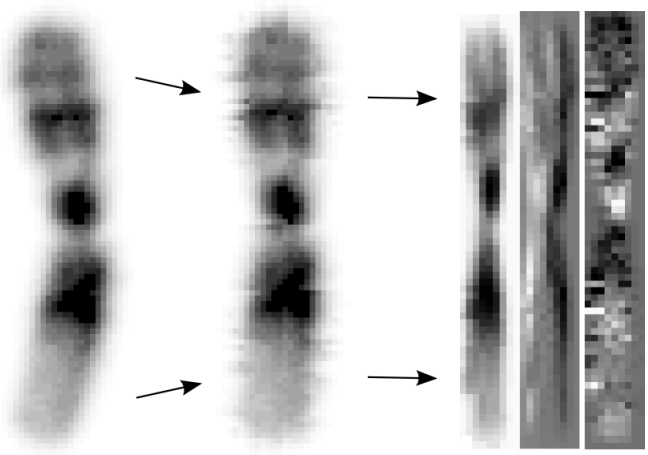
$$\Delta_h(i,j) = \frac{\sum_{\theta=1}^{\Theta} \theta(c(i,j+\theta) - c(i,j-\theta))}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2.5a}$$

$$\Delta_v(i,j) = \frac{\sum_{\theta=1}^{\Theta} \theta(c(i+\theta,j) - c(i-\theta,j))}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2.5b}$$

donde $1 \leq i \leq M$, $1 \leq j \leq N$ y $1 \leq \Theta \leq 2$ (para dar contexto a las derivadas).

A partir de estas definiciones es posible calcular además los coeficientes *delta-delta* $\Delta^2(i,j)$ o de aceleración, correspondientes a las derivadas segundas, por medio de la aplicación de las ecuaciones (2.5) a los coeficientes delta.

En la Figura 2.6 se muestran las imágenes resultantes del proceso completo de extracción de características: partiendo de la imagen original (a) se obtiene el cromosoma desdoblado (b) y luego sobre cada fila (correspondiente a una cuerda desdoblada) se muestra el resultado de un ejemplo particular de características: perfil de muestreo de 9 puntos por cuerda junto a las derivadas horizontales y verticales. En estos dos últimos conjuntos, los coeficientes delta toman valores positivos y negativos que se representan por grises claros y oscuros, respectivamente, con el cero dado por el valor de gris medio (127 para una imagen de 8 bits).



(a) (b) (c)

Figura 2.6: Ilustración de los pasos generales de la extracción de características: a) imagen original de un cromosoma de clase 1 de la célula *a117-90*, b) cromosoma desdoblado y c) perfil de muestreo de 9 puntos (izquierda) más el perfil de derivadas horizontales (centro) y derivadas verticales (derecha) .

## 2.7 Conjuntos de características seleccionadas

A partir de las medidas presentadas, es posible conformar diversos conjuntos de características:

- Perfil de densidad: unidimensional.
- Perfil de densidad+gradiente+forma.
- Perfil de muestreo, variando la cantidad de puntos por cuerda.
- Variaciones sobre los perfiles de muestreo.

En los perfiles de muestreo, se establece un valor mínimo de 3 puntos a fin de capturar los grises sobre ambas cromátidas más el nivel de gris en el espacio intermedio, en caso de que fuera visible. El máximo de puntos puede ser fijado arbitrariamente en función del ancho promedio de los cromosomas (20 pixeles), a fin de no generar un sobremuestreo de los grises teniendo en cuenta el tamaño de la máscara **m** empleada en el muestreo.

Es posible, además, combinar los perfiles de muestreo de grises y los coeficientes delta y aceleración en ambos sentidos para conformar conjuntos alternativos de características. Por ejemplo: grises+$\Delta_h$, grises+$\Delta_h$+$\Delta_h^2$, grises+$\Delta_h$+$\Delta_v$, $\Delta_h$+$\Delta_v$ (sin grises), etc.

De esta manera, en la experimentación se probaron diferentes conjuntos de características compuestos por combinaciones de estas medidas, tratando de encontrar una representación local que capture adecuadamente las variabilidades en los niveles de gris de las bandas.

## 2.8 Comentarios de cierre del capítulo

En este capítulo se presentaron las técnicas empleadas en el preprocesamiento y extracción de características de las imágenes de cromosomas. Se expusieron los nuevos conjuntos de características propuestos que se obtienen al realizar el tratamiento símil temporal de las imágenes desdobladas, recorriéndolas sobre su eje.

En el capítulo siguiente se detallan los experimentos de clasificación y resultados obtenidos en el reconocimiento de cromosomas aislados y en contexto celular.

CAPÍTULO 3

# Clasificación de cromosomas y cariotipado automático

En este Capítulo se reseñan, en primer lugar, las particularidades de los subconjuntos de entrenamiento, validación y prueba que fueron empleados en los experimentos. A continuación se detallan las pruebas realizadas con ambos tipos de modelos, las redes neuronales y los modelos ocultos de Markov. La clasificación obtenida se considera *aislada* en el sentido de que cada objeto es etiquetado de manera independiente.

Luego se expone la fundamentación de un método de contextualización de los resultados previos al número esperado de cromosomas por célula. Se emula, así, el trabajo del citogenetista en la rutina clínica, quien clasifica todos los cromosomas en su conjunto.

## 3.1 Consideraciones preliminares

La implementación de los algoritmos de entrenamiento y prueba con redes neuronales se realizó mediante la herramienta SNNS (Stuttgart Neural Network Simulator), un simulador software de redes neuronales desarrollado en la Universidad de Stuttgart. La herramienta provee un ambiente eficiente y flexible para el diseño y prueba tanto del Perceptrón multicapa como de las redes neuronales parcialmente recurrentes [49].

#### Métrica de evaluación del desempeño

En todos los experimentos, el desempeño en la clasificación fue evaluado a través de la tasa de *error de clasificación* (EC), considerando el error sobre un conjunto de prueba dado por $\mathrm{EC} = w/(w+r) \times 100\,\%$, donde $w$ es el número de patrones mal clasificados y $r$ es el número de patrones correctamente clasificados.

## 3.2 Marco experimental con Perceptrón multicapa

La arquitectura y funcionamiento de un PMC involucra fijar el número de nodos de entrada *a priori*. Por esta razón, los patrones de entrenamiento deben ser de longitud fija. Para la tarea de interés, sin embargo, este punto representa un inconveniente ya que la longitud promedio de los cromosomas es dependiente de la clase. Una manera convencional de sortear esta limitación consiste en definir una capa de entrada con suficientes nodos para contener al patrón más largo del corpus, y luego extender con entradas ceros al resto de los patrones. Esta aproximación, no obstante, presenta la desventaja de aumentar los requerimientos computacionales de tiempo para el entrenamiento y memoria para almacenar la gran cantidad de pesos que conforman la red.

Como alternativa al procedimiento descripto anteriormente, en la experimentación se aplicó una técnica de interpolación denominada *Segmentación de Traza* [50], la cual ajusta el largo de todos los patrones al valor de longitud promedio de los patrones presentes en la partición de entrenamiento. Luego,

este mismo valor es el que se fija para la cantidad de nodos de la capa de entrada del PMC.

El PMC reportado consiste en una red neuronal anteroalimentada de una capa oculta. El número de unidades en esta capa fue variable y objeto de ajuste en la experimentación.

### 3.2.1 Diseño de los experimentos

#### Conjuntos de características

En los experimentos con PMC se emplearon dos conjuntos de características basados en los perfiles clásicos [43]:

1. *Perfil de densidad*: vectores que contienen el valor de gris promedio sobre la cuerda calculado según la ecuación (2.1).

2. *Perfil de densidad+gradiente+forma*: vectores que contienen –en forma consecutiva– la característica anterior más su derivada primera y el perfil de forma, estas dos últimas calculadas mediante las ecuaciones (2.2) y (2.3) respectivamente.

#### Partición del corpus

En los experimentos iniciales para ajuste de arquitectura, el corpus completo de 2800 células fue dividido de la siguiente manera:

- una partición de 2400 células de las cuales se extrajeron 3000 patrones de entrenamiento aleatoriamente de entre 740 células, a fin de conformar un conjunto de entrenamiento de tamaño medio que fuera adecuado para experimentación exhaustiva;

- un conjunto de validación de 2000 cromosomas no presentes en la partición de entrenamiento; y

- un conjunto de prueba de la misma cantidad anterior, sin repetir patrones previos.

### 3.2.2 Experimentos y resultados

La aplicación de la interpolación mediante segmentación de traza determinó una longitud promedio de 44 puntos. El conjunto de características 1, por lo tanto, consiste en patrones representando un perfil de densidad con 44 valores, mientras que el conjunto de características 2 consta en total de 132 datos por patrón. Desde el punto de vista del reconocimiento de patrones se considera aquí a los cromosomas sexuales X e Y como dos clases informacionales diferentes, lo que hace un total de 24 clases. La información de la salida deseada, entonces, consiste en un valor fijado a 1 en la posición que representa a la clase del cromosoma y 23 valores fijados a 0 en las restantes posiciones.

La función de aprendizaje utilizada fue el algoritmo de retropropagación con término de momento, como fuera introducido en la Sección 1.2.3. El conjunto de los parámetros empleados tiene los siguientes valores:

- Inicialización de pesos de forma aleatoria en el intervalo $[-1, 1]$.
- Parámetro de aprendizaje $\eta = 0{,}2$.
- Término de momento $\mu = 0{,}1$.
- Valor de salteo de zonas planas en la curva de error $c = 0$.
- Máxima diferencia $d_j = t_j - o_j$ tolerada entre una salida deseada $t_j$ y la salida de la red $o_j$ para el $j$-ésimo nodo de salida: $d_{max} = 0{,}1$.
  Esta diferencia implica que valores superiores a 0,9 fueron tratados como 1, mientras que valores inferiores a 0,1 fueron tratados como 0. La elección de estos umbrales previene el sobreentrenamiento de la red.

En todos los experimentos, el proceso de entrenamiento fue detenido en el pico de generalización medido respecto al conjunto de patrones de validación. El error reportado se obtuvo mediante el método de estimación por conjunto de prueba (Sección 1.4).

#### Resultados con perfiles de densidad

La topología del PMC empleado fue la siguiente:

- 44 unidades en la capa de entrada.

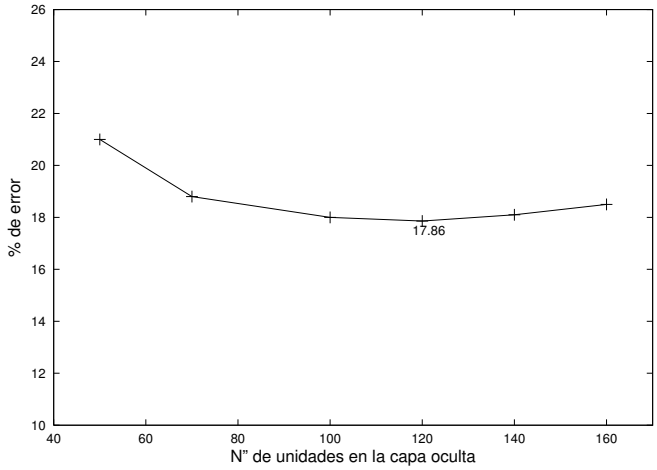



Figura 3.1: Evolución del error en la clasificación de patrones de densidad en función del número de nodos en la capa oculta.

- 1 capa oculta de número variable de unidades.
- 24 unidades en la capa de salida.

La Figura 3.1 resume los resultados obtenidos por el PMC en la clasificación de patrones de densidad, al variar la cantidad de nodos en la capa oculta. Se observa que el mejor desempeño se obtiene para la topología 44/120/24, con la cual se alcanza un valor de EC mínimo.

### Resultados con perfiles de densidad, gradiente y forma

Para esta serie de experimentos, la topología del PMC se detalla a continuación:

- 132 unidades en la capa de entrada.
- 1 capa oculta de número variable de unidades.
- 24 unidades en la capa de salida.

La Figura 3.2 muestra los valores de error obtenidos en clasificación al variar el número de unidades en la capa oculta, donde el mejor desempeño se obtiene con la topología 132/100/24.

La Tabla 3.1 resume el mejor desempeño obtenido con los 2 conjuntos de perfiles clásicos en la aproximación mediante PMC. La carga computacional, en ambos casos, prácticamente es la misma.

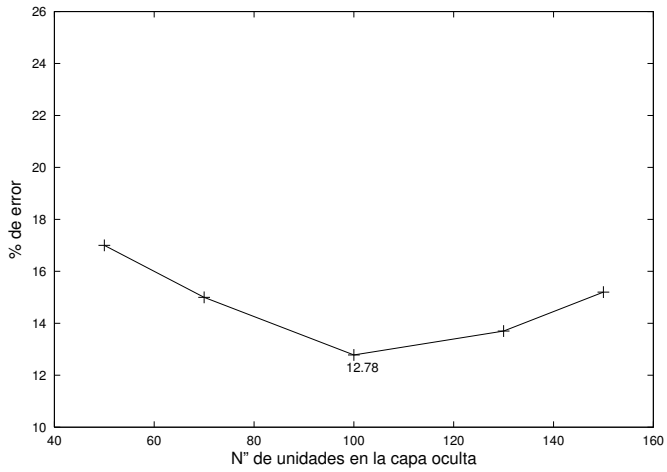



Figura 3.2: Evolución del error en clasificación de patrones de densidad+gradiente+forma

Tabla 3.1: Comparación de desempeños para Perceptrón multicapa.

| Característica | Tasa de error en % |
|---|---|
| Densidad | 17,86 % |
| Densidad+Gradiente+Forma | 12,78 % |

## 3.3 Marco experimental con redes de Elman

### 3.3.1 Aspectos de funcionamiento y operación

#### Conformación de la entrada

Las redes de Elman (RE) tienen la capacidad de recorrer los patrones por ventanas de longitud fija, alojando en la capa de entrada en cada instante de tiempo uno o más vectores de características. De esta manera, no es necesaria la aplicación de la Segmentación de Traza para unificar la longitud de todos los patrones, poniéndose en juego ahora también la variabilidad de longitud entre clases.

La capa de entrada puede ser alimentada no sólo con la cuerda actual-

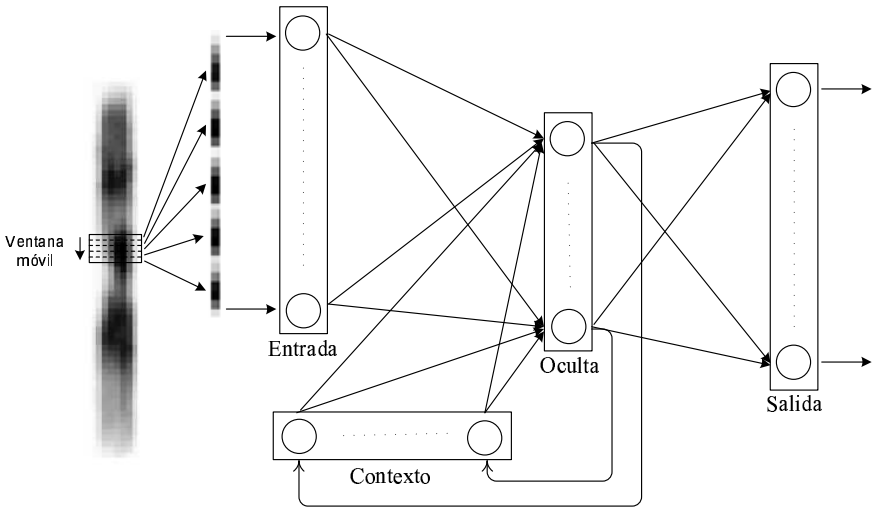



Figura 3.3: Configuración empleada para la clasificación de cromosomas mediante redes de Elman.

mente en proceso, sino también con una pequeña porción del patrón de bandas constituída por un número de cuerdas consecutivas, de manera de incorporar un contexto local del patrón en la operación de la misma. En adelante, esta porción del patrón será denominada *ventana móvil* sobre la imagen. Esta ventana móvil se construye reacomodando las cuerdas de la porción de manera lineal, esto es, se forma un vector de grises donde, por ej., los índices 1 a 5 de la ventana contienen la primer cuerda, los índices 5 a 10 contienen la segunda cuerda, y así sucesivamente. El comportamiento "temporal" del procesamiento del patrón de entrada es asemejado por el movimiento deslizante de la ventana móvil: una vez que una ventana ha sido procesada, una nueva ventana es tomada de las cuerdas siguientes. La Figura 3.3 ilustra el uso de las redes de Elman para la clasificación de cromosomas, como fueron empleadas en esta tesis. La ventana móvil se muestra con las cuerdas rotadas 90º con respecto al eje medio del cromosoma, conformando el vector de entrada a la red.

### Entrenamiento

Las activaciones de los nodos ocultos son copiadas a través de los enlaces con pesos fijos a las unidades de contexto de forma uno-a-uno. En el instante de tiempo siguiente, las unidades de contexto realimentan a las unidades ocultas de forma distribuida.

Si se eliminan todos los enlaces recurrentes en una red parcialmente recurrente, se obtiene una red anteroalimentada simple y las unidades de contexto tienen ahora la función de unidades de entrada, de manera que el algoritmo de retropropagación puede ser modificado para el entrenamiento. De esta manera, la función de aprendizaje de SNNS `JE_BPMomentum` fue utilizada, la cual constituye el algoritmo estándar de retropropagación con término de momento para las redes neuronales recurrentes. Asimismo, la función de actualización provista por SNNS es `JE_Order`, la cual propaga un patrón desde la capa de entrada a las capas ocultas (en orden, si hubiera más de una) y luego a la capa de salida, seguido de una actualización sincrónica de todas las unidades de contexto.

Al finalizar la presentación de cada patrón, las activaciones de los nodos de contexto son puestos a cero, para no interferir en el aprendizaje del siguiente patrón.

### Obtención de la salida

La salida que se obtiene de la RE en clasificación difiere en su forma de la salida que se obtiene del PMC. En el PMC, por cada patrón de entrada, se obtiene una sola distribución de probabilidades, y por lo tanto, un único valor de salida que se compara con la salida deseada para obtener la tasa de error. En la RE, por el contrario, cada patrón del conjunto de prueba es analizado por segmentos con la ventana de la capa de entrada, obteniéndose una distribución de probabilidades en la capa de salida para cada ventana de análisis. Por lo tanto, debe proponerse un método para decidir cuál es la clase del patrón de entrada (luego del análisis del patrón completo), y hacer posible la comparación con la salida deseada a fin de obtener el error para el patrón.

Para obtener la clasificación que realizan las redes se propusieron dos métodos: el método de cota mínima y el método del promedio. A continuación se explican las particularidades de cada uno de ellos.

El método de cota mínima fija un porcentaje `MIN` a priori de vectores individuales correctamente clasificados, y un cromosoma se dice correctamente clasificado si más del `MIN` por ciento de los vectores fueron correctamente clasificados de acuerdo a la salida deseada. Si, por el contrario, el porcentaje de vectores correctamente clasificados no supera el valor de `MIN`, el cromosoma se cuenta como error en clasificación. El valor de `MIN` se fijó en 70 % [51].

El método del promedio, propuesto posteriormente, guarda en una tabla

la distribución de probabilidad en la salida para cada ventana de análisis en la capa de entrada. Al finalizar la Perceptrón del patrón, se promedian las probabilidades de cada clase considerando la longitud del patrón, y se decide la clase del cromosoma en base a la máxima probabilidad resultante, sin establecer cota mínima. Este método refleja más fielmente la tasa de error real de la RNR, ya que equivale a encontrar la clase con máxima verosimilitud [52].

### 3.3.2 Experimentos y resultados

A fin de implementar un análisis local de los grises de las bandas del cromosoma, en esta aproximación se utilizó como entrada el perfil de muestreo de grises. El número de puntos por cuerda se fijó en 9, ya que constituye un término medio teniendo en cuenta el ancho medio de los cromosomas.

Las redes de Elman implementadas tienen la siguiente topología:

- 1 capa de entrada de 45 unidades que aloja a una ventana de vectores consecutivos de características en cada instante de tiempo. El número de vectores que conforma la entrada se determinó mediante experimentación preliminar y fue fijado en 5 para el resto de la experimentación.
- 1 capa oculta, junto a su capa de contexto (número variable de unidades). Los pesos de los enlaces autorecurrentes en las unidades de contexto fueron fijados a 0. Los pesos de los restantes enlaces recurrentes, de las unidades ocultas a las unidades de contexto, fueron fijados a 1 [25].
- 1 capa de salida de 24 unidades, sin capa de contexto.

Diferentes configuraciones de los parámetros de inicialización y aprendizaje fueron extensamente probados, y los mejores desempeños se obtienen para la siguiente configuración, con la cual se reportan los resultados:

- Inicialización de pesos aleatoria, en conexiones entrenables, dentro del intervalo $[-1, 1]$.
- Activación inicial de las unidades de contexto fijada a 0.
- Parámetro de aprendizaje $\eta = 0{,}2$.
- Término de momento $\mu = 0{,}1$.

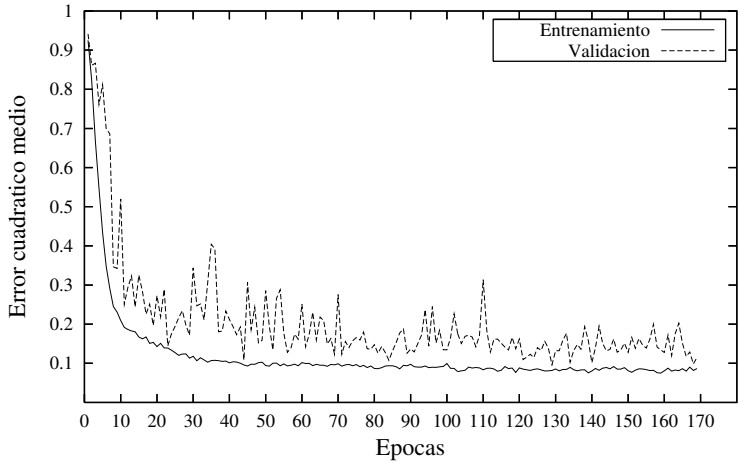



Figura 3.4: Curvas de evolución del error de entrenamiento y prueba en una red de Elman.

- Valor de salteo de zonas planas en la curva de error $c = 0$.
- Máxima diferencia tolerada entre una salida deseada y la salida de la red $d_{max} = 0{,}1$

El número de unidades en la capa oculta fue variado desde 50 a 250 (con el mismo número de unidades en la capa de contexto). En la Figura 3.4 se observa un ejemplo de evolución de las curvas de error cuadrático medio durante el entrenamiento, con las típicas formas de descenso asintótico a medida que progresa el aprendizaje.

La Tabla 3.2 lista los resultados obtenidos para diferentes configuraciones de capa oculta, donde el mínimo error en clasificación es alcanzado por una red con 200 nodos en la capa oculta.

El conjunto de prueba se aplicó a la configuración 45/200/24, que logra el pico de generalización para el modelo. La clasificación por el método de cota fija logra una tasa de EC de 5,63 % para las 24 clases. El error de clase individual mínimo fue de 3,1 %, obtenido para la clase 3. El error máximo para una clase fue de 40,0 %, correspondiente a la clase 24 (cromosoma sexual Y). Este es un resultado esperado dada la baja probabilidad a priori de este cromosoma : 1/92 vs. 3,92 para la clase 23 (cromosoma sexual $X$, presente en células masculinas y femeninas), y 4/92 para los autosomas (cromosomas de clase 1 a 22).

El experimento con la base de datos completa se realiza en dos etapas. La primera consiste en la clasificación de cromosomas de manera independiente

Tabla 3.2: Resultados en clasificación de perfiles de muestreo (9 puntos por cuerda) para diferentes configuraciones de la capa oculta.

| Número de nodos | Tasa de EC |
|---|---|
| 50 | 22,8 % |
| 100 | 15,3 % |
| 150 | 9,5 % |
| 200 | 5,7 % |
| 250 | 6,7 % |

Tabla 3.3: Resultados para el experimento con validación cruzada no-contextual con redes de Elman.

| Partición | Tasa de EC |
|---|---|
| 1 | 5,92 % |
| 2 | 6,72 % |
| 3 | 6,03 % |
| 4 | 8,29 % |
| 5 | 7,41 % |
| 6 | 5,63 % |
| 7 | 6,64 % |
| Media | 6,67 % |

del contexto, tal como se obtiene en la salida de la red ante la presentación de cada patrón. La segunda etapa consiste en la restricción a clasificación contextual y será expuesta en la Sección 3.5.

En la etapa de clasificación no-contextual, una estrategia de validación cruzada en 7 bloques fue aplicada, con el objetivo de reducir el desvío en la estimación del desempeño del clasificador. Cada partición consistió en 400 células para prueba, mientras que las restantes 2400 células fueron usadas para el conjunto de entrenamiento. El costo computacional del entrenamiento para cada partición fue aproximadamente de 600 horas de CPU sobre una PC Pentium IV con 1 Gb RAM.

La Tabla 3.3 resume los resultados obtenidos para cada una de las 7 particiones del experimento completo no-contextual. El error final fue de 6,7 %, calculado como el EC medio entre todas las particiones.

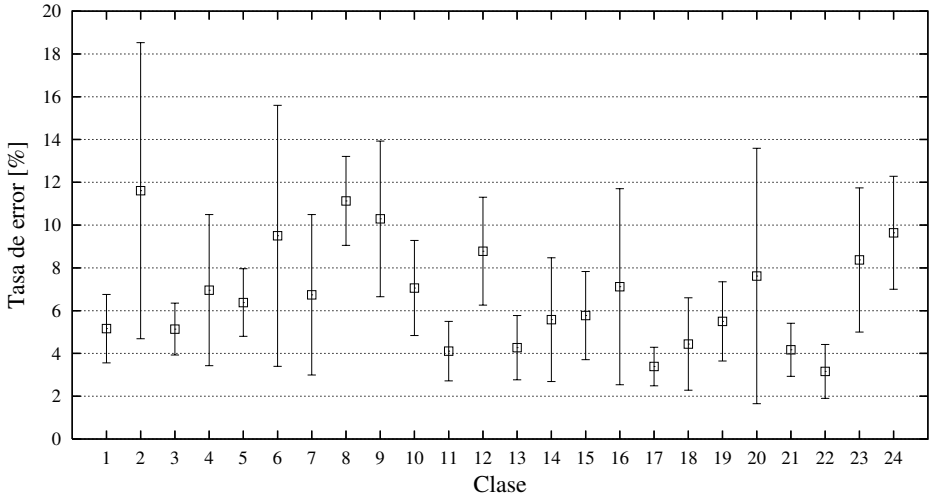



Figura 3.5: Error medio y desvío estándar (marca central y extremos) para cada clase de cromosoma, como fueron obtenidos para las 7 particiones del experimento de validación cruzada independiente del contexto.

La Figura 3.5 resume algunas estadísticas sobre los valores de EC obtenidos por bloque. Para cada clase, las barras de error indican la media aritmética del EC (marca central) y el desvío estándar (marcas finales) de las 7 particiones. Esta figura muestra claramente la dificultad natural de la tarea de clasificación de cromosomas para la red neuronal, donde algunas clases (por ejemplo, 17 y 22) son considerablemente mejor estimadas que otras (por ejemplo, 2 y 20). Además, puede observarse la alta dispersión en los resultados entre particiones para algunas clases, donde se alcanzan valores de EC por encima de 20 %.

### 3.3.3 Discusión de resultados

En primer lugar se experimentó un modelo estándar reportado en otros trabajos –el Perceptrón multicapa– con los perfiles clásicos. El mejor resultado obtenido para los patrones de densidad fue una tasa de error de 17,86 % sobre la clasificación de los 2000 cromosomas del conjunto de prueba, la cual fue mejorada agregando la información de gradiente de densidad y la forma del cromosoma. En este caso, el PMC logró una tasa de error de 12,78 %.

Este mejor desempeño en clasificación se logra incluso con menor cantidad

de unidades en la capa oculta (100 unidades vs. 120 unidades en el primer caso), debido a la mayor información contenida en los patrones, por ej., la información de la localización centromérica contenida en el mínimo global del perfil de forma. Además, el agregado de información mejora la tasa de error a pesar de los mayores costos computacionales de entrenamiento y la mayor cantidad de parámetros a estimar (15600 vs. 8160 del primer caso).

Las redes de Elman logran una reducción significativa en las tasas de error del PMC, trabajando solamente con información de grises. Así, se demostró que el empleo de redes recurrentes es más adecuado para la tarea que la aproximación anterior por el tratamiento dinámico que efectúan a lo largo de los patrones. Esta forma de operación les permite sortear el inconveniente que presentan las redes estáticas con el procesamiento de patrones de longitud variable, a costa de un aumento en el costo computacional y de cantidad de pesos a estimar (54000).

Más allá de que las redes de Elman capturan la variabilidad temporal de los eventos con su capa de realimentación interna, se demostró también que cuando se alimenta a estos modelos con una ventana móvil de unos pocos vectores de características los resultados son mejorados, ya que se provee a la red con un contexto espacial del patrón.

## 3.4 Marco experimental con modelos ocultos de Markov

En el Capítulo 1 se introdujeron los conceptos y formulación matemática de los modelos ocultos de Markov (MOM), en particular de la aproximación continua tratada en esta tesis. A continuación se exponen las consideraciones de índole conceptual y práctica supuestas para la implementación.

Para la experimentación se utilizó la herramienta HTK [53], un paquete software que permite construir y entrenar modelos ocultos de Markov (MOM). Aunque el HTK está diseñado principalmente para aplicaciones en reconocimiento de habla, es posible modelar cualquier serie temporal, por lo cual se considera una herramienta de propósito general.

El clasificador basado en MOM consta de 24 modelos: uno por cada clase de los autosomas, numerados del 1 a 22, más un modelo para el cromosoma sexual $X$ y otro modelo para el cromosoma sexual $Y$. Los cromosomas son modelados por medio de MOM continuos con topología izquierda-derecha,

permitiendo solamente las transiciones al estado consecutivo y las realimentaciones al mismo estado. Esta arquitectura permite modelar la sucesión, a lo largo del eje longitudinal, de vectores de características extraídos a partir de ejemplos de los cromosomas de esa clase.

Un esquema gráfico de los modelos se presenta en la Figura 3.6. Aquí es posible observar un patrón genérico de grises y derivadas (explicado en detalle más adelante) en la parte superior, con una correspondencia a un segmento de un MOM. Con una serie de ventanas marcadas sobre el patrón, se muestra conceptualmente la porción del mismo que tiene máxima verosimilitud por cada estado.

La asociación entre secuencia de vectores y estados también puede entenderse si se ve al modelo entrenado como generador de porciones del patrón. Cada estado del MOM genera vectores de características de acuerdo a una ley de probabilidad, típicamente una mezcla de densidades Gaussianas. El número de estados y el número de densidades por estado que son apropiados para modelar cada clase dependen de la variabilidad de la clase y la cantidad disponible de datos de entrenamiento, siendo parámetros que necesitan un ajuste empírico.

Una consideración importante debe ser realizada en cuanto a la arquitectura de los modelos. Por un lado, el modelado de los cromosomas se realiza con un único modelo por clase, sin considerar una subdivisión en modelos más cortos. De este modo, se espera que cada modelo capture mediante su arquitectura de estados y transiciones, la secuencia particular de bandas alternantes que caracteriza a los cromosomas de cada clase. Por otro lado, las clases del cariotipo están ordenadas por las características morfológicas y tamaño de los cromosomas, siendo fácilmente visibles grandes diferencias en la longitud entre clases. Del análisis se desprende que el número de estados para cada modelo no será el mismo. A fin de tener en consideración la longitud media de las clases, se introduce el siguiente cálculo de la cantidad de estados emisores para cada MOM:

$$s_i = \frac{f_i}{k}, \tag{3.1}$$

donde

$s_i$: número de estados del MOM,

$f_i$: longitud media de la secuencia de vectores de características utilizados para entrenar el modelo $M_i$,

$k$: número promedio de vectores de características modelados por estado, valor que será denominado *factor de carga de estado*.

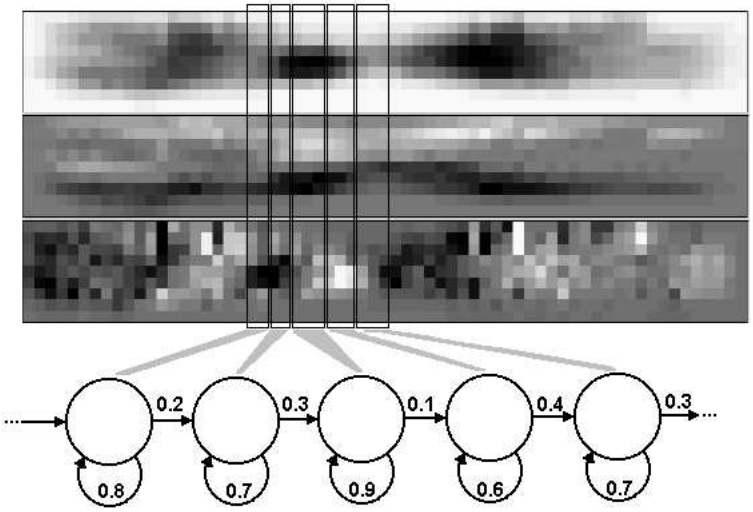



Figura 3.6: Ejemplo de la topología izquierda-derecha de los MOM. Se observa un segmento del modelo junto a la representación de un patrón y una secuencia de ventanas de vectores con máxima verosimilitud para un estado particular.

Esta manera de calcular el número de estados intenta balancear los esfuerzos de modelado entre estados a la vez que el modelo captura información discriminativa importante sobre la longitud típica de cada clase.

En la generación de los modelos de cada clase, primeramente se crean los *prototipos*: definiciones de la arquitectura (número de estados), cantidad de distribuciones Gaussianas en la mezcla de cada estado (idéntica para todos los estados) y los valores entrenables –no cero– de la matriz de probabilidades de transición, que fijan las transiciones permitidas en el modelo.

Los prototipos son inicializados siguiendo el diagrama de flujo que muestra la Figura 3.7, el cual es aplicado para cada clase. En el $i-$ésimo modelo $M_i$, el paso de *segmentación uniforme* consiste en la división de los patrones en $s_i$ porciones de aproximadamente la misma longitud. Estos segmentos son asignados a los estados del modelo de manera consecutiva, cada uno de los cuales es empleado en la *inicialización* de los parámetros. Un esquema iterativo de refinamiento es luego aplicado, donde el paso fundamental está dado por el bloque de *segmentación por Viterbi*, el cual encuentra la secuencia de estados más probable para cada patrón de entrenamiento. Con esta nueva asignación de estados, los parámetros del modelo son actualizados y el ciclo se repite hasta satisfacer un criterio de convergencia.

Una vez inicializados los modelos se lleva a cabo la parte más importante

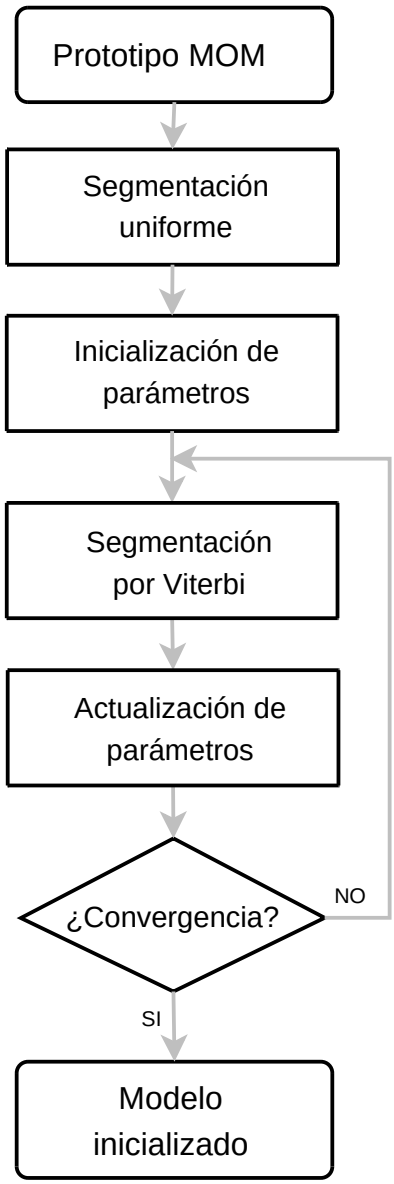



Figura 3.7: Diagrama de flujo de la inicialización de los modelos. La herramienta HTK provee el programa HInit con el cual se realiza este proceso [53].

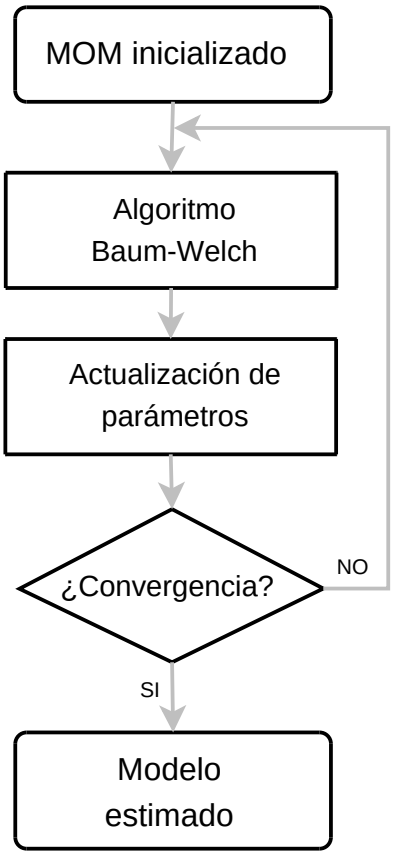



Figura 3.8: Diagrama de la reestimación de los parámetros, llevado a cabo mediante el programa HRest de HTK [53].

de todo el entrenamiento, consistente en la reestimación de los parámetros, como muestra la Figura 3.8. Es también el proceso más costoso computacionalmente, ya que básicamente es similar a la iteración en el proceso de inicialización, pero cambiando la estimación por Viterbi por el algoritmo de Baum-Welch. Este esquema se aplica directamente a los prototipos inicializados o, como se verá en la etapa de experimentos, a nuevos prototipos que resultan de incrementar el número de gaussianas en cada mezcla. Esta estrategia responde a la filosofía de HTK de construir los sistemas incrementalmente: cada paso involucra aumentar el número de parámetros y reestimarlos.

En clasificación, el sistema completo consta de los 24 modelos que analizan cada uno el patrón de entrada desconocido siguiendo el esquema conceptual expuesto en la Figura 3.9. Cada modelo obtiene, por máxima verosimilitud, la probabilidad de que la secuencia haya sigo generada por ese modelo. En base a estos valores, la decisión final sobre la etiqueta de clase que se asocia

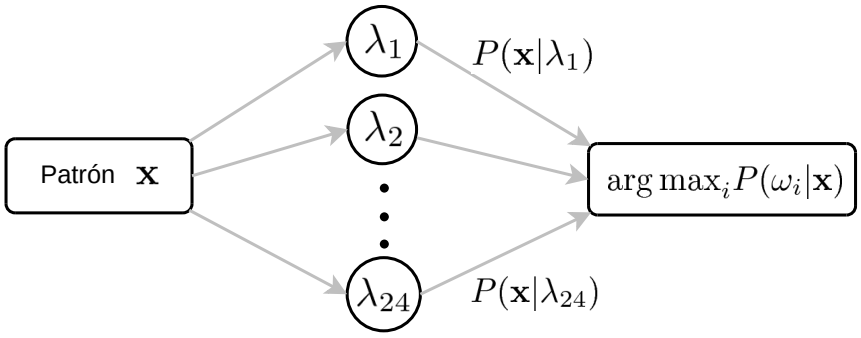



Figura 3.9: Esquema de la clasificación mediante MOM. El patrón de entrada es analizado por todos los modelos, y la decisión sobre la clase asignada se realiza por máxima verosimilitud. La herramienta HTK dispone del programa HVite, el cual implementa el cálculo de la probabilidad de Viterbi [53].

al patrón es tomada por un operador máximo, como fuera expuesto en el Capítulo 2.

### 3.4.1 Diseño de los experimentos

#### Conjuntos de características

Se diseñaron una serie de conjuntos de características, los primeros de ellos con los perfiles clásicos reportados en [43] para comparación con los modelos conexionistas, y los últimos con la propuesta de muestreo de grises y algunas variantes relacionadas:

- Perfil de densidad: vectores unidimensionales conteniendo el valor de gris promedio sobre la cuerda según (2.1).
- Perfil de densidad+gradiente+forma: vectores de dimensión 3, considerando la característica anterior más su vector gradiente y el perfil de forma según (2.2) y (2.3), respectivamente.
- Perfiles de muestreo: construídos con 3, 5 y 9 puntos por cuerda.
- Perfiles de muestreo más derivadas: corresponde a la versión previa de perfiles, con el agregado de las derivadas horizontales y/o verticales como fueran calculadas en el Capítulo 2.
- Variantes sobre los perfiles de muestreo, con ventanas que alojan a un subconjunto de vectores de características por vez.

### Aspectos de implementación

A continuación se exponen generalidades de la implementación, válidas para todos los experimentos:

- El número de distribuciones gaussianas por estado se fijó para los prototipos iniciales igual a 1. Para aumentar la complejidad de los modelos, se crearon mezclas de gaussianas en cada estado, con un número sucesivo de distribuciones en potencia de 2. En cada experimento se fijó el mismo número de gaussianas por mezcla para todos los estados de los 24 modelos.

- La inicialización se fijó a un número máximo de 100 iteraciones de la segmentación por Viterbi, con un factor de convergencia $e = 0{,}0001$, esto es, cambio relativo entre valores sucesivos de $P_{max}(O|\lambda)$. Los experimentos lograron la convergencia, en todos los casos, entre las 20 y las 60 iteraciones.

- La reestimación por Baum-Welch se fijó a un número máximo de iteraciones, diferente según la carga del experimento.

- Los valores de factor de carga $k$ cercanos a la unidad producen MOM con igual cantidad de estados que el promedio de longitud de las secuencias de características para ese modelo. Esta situación implica descartar para la inicialización, una gran cantidad de secuencias cuya longitud es menor a la cantidad de estados de los MOM, lo cual genera una condición de error en la segmentación uniforme. En todos los modelos se encuentra que el valor $k = 1{,}5$ genera una cantidad de estados igual o superior a la longitud mínima de los patrones, y por lo tanto se toma como cota inferior del parámetro.

- La tasa de error en clasificación (EC) porcentual reportada para un clasificador basado en MOM se calcula como

$$\mathrm{EC} = 100 \cdot \frac{\sum_{i=1}^{24} w_i/(w_i + r_i)}{24}, \tag{3.2}$$

  donde $w_i$ es la cantidad de patrones de prueba de la clase $\omega_i$ correctamente clasificados y $r_i$ la cantidad de patrones incorrectamente etiquetados. Esta expresión resulta similar a la dada por la ecuación (1.60).

Tabla 3.4: Resultados para perfiles clásicos con MOM de una distribución gaussiana por estado.

| Partición | Perfil | EC (en %) |
|---|---|---|
| 400/42 | Densidad | 80 % |
| | Densidad+gradiente+forma | 47 % |
| 2400/400 | Densidad | 60 % |
| | Densidad+gradiente+forma | 32 % |

### 3.4.2 Experimentos con perfiles clásicos

Con las características de perfiles de densidad, gradiente y forma de [43] se realizó la primer serie de experimentos, utilizando las siguientes particiones del corpus Cpa:

- Partición 1: 400 células para entrenamiento (18400 patrones) y 42 células para prueba (1839 patrones).
- Partición 2: 2400 células para entrenamiento (110300 patrones) y 400 células para prueba, utilizándose en este caso el corpus completo.

En estos experimentos iniciales se probaron MOM con una distribución gaussiana por estado, cuyos resultados se resumen en la Tabla 3.4. Se puede observar una reducción significativa aproximadamente a la mitad en el error de clasificación cuando se agrega información a los patrones de densidad. La partición 2400/400 del corpus completo fue conformada a fin de mejorar los resultados de la partición más pequeña, ya que un EC de 80 % es muy alto todavía comparado al 17 % obtenido con el Perceptrón multicapa.

Posteriormente se evaluaron sistemas en donde se aumentó la cantidad de distribuciones gaussianas por mezcla. Sin embargo, los resultados obtenidos no mejoraron los previamente reportados, incluso se obtuvieron tasas de error ligeramente más altas.

Para estudiar este fenómeno, se debe notar que empleando el conjunto de perfiles de densidad y una única gaussiana por estado, la probabilidad de emisión consistirá justamente en el gris medio estimado a partir de las porciones del vector de características más probables para cada estado. Luego de la clasificación, y como subproducto del algoritmo de Viterbi, se tiene el

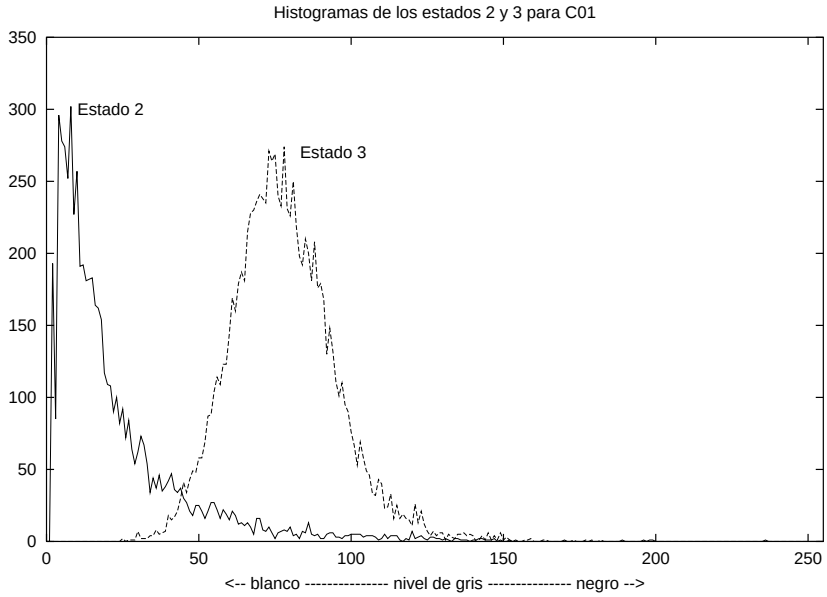



Figura 3.10: Histograma para los dos primeros estados emisores (estado 2 y 3) de un MOM de clase 1, denotado como C01. Nótese que el nivel de gris en el eje de abscisas sigue la convención: 0=blanco, 255=negro.

camino más probable para las secuencias. A partir de esta información se calcula el histograma para los dos primeros estados emisores[1].

La Figura 3.10 muestra los histogramas calculados sobre 1000 células de prueba, donde se evidencia la naturaleza de la distribución de los valores de gris aprendidos en cada estado. El primer estado emisor (Estado 2) corresponde al inicio del cromosoma, por lo que consta de un valor medio cercano al gris máximo, con una distribución que se asemeja a una Laplaciana. En el caso del otro estado emisor graficado (Estado 3), se observa una distribución típicamente Gaussiana, correspondiente a la primera banda –más oscura– del cromosoma.

El análisis explica los resultados insatisfactorios obtenidos con más de una distribución gaussiana por estado, ya que las distribuciones se solapan en cada mezcla. Esta superposición de gaussianas da lugar a valores medios y matrices de covarianza similares, pero con una mayor cantidad de parámetros a estimar.

[1]Correspondientes a los estados 2 y 3 de los MOM implementados, ya que HTK considera dos estados extra no-emisores al inicio y al final.

Tabla 3.5: Ajuste del factor de carga para fijar la longitud de los MOM.

| Factor $k$ | 3 puntos | 5 puntos |
|---|---|---|
| 1,5 | 29 % | 31 % |
| 2 | 42 % | 40 % |
| 3 | 49 % | 52 % |

### 3.4.3 Experimentos con perfiles de muestreo

Con el objetivo de mejorar estos resultados y demostrar las ventajas de la representación basada en muestreo de grises, la experimentación exhaustiva de ajuste de arquitectura y características fue realizada utilizando la partición 1 anterior.

Una vez ajustado el clasificador para un desempeño óptimo, los experimentos finales serán realizados siguiendo el esquema de validación cruzada en 7 bloques con el corpus completo.

Los datos empleados en todas las pruebas corresponden a las mismas particiones aplicadas en la experimentación con modelos conexionistas, como fueran detallados en la sección 3.2.1.

### 3.4.4 Estudio del factor de carga

Los primeros experimentos fueron realizados con 1 distribución gaussiana por estado, y se probaron los valores de factor de carga $k$ entre 1,5 y 4, a efectos de ajustar el largo de los modelos de manera óptima para los experimentos siguientes. En la Tabla 3.5 se muestra el error de clasificación al utilizar los perfiles de 3 y 5 puntos por cuerda, para diferentes valores de $k$.

El comportamiento general obtenido muestra un decaimiento en el desempeño a medida que el factor de carga aumenta, por lo cual en adelante se fijará el valor de $k = 1{,}5$ como estándar.

#### Estudio de los perfiles de muestreo

Luego de decidir el valor de $k$, una segunda serie de experimentos fue llevada a cabo sobre la misma partición para estudiar el desempeño del clasi-

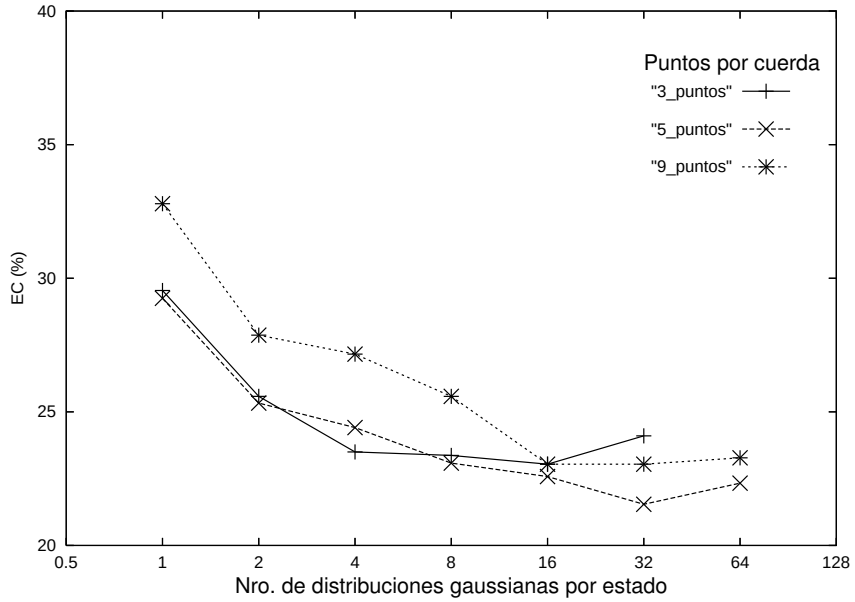



Figura 3.11: Evolución del error de clasificación para diferentes conjuntos de características: perfiles de 3, 5 y 9 puntos por cuerda.

ficador sobre diferentes conjuntos de características. La Figura 3.11 presenta los resultados obtenidos con los perfiles de muestreo de 3, 5 y 9 puntos por cuerda, al variar el número de distribuciones gaussianas por mezcla en cada estado de los MOM.

El comportamiento obtenido muestra un mejor desempeño para el conjunto de 5 puntos por cuerda, el cual es comparado con otros conjuntos en la Figura 3.12 para secuencias de vectores de 1, 3, 5 y 9 puntos fijando el número de distribuciones Gaussianas por mezcla a 16 y el factor de carga $k$ a 1.5.

Para finalizar la comparativa de los conjuntos de características, la Tabla 3.6 resume los errores mínimos en clasificación logrados por cada configuración, la cual muestra ventajas del conjunto de 5 puntos por cuerda respecto a los restantes.

### Variantes de los perfiles de muestreo

La extracción de características para la caracterización de cromosomas constituye un problema abierto de interés en la investigación actual [54, 36]. Hasta aquí, los perfiles de muestreo demostraron ser más adecuados que los perfiles clásicos para la clasificación. Sin embargo, los resultados todavía

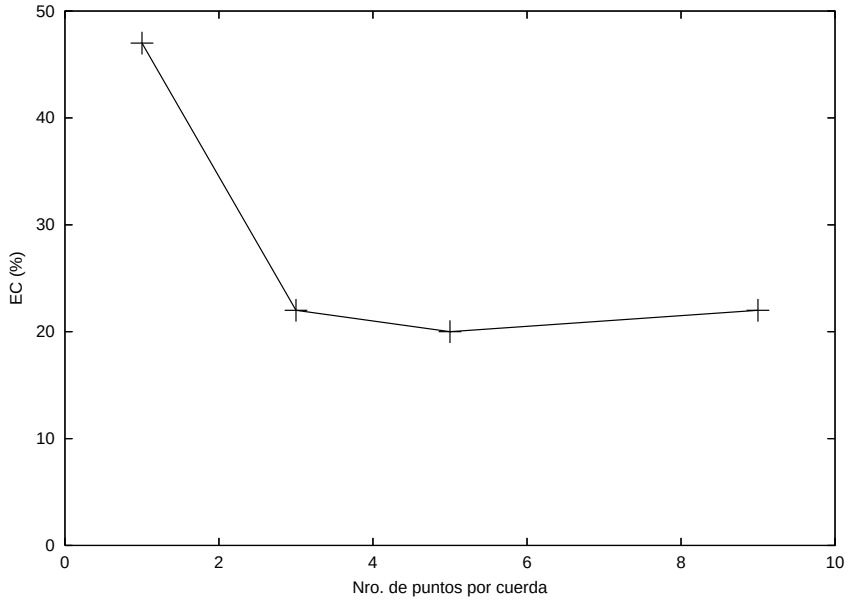



Figura 3.12: Evolución del error con el número de características, para mezclas de gaussianas de igual dimensión y factor de carga fijo.

distan de aquéllos obtenidos mediante modelos conexionistas, y con el ánimo de mejorar el desempeño de los MOM, en esta sección se consideran algunas variantes de los perfiles mediante el agregado de derivadas y aceleraciones en horizontal y vertical.

Considerando un perfil de muestreo inicial de 9 puntos por cuerda, valor que se elige igual que en modelos conexionistas y que facilita la comparación de resultados, se consideran los siguientes tipos de características:

- 9p: grises muestreados.
- D: derivada horizontal (a lo largo del cromosoma).

Tabla 3.6: Mejor desempeño en clasificación para perfiles de muestreo de diferente dimensión.

| Puntos por cuerda | Nro. de gaussianas por mezcla | EC (%) |
|---|---|---|
| 3 puntos | 16 | 23,04 % |
| 5 puntos | 32 | 21,54 % |
| 9 puntos | 32 | 23,10 % |

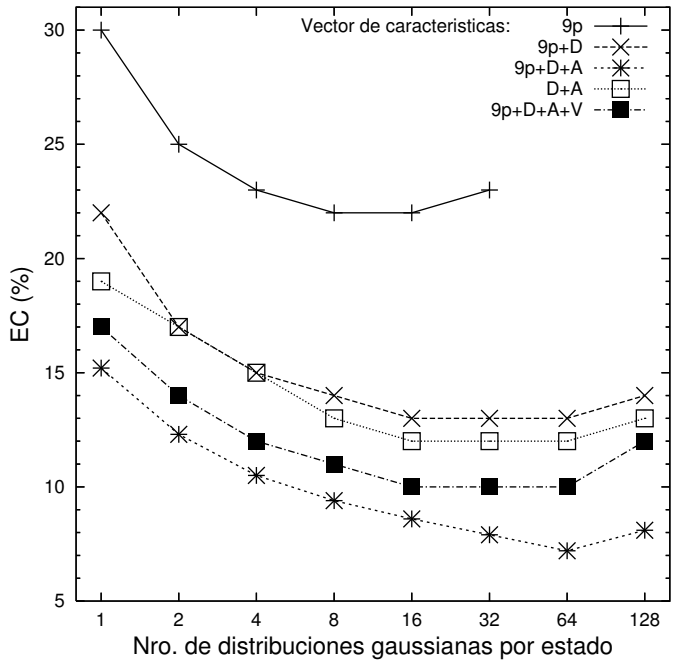



Figura 3.13: Error de clasificación como función del número de distribuciones gaussianas por estado, para variantes de conjuntos de características.

- A: aceleración horizontal.
- V: derivada vertical (a lo largo de la cuerda).

Las derivadas y aceleraciones fueron calculadas según las expresiones (2.5).

Variando el número de distribuciones gaussianas por mezcla, una serie de experimentos se llevó a cabo cuyos resultados se muestran en la Figura 3.13. Puede observarse que el conjunto de características más apropiado consiste en usar los grises muestreados más las derivadas horizontales y verticales (9p+D+V), donde 64 es un número adecuado de distribuciones gaussianas por estado.

En las redes neuronales recurrentes, una ventana móvil conteniendo una porción del patrón es procesada en cada instante de tiempo, lo cual demostró ser efectivo y mejorar la clasificación respecto a emplear solamente un vector por vez. En un intento de reproducir este comportamiento, una serie posterior de experimentos fue realizada empleando ventanas de vectores, con diferente longitud.

La Figura 3.14 muestra el error en clasificación obtenido al variar la cantidad de distribuciones gaussianas por mezcla, para ventanas conteniendo 1 a 4

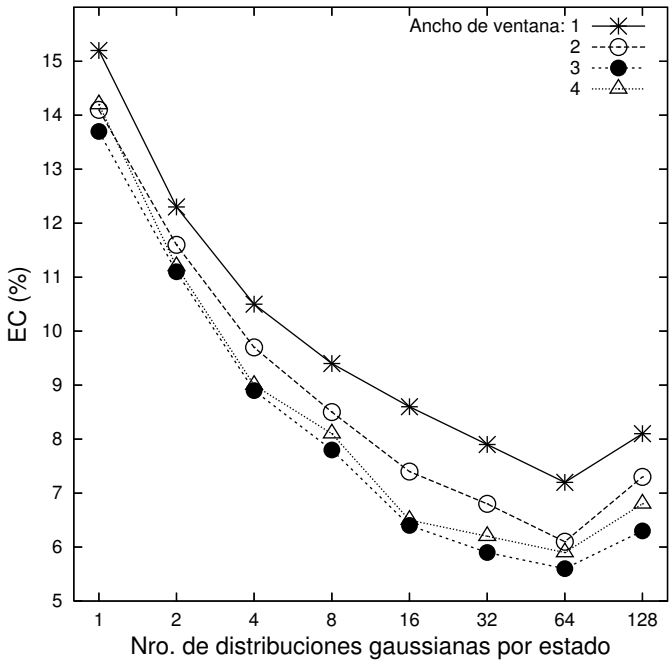



Figura 3.14: Error en clasificación como función del número de distribuciones gaussianas por estado, para diferentes anchos de ventanas.

cuerdas cada una. El mejor resultado obtenido (error de 5,6 %) fue obtenido usando una ventana de 3 vectores por ventana y 64 gaussianas por mezcla.

### 3.4.5 Experimento completo no-contextual

Todos los experimentos reportados hasta aquí fueron desarrollados utilizando una partición simple del corpus. Con el objetivo de obtener resultados más precisos evitando desvíos por la elección particular de los patrones, y que a la vez fueran comparables al experimento completo con redes de Elman, el error final en clasificación para la aproximación con MOM continuos fue estimado mediante el mismo esquema de validación cruzada en 7 bloques. Cada bloque contiene 400 células de prueba y las restantes 2400 de entrenamiento, sin solapamiento de células entre conjuntos de prueba.

En la Tabla 3.7 se exponen las tasas de error en clasificación como función del número de distribuciones gaussianas por estado, siendo el resto de los parámetros mantenido como los que lograran el mejor desempeño en los experimentos previos. Para este experimento, el rendimiento óptimo se alcanza

Tabla 3.7: Tasa de error de clasificación estimada usando el procedimiento de validación cruzada en 7 bloques. En cada caso, el error reportado corresponde a la media aritmética entre los errores obtenidos para todos los bloques.

| N° de gaussianas por estado | Error (%) |
|---|---|
| 1 | 15,7 |
| 2 | 12,5 |
| 4 | 10,4 |
| 8 | 9,3 |
| 16 | 8,4 |
| 32 | 7,9 |
| 64 | 7,5 |
| 128 | 7,7 |

con un error medio de 7,5 % para 64 Gaussianas por mezcla.

Los resultados obtenidos en la variación del factor de carga de estado muestra que el desempeño es óptimo si la longitud de los modelos es la más larga posible, correspondiente a valores pequeños del parámetro. Esta longitud permite a los MOM aproximar de manera más adecuada toda la variabilidad presente en las bandas claras y oscuras y las transiciones entre ellas.

Asimismo, en la comparación entre perfiles clásicos y de muestreo, se demostró la capacidad de representación que tienen los perfiles de muestreo respecto a los perfiles clásicos, al disminuírse el error de clasificación.

Un punto que apoya la conclusión anterior puede estudiarse haciendo un análisis de las imágenes del corpus. En los cromosomas cuyas cromátidas son visibles en la metafase, en el momento anterior a la separación, se tiene entre ambas estructuras puntos claros que visualmente se identifican como el fondo de la imagen. Sin embargo, los pixeles de esta separación no tienen valores iguales al brillo de fondo por cuestiones de digitalización, o por encontrarse demasiado cercanos. De esta manera, el borde calculado en el preproceso engloba a las dos cromátidas sin presentar la cisura correspondiente, y una cuerda típica de estos cromosomas mostrará zonas altamente oscuras (la tinción de los cromosomas) cercana a los bordes, con una zona clara en el centro. Si solamente se calcula un valor de gris promedio por cuerda, se está perdiendo información acerca de estas características de los cromosomas, problema que el muestreo mejora ya que rescata detalles más finos sobre porciones de

la cuerda.

Comparando entre sí a los perfiles de muestreo con diferente cantidad de puntos, los experimentos con 3 puntos lograron tasas de error mayores que el caso de 5 puntos por cuerda, lo cual indica que la información de 3 grises es insuficiente para modelar la variabilidad vertical sobre la cuerda. Por otro lado, los experimentos con 9 puntos por cuerda obtuvieron tasas de error mayores que en el caso de 5 puntos. Este comportamiento estaría indicando que la información extra dada a los modelos no agrega información significativa para aproximar las distribuciones de probabilidad, produciéndose el decaimiento por la mayor cantidad de parámetros a estimar.

El agregado de información "temporal", tal como la derivada y la aceleración, conduce a una reducción significativa en la tasa de error (de 22 % a 7,2 %). Sin embargo, la derivada vertical (considerada a lo largo de las cuerdas) no ayuda en igual manera a reducir el error, sino que provoca un aumento en esta tasa de 7,2 % a 10 %. Este efecto es producido por la variabilidad extra sobre las cuerdas mencionada previamente, cuya derivada podría estar agregando información no discriminativa a los modelos, explicando el aumento en la tasa de error.

Al igual de lo que sucede con las redes de Elman, los MOM capturan la variación temporal de los eventos en la entrada con la topología y parámetros de cada estado. No obstante, se demostró aquí que la construcción de una entrada aumentada en longitud demostró ser muy efectiva a los efectos de la clasificación. El uso de ventanas de 3 cuerdas ayuda a obtener una reducción en el error de 7,2 % a 5,6 %.

Finalmente fue reportado el experimento completo para clasificación de cromosomas sin restricciones al contexto celular, con las 7 particiones del corpus en las mismas condiciones que en redes de Elman. En la sección siguiente se restringen estos resultados a un contexto celular, donde se espera que el agregado de información del cariotipo logre mejorar más aún el desempeño de los clasificadores.

## 3.5 Clasificación contextual iterativa

### 3.5.1 Formulación del algoritmo contextual

La idea principal del algoritmo consiste en la reasignación de las clases previamente asignadas por el clasificador no-contextual, en aquellos cromosomas que hubieran sido clasificados con baja confiabilidad. El proceso se basa en la exploración, para cada patrón de entrada, de las probabilidades de salida de cada clase obtenidas con el clasificador no-contextual, y la reasignación es llevada a cabo por medio de una búsqueda iterativa de cromosomas homólogos en los autosomas (clases 1 a 22), y la búsqueda de los dos cromosomas sexuales. De esta manera, la información del cariotipo es tenida en cuenta con el objeto de arreglar algunas asignaciones erróneas de clase que pudieran haber tenido lugar en la clasificación inicial. El algoritmo fue denominado Clasificación Contextual Iterativa, y se denota por su sigla en inglés ICC (*Iterative Contextual Classification*).

La entrada del ICC consiste en la clasificación no-contextual de cada cromosoma para una célula completa. Esta información se organiza, para cada cromosoma $p$, como un vector $V_p$ de dimensión $C$ (número de clases) conteniendo las salidas del clasificador normalizadas como probabilidad.

En el caso de las redes neuronales tipo perceptrón multicapa, las probabilidades son obtenidas de cada nodo $k$ de la capa de salida haciendo: $V_k = \frac{O_k}{\sum_{k=1}^{C} O_k}$, donde $O_k$ corresponde a las activaciones en los nodos de salida obtenidas mediante el algoritmo de retropropagación.

En el caso de las redes de Elman, cada patrón es analizado mediante las ventanas móviles, donde para cada una de ellas se obtiene un vector de 24 componentes con las activaciones de salida correspondientes a cada clase. De este modo, una vez clasificado cada cromosoma se obtiene una serie de $A$ vectores de activaciones, donde $A$ corresponde al número de ventanas de análisis para el patrón de entrada. Luego, las probabilidades de cada nodo de salida se calculan como $V_k = \frac{\sum_i w_{ik}}{A}$.

En el caso de los MOM, la entrada del ICC se toma directamente de los valores de verosimilitud obtenidos para cada modelo luego de la clasificación del patrón.

La información obtenida es ordenada en una matriz $\mathbf{M}$ de dimensiones

$N \times C$, donde $N$ es el número de cromosomas en la célula y $C$ es el número de clases a encontrar. Las probabilidades de salida son colocadas en orden decreciente (acción denotada en el Algoritmo 1 por la función 'reordenarProbabilidades()'), de manera que la matriz tiene las siguientes propiedades:

$$\sum_{k=1}^{C} \mathbf{M}(p,k) = 1, \quad 1 \leq p \leq N,$$
$$\mathbf{M}(p,1) > \mathbf{M}(p,2) > \ldots > \mathbf{M}(p,C), \quad 1 \leq p \leq N.$$

La parte central del algoritmo es la búsqueda de las denominadas *clases resueltas*: son aquéllas clases con solamente 2 patrones en la célula teniendo su máxima probabilidad de salida para esa clase, para el caso de los autosomas, o el par de cromosomas sexuales. Como un requisito adicional, un umbral mínimo de ambas probabilidades de 0,8 es fijado con el objeto de asegurar que la clasificación no-contextual haya sido realizada con alta confiabilidad. De esta manera, el ICC considera que los patrones de las clases resueltas fueron correctamente clasificados y, por lo tanto, no son incluídos en la etapa posterior de reasignación de clases.

Antes de comenzar la búsqueda iterativa, el ICC registra todas las clases resueltas provenientes de la clasificación no-contextual, y esos patrones no se consideran en adelante. En ciclos sucesivos, el algoritmo encuentra el resto de pares resueltos por medio del método de *renormalización*: para cada patrón, el método recalcula las probabilidades de salida de acuerdo a la cantidad de clases todavía no resueltas. A continuación, se aplica nuevamente la subrutina de búsqueda de nuevas clases resueltas. Si no se encontrara ninguna en esa iteración en particular, el ICC busca el par de patrones cuyo producto de probabilidades es máximo para una clase dada, y reasigna esos patrones al conjunto de clases resueltas. Este esquema es repetido hasta que todas las clases han sido resueltas.

El ICC realiza un tratamiento separado para los cromosomas sexuales, ya que las células normales tienen dos cromosomas de este tipo: la combinación $X - X$ en células femeninas o $X - Y$ en masculinas. Así, la búsqueda es restringida a 23 clases: 22 autosomas más la clase sexual, permitiendo las combinaciones mencionadas anteriormente. Para la clasificación de células anormales, el algoritmo toma en cuenta los casos de células con un cromosoma faltante, y con un cromosoma extra. En el primer caso, se buscan los mejores 22 pares de patrones y el patrón restante se reasigna a la clase restante no resuelta. En el segundo caso, una vez que todas las células han sido resueltas con su par correspondiente, el cromosoma extra es asignado a la clase con

mayor probabilidad de salida.

La Figura 3.15 lista el proceso completo para el caso genérico de una célula normal,mientras que la subrutina para la búsqueda de clases resueltas se lista en la Figura 3.16.

### 3.5.2 Experimentos y resultados

La Figura 3.17 presenta una serie de medidas sobre el EC obtenidos por bloque en el experimento completo de validación cruzada con redes de Elman[2]. El análisis se realiza por clase, donde las barras de error muestran la media aritmética del EC (marca central) y el desvío estándar (marcas finales), para las 7 particiones. Es de notar la alta variabilidad en los resultados para algunas clases, llegándose a alcanzar valores de EC por encima de 20 %.

El algoritmo ICC fue aplicado a los resultados no-contextuales, obteniéndose un EC medio de $3,9\,\%$. La Tabla 3.8 presenta los resultados obtenidos para cada partición, contrastando las dos aproximaciones. La Figura 3.18 muestra la media y el desvío estándar de las particiones, siguiendo los lineamientos de la Figura 3.5. Puede observarse fácilmente las mejoras en el EC para cada clase, y la significante reducción en la dispersión de los resultados entre particiones.

La aplicación del algoritmo ICC toma un tiempo promedio de 1,6 milisegundos por célula, dependiendo de la clasificación no-contextual.

Con el objeto de validar los resultados y el desempeño del algoritmo ICC propuesto, se evaluó la significancia estadística de las tasas de error en clasificación independiente del contexto ($\varepsilon_{ci}$) versus los valores obtenidos luego de la aplicación del algoritmo ICC ($\varepsilon_{cd}$) [55]. Se obtuvo que $\Pr(\varepsilon_{ci} > \varepsilon_{cd}) > 99{,}999\,\%$, mostrando el importante beneficio en la aplicación de este algoritmo de post-clasificación.

Con propósitos de comparación con otras aproximaciones basadas en redes neuronales aplicadas al corpus Copenhagen, se presenta en la Tabla 3.9 el resumen de resultados de este trabajo junto a un Perceptrón multicapa con una capa de salida de tamaño reducida propuesto en [31], un Perceptrón multicapa estándar con una capa oculta propuesto en [56] y un sistema jerárquico que clasifica los patrones primeramente en 7 grupos de clases principales y luego en una de las 24 clases del cariotipo, propuesto en [32].

[2]Se incluye esta figura para claridad en la comparativa, aunque ya fuera introducida como Figura 3.5.

**Datos**:
$N$: número de patrones de la célula,
$C$: número de clases,
$\mathbf{M}$: $N \times C$ matriz de probabilidades de salida no-contextual.
**Resultados**:
$\mathbf{R} = \{(t, o)\}$: $N \times 2$ matriz de pares (clase deseada,clase reasignada) para cada patrón.
$V$: vector conteniendo las clases resueltas.
$T$: vector conteniendo los patrones resueltos.
**inicialización**
$\mathbf{R}, T, V \Leftarrow \emptyset$
**begin**
reordenarProbabilidades($\mathbf{M}$)
buscarCR($\mathbf{M}, V, T, c$) `/* búsqueda de clases resueltas */`
**mientras** $\#V < C$ **hacer** `/* -->Búsqueda Iterativa <-- */`
**para** $1 \leq p \leq N,\ p \notin T$ **hacer** `/* renormalización */`
$s \leftarrow \sum_{k=1, k \notin V}^{C} \mathbf{M}(p, k)$
$\mathbf{M}(p, :) \leftarrow \mathbf{M}(p, :)/s$
**fin para**
buscarCR($\mathbf{M}, V, T, c$) `/* búsqueda de nuevas clases */`
**si** $c$=falso **entonces** `/* no hay clases resueltas */`
reordenarProbabilidades($\mathbf{M}$)
$P, Q \Leftarrow \emptyset$
**para** $1 \leq k \leq C,\ k \notin V$ **hacer** `/* selecciona pares de patrones */`
`/* de clases no resueltas */`
$Q(k) \Leftarrow$ los dos patrones con máxima probabilidad entre todos los patrones de clase $k$
$P(k) \Leftarrow$ producto de probabilidades de los patrones en $Q(k)$
**fin para**
$i \leftarrow \text{argmax}_i\, P(i)$ `/* toma el mejor par */`
$V \Leftarrow$ clase deseada de los patrones en $Q(i)$
$T \Leftarrow$ patrones en $Q(i)$
$\mathbf{R} \leftarrow$ {clase deseada de patrones en $Q(i)$,clase de salida de patrones en $Q(i)$}
**fin si**
**fin mientras**
**end**

Figura 3.15: Algoritmo ICC de clasificación contextual iterativa).

**Datos**:
$\mathbf{M}, V, T$;
**Resultados**:
Bandera $c$ indicando la presencia de clases resueltas.
$\mathbf{R}, V, T$: actualizadas con las nuevas clases resueltas encontradas.
**begin**
  $c \leftarrow$ falso ;
  **para** $1 \leq k \leq C$ **hacer**
    **si** [ (número de patrones de clase $k$) =
    2 $\wedge$ (ambas probabilidades $>$ 0,8)] **entonces**
      $T \Leftarrow$ ambos índices de patrón $p'$ ;
      $V \Leftarrow$ clase de los cromosomas $k$ ;
      $\mathbf{R} \Leftarrow \{$salida deseada de los patrones $p', k\}$ ;
      $c \leftarrow$ verdadero ;
  **return** c
**end**

Figura 3.16: Subrutina buscarCR()

Para completar el análisis se incluyen los resultados de la aplicación del ICC cuando los modelos ocultos continuos de Markov fueron empleados como clasificador no-contextual. En este caso, el error previo de 7,5 % para modelos de 64 distribuciones gaussianas por estado se reduce luego de la aplicación del ICC a 4,6 %, en ambos casos como error medio para los 7 bloques del experimento completo de validación cruzada. En esta comparativa puede observarse que el sistema basado en redes neuronales parcialmente recurrentes mejora la tasa de error de los restantes sistemas.

Se mostró que la red neuronal recurrente claramente mejora el desempeño del Perceptrón multicapa cuando se utiliza solamente el patrón de bandas como entrada. La clasificación independiente del contexto podría ser mejorada si otras características ampliamente difundidas (por ej., medidas globales, índice centromérico u otras) fueran utilizadas conjuntamente con otros clasificadores reportados para esta tarea. En este sentido, la clasificación inicial de los cromosomas es relativamente más importante para la tasa final de error más que la etapa de clasificación contextual, ya que el desempeño del ICC depende exclusivamente de las probabilidades de salida asignadas para cada clase.

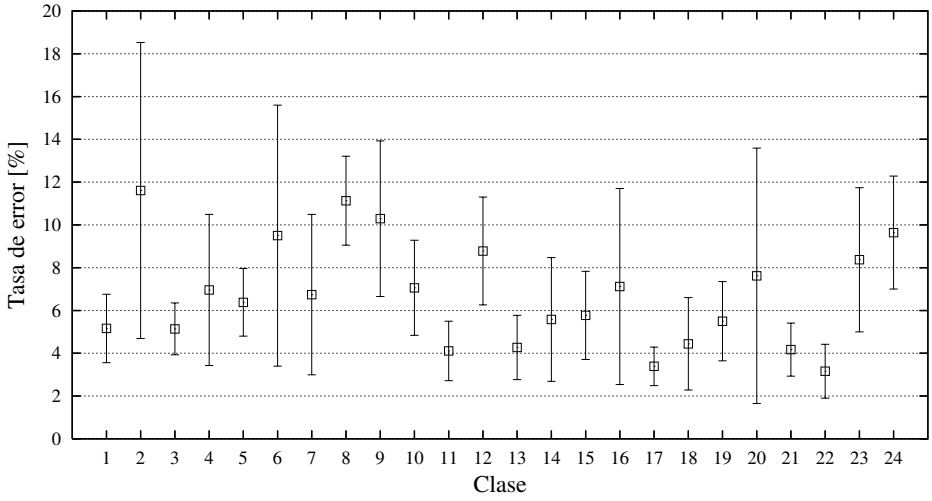



Figura 3.17: Error medio y desvío estándar (marca central y extremos) para cada clase, como fueron obtenidos para las 7 particiones del experimento de validación cruzada independiente del contexto con redes de Elman.

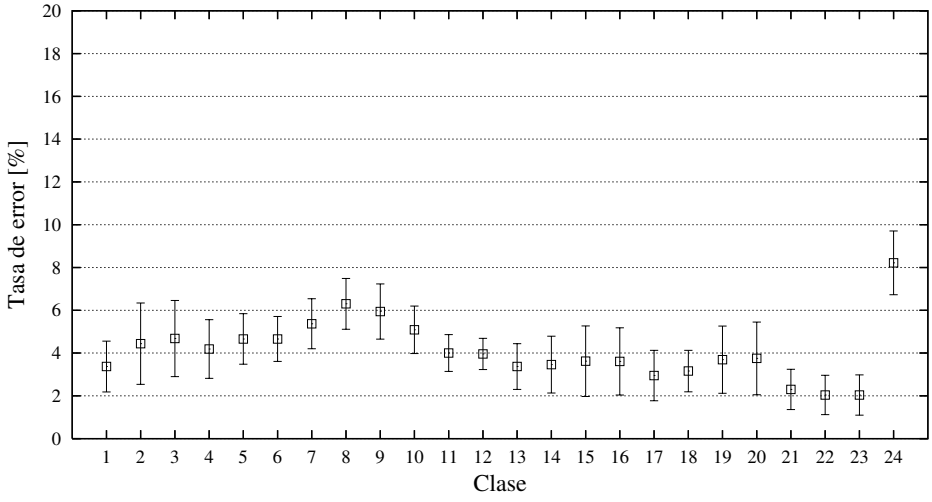



Figura 3.18: Error medio y desvío estándar (marca central y extremos) obtenidos luego de la aplicación del algoritmo ICC en el experimento con redes de Elman.

Tabla 3.8: Resultados para el experimento completo de validación cruzada con redes de Elman. En filas, el EC para la clasificación independiente del contexto y luego de la aplicación del algoritmo ICC. En columnas, cada una de las $p$ particiones del experimento.

| Modo | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ | $p_7$ | Media |
|---|---|---|---|---|---|---|---|---|
| Independiente del contexto | 6,0 | 6,7 | 6,1 | 8,3 | 7,5 | 5,6 | 6,8 | 6,7 |
| Dependiente del contexto | 3,6 | 3,8 | 3,7 | 4,9 | 4,8 | 2,7 | 4,0 | 3,9 |

Tabla 3.9: Resultados para diferentes clasificadores sobre el corpus Copenhagen.

| Sistema | Tasa de error |
|---|---|
| Perceptrón multicapa con capa de salida reducida [31] | 11,7 % |
| Perceptrón multicapa estándar [56] | 6,5 % |
| Red neuronal jerárquica [32] | 5,6 % |
| Modelo oculto continuo de Markov-ICC | 4,6 % |
| Redes de Elman-ICC | 3,9 % |

### 3.5.3 Discusión de resultados

El comportamiento del algoritmo depende, en gran medida, no solamente de la clase ganadora obtenida por el clasificador no-contextual sino también del vector completo de *probabilidades de salida* de la matriz $M$. Así, la clasificación de la célula completa debe haber sido realizada con una buena precisión, principalmente en los primeros lugares del vector de probabilidades, para lograr que el reacomodo subsiguiente logre su cometido de disminuir la tasa de error de manera significativa.

El método contextual logra, para el 68 % de las células, una clasificación con el 100 % de exactitud, con reducciones del error de hasta más del 10 %. El 21 % de las células no pudieron ser clasificadas con alta exactitud por el clasificador no-contextual, y el método logra reducir el error pero sin lograr una exactitud del 100 %. Por otro lado, para el 7 % de las células, la aplicación del ICC obtiene la misma tasa de error que la obtenida en la etapa no-contextual, mientras que en el 4 % de las células la tasa de error final obtenida

es mayor que la previa a la aplicación del ICC.

Las Figuras 3.19-3.22 muestran las matrices de confusión antes y después de la aplicación del algoritmo ICC, ilustrando ejemplos de las situaciones previamente mencionadas cuando se utiliza una red de Elman como clasificador no-contextual. En las gráficas, en las filas se denotan los cromosomas y en columnas las clases del cariotipo, con los cuadradros negros significando una clasificación correcta y un cuadrado gris un error. La Figura 3.19 muestra el resultado para una célula donde la red neuronal obtiene una tasa de acierto del 89 % y el ICC es capaz de corregir todos los errores previos. En el segundo caso, ejemplificado por la Figura 3.20 la red clasifica erróneamente 6 cromosomas (números 12 a 17) y el ICC logra corregir los errores de 4 cromosomas (números 13, 15, 16 y 17). Esta vez, la tasa de acierto es mejorada casi un 10 % (de manera similar al caso anterior), pero desde un valor inicial de 86 % a 95 %. La Figure 3.21–en el tercer caso– muestra que los cromosomas número 19 y 20 son inicialmente mal clasificados por la red neuronal. El ICC elige como componentes del par a reasignar a los cromosomas número 18 y 20. Luego de la aplicación del algoritmo, ambos cromosomas son asignados a clases erróneas, lo que origina que la tasa de acierto antes y después del ICC sea la misma, en este caso 95 %. En el último caso, mostrado en la Figura 3.22, la red neuronal asigna dos cromosomas extras a la clase 7 (cromosomas número 12 y 15). Luego, el ICC elige dos de los cuatro cromosomas para el reacomodo, pero el par incorrecto es elegido (cromosomas número 14 y 15), debido a los errores presentes en el vector de probabilidades no-contextuales. Esta elección introduce un error en el cromosoma 14 –correctamente clasificado por la red neuronal– mientras que mantiene los errores previos en los cromosomas 12 y 15, de manera que la exactitud de la clasificación es reducida de 95 % a 93 %.

El algoritmo ICC presentado aquí obtiene una reducción media del 50 % sobre el error independiente del contexto (3 puntos en el caso de las redes de Elman), como puede verse en la Tabla 3.8. Este resultado constituye un buen desempeño para un método de reasignación, dado que en [57] los autores encontraron que el algoritmo de transporte logró reducir la tasa de error pero en una cantidad pequeña (de 6,5 % to 4,4 % sobre el corpus Copenhagen). La aproximación propuesta más recientemente en [33] logra mejorar el desempeño del algoritmo de transporte (cerca del 33 %, debido a que este último método no explota las similaridades de características dentro de una célula), obteniendo reducciones de aproximadamente 3 puntos.

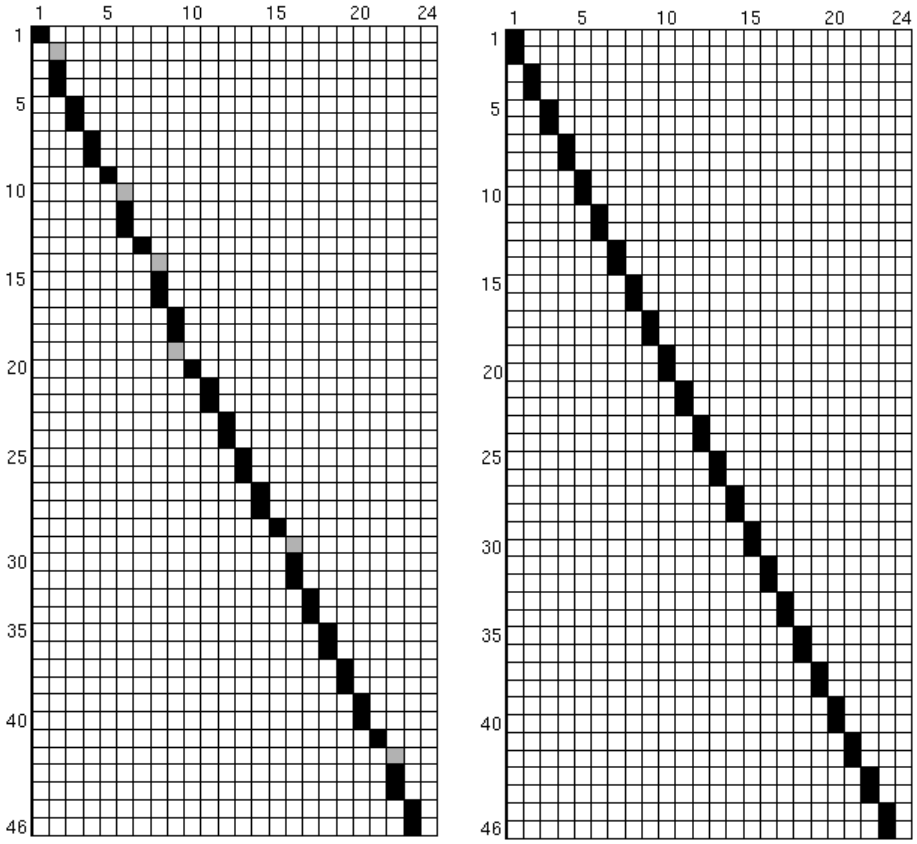



Figura 3.19: Matriz de confusión para una célula, mostrando la clasificación no-contextual a la izquierda y la dependiente del contexto a la derecha. La célula fue clasificada con una tasa de acierto del 89 % por una red neuronal (a), luego corregida a un 100 % por el algoritmo ICC (b).

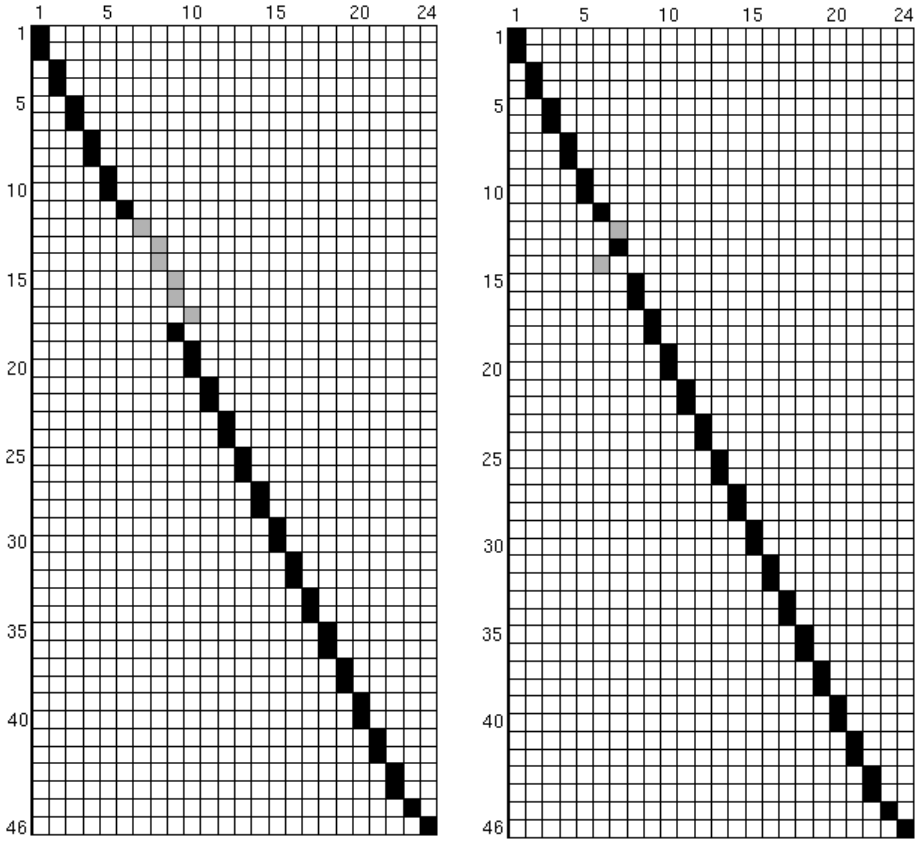



Figura 3.20: Clasificación no-contextual con acierto del 86 % (a), luego corregida a 95 % por el algoritmo ICC (b).

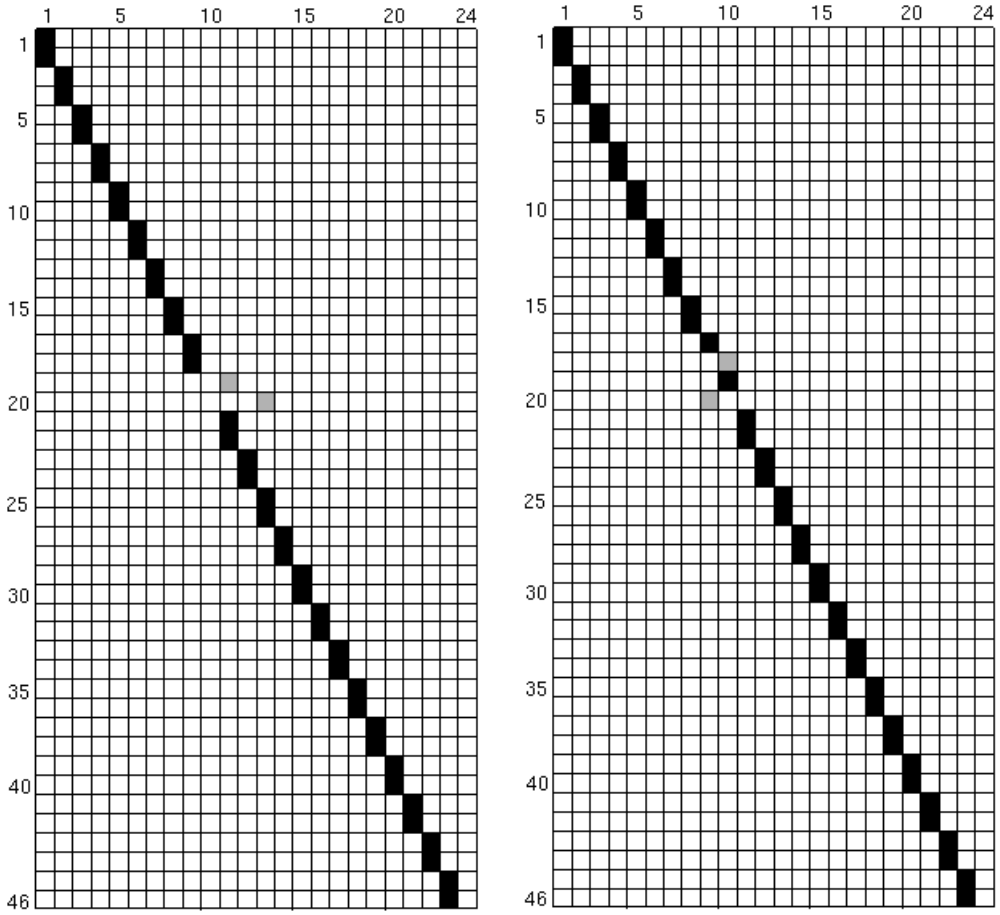



Figura 3.21: La célula fue clasificada con un 95 % de acierto por la red neuronal(a), obteniendo la misma tasa de error luego de la aplicación del algoritmo ICC (b).

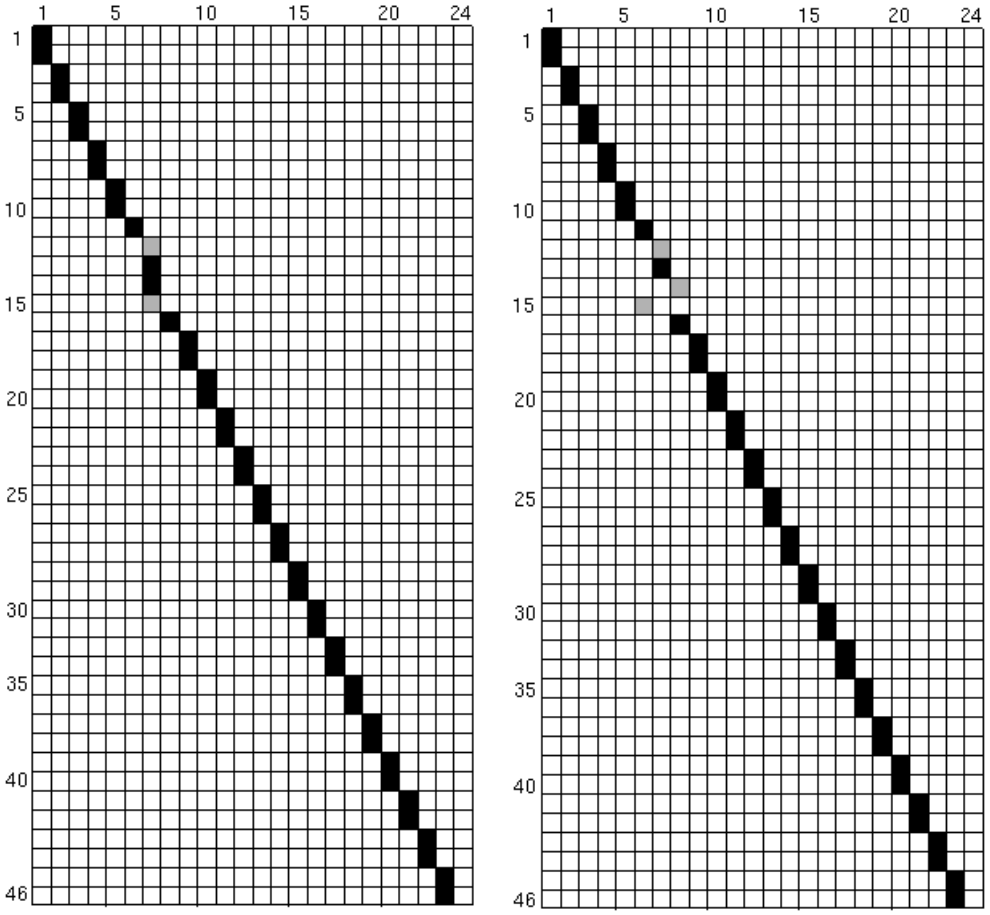



Figura 3.22: En un caso de aplicación negativo, la célula había obtenido 95 % por la red neuronal (a). Luego de la aplicación del ICC, la tasa de acierto es reducida a 93 % (b).

## 3.6 Comentarios de cierre de capítulo

En este Capítulo se presentó la experimentación sobre clasificadores de distinto tipo, las redes neuronales tipo perceptrón multicapa y los modelos ocultos de Markov, sobre los patrones extraídos usando las técnicas propuestas en el Capítulo previo.

Se obtuvieron resultados que mejoran los del estado del arte en la tarea de clasificación aislada de cromosoma, así también como dependiente del número presente en células sanas y con un cromosoma faltante, mediante un algoritmo de contextualización propuesto a tal fin.

A continuación, el documento se dedica a exponer el trabajo realizado en la línea de investigación sobre representación cortical y clasificación robusta de habla.

# Parte II

# Reconocimiento del habla

CAPÍTULO 4

# Representación cortical de la señal de habla

En este capítulo se exponen los conceptos teóricos del método de extracción de características que se plantea aquí como una alternativa a las técnicas convencionales de parametrización del habla: las representaciones ralas.

En particular, se trabajará sobre la detección de pistas acústicas sobre un análisis tiempo-frecuencia obtenido a nivel de corteza auditiva primaria, el espectrograma auditivo. Se mostrará como es posible estimar un conjunto de señales básicas a partir de este espectrograma, con las cuales se puede luego analizar cualquier señal de prueba y obtener una representación en forma de activación de coeficientes. Este nuevo enfoque generaliza, de este modo, el análisis clásico basado en la transformación de una señal en función de una base de elementos del mismo tipo, como las exponenciales complejas de Fourier.

Finalmente, se muestran experimentos con señales sintéticas a fin de mostrar las capacidades del método en la extracción de características significativas de la señal, las que se mantienen aún en condiciones adversas de baja relación señal/ruido.

## 4.1 Consideraciones preliminares

Como se explicó en el Capítulo 1, los sistemas artificiales de clasificación de patrones incorporan, en general, una etapa de adecuación de la señal de entrada y su expresión en alguna otra forma –*características*– más manejable por el clasificador en sí mismo. En el campo del reconocimiento automático del habla (RAH), el análisis espectral de tiempo corto se ha establecido como una técnica clásica de representación, la cual es aplicada a la señal acústica del habla. Es así que los métodos espectrales fueron ampliamente estudiados y aplicados como paso inicial en la etapa de extracción de características.

El enfoque tradicional mencionado supone como hipótesis una serie de hechos necesarios para la correcta obtención de la representación; entre los cuales se pueden mencionar la linealidad del sistema de procesamiento, la invarianza temporal de los sistemas de producción de la voz, la estadística significativa hasta segundo orden y la ortogonalidad de la representación [26]. Justamente esta última condición conlleva la obtención de una descripción matemática más simple y manejable que la señal original desde el punto de vista del análisis, mediante un cambio de espacio a otro de igual o de menor dimensión, induciendo en este caso un error que puede ser despreciado.

Históricamente, las señales de habla fueron registradas en ambientes acústicamente aislados, como cámaras sonoamortiguadas o anecoicas, lo que hizo posible el desarrollo de la disciplina. Más recientemente, las aplicaciones de interés se extendieron al trabajo con señales registradas en ambientes reales, en condiciones que se alejan de las ideales encontradas en los ambientes aislados, dando como resultados señales inmersas en ruidos no deseados. Ejemplos de sistemas que trabajen en estos ambientes pueden se los reconocedores de palabras para telefonía móvil, procesadores adaptativos de señal para prótesis auditivas, sistemas *manos libres* para automóviles, etc. Sin embargo, estos aún distan de alcanzar la eficiencia que logran los sistemas sensoriales naturales en tareas complejas con múltiples hablantes, vocabulario extenso, habla natural y ambientes con ruido de fondo, entre otras condiciones adversas [58, 59].

En un ambiente no controlado, muchas de las consideraciones iniciales realizadas acerca de la generación y tratamiento de la señal acústica ya no se cumplen o presentan grandes limitaciones en su ámbito de aplicabilidad. Debido a esto surgió el interés en desarrollar métodos alternativos a los tradicionales, cuyos enfoques sean más generales y abarcativos que los iniciales.

Recientemente se ha retomado el interés en el desarrollo de los denominados *sistemas bioinspirados*. De acuerdo a estas ideas, es posible obtener la representación de la señal de la voz mediante una transformación que obtiene señales con características similares a las obtenidas experimentalmente a nivel de corteza auditiva primaria: los campos receptivos espectro-temporales. Esta representación posee características ralas, las que han sido empleadas en el procesamiento de señales obtenidas de sistemas no lineales [60]. La representación rala de la señal permite la obtención de los patrones para clasificación, contribuyendo a la extracción de pistas útiles para el reconocimiento.

Este es el contexto principal que da cabida a los métodos alternativos de representación, entre los cuales se encuentra el método desarrollado en esta tesis. El resto del capítulo presenta formalmente estas ideas y la experimentación respectiva.

## 4.2 Representación basada en diccionarios discretos

### 4.2.1 Análisis clásico

El habla se encuentra disponible como una señal acústica en el dominio temporal, $x(t)$. La transformada de Fourier de tiempo corto (STFT, por su sigla en inglés *short-time Fourier transform*) se ha constituido como la herramienta de referencia en el análisis del habla, la cual consiste en una transformación integral cuyo núcleo de descomposición son funciones exponenciales complejas de diferente frecuencia. El método de representación supone la estacionariedad de la señal en el tramo de señal considerado[1]:

La descomposición se aplica sobre la señal ventaneada

$$x_v(t;\tau) = x(t)\ v(t-\tau), \tag{4.1}$$

donde $v(t)$ es una ventana móvil de soporte compacto desplazada un tiempo $\tau$. Suponiendo estacionario el tramo analizado, la transformada de Fourier de

[1]Para la transformada de Fourier se demuestra que las exponenciales complejas de la base constituyen las autofunciones de un sistema lineal e invariante en el tiempo (LTI, por su sigla en inglés *linear, time invariant*).

la secuencia de segmentos ventaneados constituye la transformada de Fourier de tiempo corto:

$$S_F(\tau, f) = \int_{-\infty}^{\infty} x(t)\, v^*(t-\tau)\, e^{-j2\pi ft} dt. \tag{4.2}$$

Se obtiene así una representación de la señal $x(t)$ en el plano tiempo-frecuencia $(\tau, f)$. Una interpretación que permitirá contrastar a la STFT con los métodos desarrollados posteriormente es tratar a la descomposición obtenida como una comparación de la señal $x(t)$ con un diccionario de señales $\phi_{\tau,f}(t) = v(t-\tau)\, e^{j2\pi ft}$, también denominados átomos tiempo-frecuencia:

$$S_F(\tau, f) = \left\langle x(t), v(t-\tau)\, e^{j2\pi ft} \right\rangle.$$

Los términos indicados por el producto interno $\left\langle x(t), \phi_{\gamma=(\tau,f)}(t) \right\rangle$ pueden ser vistos como proyecciones en el plano $(t, f)$ que brindan información acerca del contenido frecuencial del tramo considerado en el sector conocido como *rectángulo de Heisenberg* de $\phi_\gamma(t)$.

### 4.2.2 Análisis no convencional

En el análisis realizado por la STFT, las características de la ventana $g(t)$ se eligen al inicio (tipo de ventana y ancho) y se mantienen constantes durante todo el proceso. De esta manera, queda fija la resolución temporal y frecuencial, y al ser dos magnitudes proporcionalmente inversas se hace evidente la desventaja enunciada en el *principio de incertidumbre de Heisenberg*: no es posible tener buena resolución en ambas dimensiones simultáneamente.

Por otro lado, el análisis clásico realiza la factorización espectral con el mismo conjunto de átomos $\phi_{\gamma=(\tau,f)}$ sin tener en cuenta ningún aspecto morfológico de la señal estudiada. Es así que para señales formadas por combinaciones lineales de los átomos de la base (por ej. una sinusoidal analizada con una base de exponenciales complejas), la STFT obtiene una representación adecuada en términos de la –baja– cantidad de componentes significativas; mientras que para señales que se alejan de las características de los átomos (por ej. una señal cuadrada analizada con la base antes citada) se obtiene una gran dispersión de energía sobre toda los elementos.

Una alternativa que trata de eliminar las restricciones presentadas está dada por el conjunto de métodos de análisis no lineal y/o no estacionario. Aquí, la idea principal es relajar la linealidad a fin de obtener una representación

que sea más efectiva en el tratamiento de señales obtenidas en condiciones adversas, como las inmersas en ruido, con baja relación señal/ruido, mezcladas con otras señales de interés, con características temporales o frecuenciales particulares, etc.

Las técnicas estudiadas en esta tesis forman parte de los *métodos de aproximación*, las cuales tratan a la señal analizada como formada por diversas componentes de interés, los átomos. Aquí, los átomos no forman necesariamente un base o son un conjunto ortogonal, además puede también existir redundancia al incluir mayor cantidad de elementos que la dimensión del espacio de la señal analizada, por lo que el conjunto de átomos se denomina *diccionario*.

El término *de aproximación* en el nombre genérico de estos métodos viene dado por la forma de obtención de la representación. Aquí se busca encontrar una combinación de los átomos del diccionario que obtenga el menor error en la aproximación de la señal, a medida que aumenta la cantidad de átomos considerados [61]. Estos métodos presentan sus ventajas al lograr una representación con átomos seleccionados en forma adaptativa, de acuerdo a características particulares de la señal analizada. De manera general, entonces, obtienen una aproximación no-lineal a $x(t)$, la cual puede ser también lineal en caso de emplearse un subconjunto ortogonal de elementos del diccionario.

### 4.2.3 Representaciones ralas

En la búsqueda de alternativas al análisis clásico empleada en el RAH, interesa que el tratamiento con representaciones basadas en diccionarios logre que ante un estímulo o patrón de interés en la entrada solamente unos pocos elementos sean capaces de representarlo. Este comportamiento es deseable ya que al activarse sólo un número reducido de átomos se condensa la "densidad" de la codificación distribuída de la STFT. Tal tipo de parametrización corresponde a las denominadas *representaciones ralas*.

Ya que en la codificación distribuída la gran mayoría de los elementos se activan, ésta se asocia con distribuciones de probabilidad de tipo gaussiana para sus coeficientes. Las representaciones ralas, en cambio, están asociadas a densidades de probabilidad de alta kurtosis, como la laplaciana. Estas características trae aparejadas algunas ventajas de la representación, como ser una mayor eficiencia de la codificación y una mejor resolución de eventos [62].

Como se adelantó previamente, la cantidad de átomos del diccionario es

un parámetro importante en la concepción de las representaciones ralas. Con diccionarios de tamaño reducido (inferior a la dimensión de los átomos), no es posible generar una codificacion rala, mientras que si se igualan estas cantidades se tiene el caso de un diccionario *completo*. Por último, si la cantidad de elementos se elige mayor a la dimensión de los mismos, el diccionario empleado se denomina *sobrecompleto*. La proyección obtenida en este tipo de diccionarios resulta ser, bajo algunas condiciones, más simple de analizar, lo que sería beneficioso para los sistemas de clasificación [63].

Para llegar a la definición de una representación rala, se expone primeramente el esquema del modelado mediante un diccionario. Sea $\mathbf{x} \in \mathbb{R}^N$ la señal analizada. El esquema plantea la construcción de la señal mediante un modelo $\mathcal{M}$, donde $\mathbf{\Phi} \in \mathbb{R}^{N \times M}$ es el diccionario y $\mathbf{a} \in \mathbb{R}^M$ los coeficientes de su representación. Además, se incorpora explícitamente el efecto del ruido $\varepsilon \in \mathbb{R}^N$ que pudiera estar contaminando la señal:

$$\mathbf{x} = \mathcal{M}(\mathbf{\Phi}, \mathbf{a}) + \varepsilon \ , \tag{4.3}$$

con $M \geq N$.

En el modelo generativo, la ecuación (4.3) toma la forma:

$$\mathbf{x} = \sum_{\gamma \in \Gamma} \phi_\gamma a_\gamma + \varepsilon = \mathbf{\Phi a} + \varepsilon, \tag{4.4}$$

con el diccionario $\mathbf{\Phi}$ siendo una colección átomos $(\phi_\gamma)_{\gamma \in \Gamma}$ y $\Gamma$ un conjunto finito de números naturales. La Figura 4.1 muestra esquemáticamente el modelo generativo que da lugar a la señal, a partir de la combinación lineal de átomos seleccionados del diccionario y la adición de ruido. En la Figura 4.2 se muestra un ejemplo de un diccionario estimado a partir de rasgos encontrados en porciones de imágenes de dígitos manuscritos, junto a la reconstrucción de un dígito de prueba, para una tarea de reconocimiento óptico de caracteres [64].

En el caso general, tanto $\vec{\Phi}$, $\vec{a}$ y $\vec{\varepsilon}$ son desconocidos, teniendo así el problema infinitas soluciones. Incluso en el caso de tratar con una señal limpia $(\vec{\varepsilon} = \vec{0})$ y siendo $\vec{\Phi}$ conocido, si éste es sobrecompleto o sus átomos no forman una base, se tendrán representaciones no únicas de la señal.

### 4.2.4 Obtención del diccionario y la representación

Si se conocieran tanto $\vec{\Phi}$ como $\vec{x}$, una posible manera de encontrar un conjunto óptimo de coeficientes $\vec{a}$ entre las diversas representaciones posibles

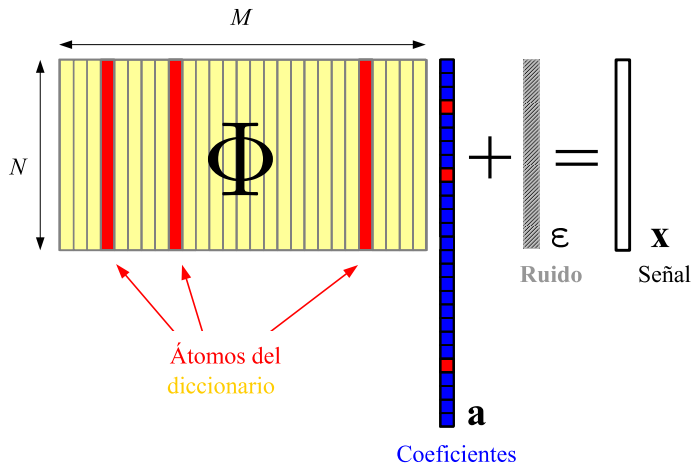



Figura 4.1: Obtención de la señal bajo análisis $\mathbf{x}$ en función de los átomos de un diccionario $\mathbf{\Phi}$ y un conjunto de coeficientes $\mathbf{a}$, considerando la presencia de ruido aditivo $\varepsilon$. En rojo se destacan esquemáticamente unos pocos átomos significativos de la descomposición, una de las características principales de las representaciones ralas. (Modificado de [65])

consiste en realizar la búsqueda de los $a_i$ que obtengan la representación más rala e independiente posible. Para ello, se puede imponer la condición de que la distribución de probabilidad de cada coeficiente tenga kurtosis positiva. Así, esta distribución *a priori* debe satisfacer

$$P(\vec{a}) = \prod_i P(a_i). \tag{4.5}$$

Siguiendo la terminología de uso común en el campo del análisis de componentes independientes (ICA), los $a_i$ pueden ser vistos como un vector de estados (fuentes) que generan la señal a través de una matriz de mezcla $\vec{\Phi}$, con la incorporación de un término de ruido gaussiano aditivo $\vec{\varepsilon}$.

El vector $\vec{a}$ puede, entonces, ser estimado a partir de su distribución de probabilidad *a posteriori* [66]

$$P(\vec{a}|\vec{\Phi}, \vec{x}) = \frac{P(\vec{x}|\vec{\Phi}, \vec{a})P(\vec{a})}{P(\vec{x}|\vec{\Phi})}. \tag{4.6}$$

Una posible estimación del máximo *a posteriori* está dada por

$$\hat{\vec{a}} = \arg\max_{\vec{a}} \left[\log P(\vec{x}|\vec{\Phi}, \vec{a}) + \log P(\vec{a})\right]. \tag{4.7}$$

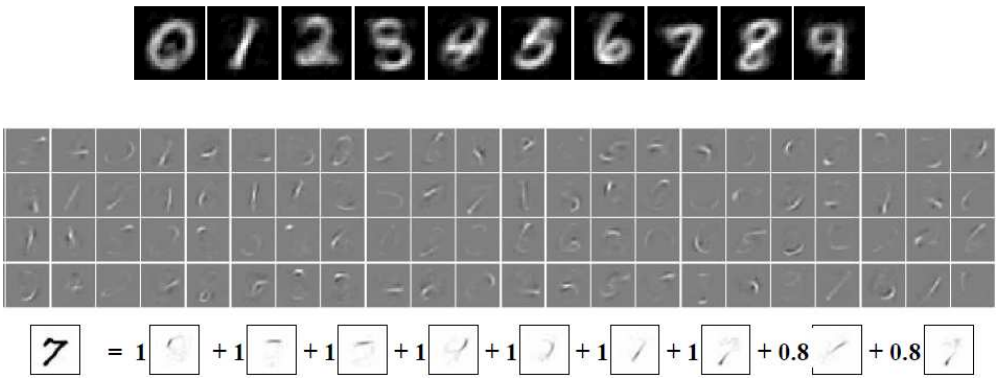



Figura 4.2: Ejemplo de diccionario para la representación de dígitos manuscritos. Arriba: conjunto de imágenes (señales dadas). Abajo: conjunto de átomos del diccionario y reconstrucción del dígito "7" mediante la combinación lineal de los átomos con coeficiente distinto de cero. (Adaptado de [64])

Si $P(\vec{a}|\vec{\Phi},\vec{x})$ es suficientemente suave, el máximo puede hallarse mediante gradiente ascendente. La forma funcional empleada corresponde a la distribución *a priori* laplaciana paramétrica [67]

$$P(a_i) = \alpha\, e^{(-\beta_i|a_i|)}, \tag{4.8}$$

de parámetro $\beta_i$, con $\alpha$ una constante de normalización. Así, la regla de actualización de los coeficientes resulta

$$\Delta\vec{a} = \vec{\Phi}^T\vec{\Lambda}_{\vec{\varepsilon}}\vec{\varepsilon} - \vec{\beta}^T\,|\vec{a}|\,, \tag{4.9}$$

donde $\vec{\Lambda}_{\vec{\varepsilon}}$ es la inversa de la matriz de covarianza $\mathcal{E}\left[\vec{\varepsilon}^T\vec{\varepsilon}\right]$, siendo $\mathcal{E}[\cdot]$ la función valor esperado.

Luego, la estimación del diccionario $\vec{\Phi}$ puede realizarse maximizando la función objetivo [67]

$$\hat{\vec{\Phi}} = \arg\max_{\vec{\Phi}} \left[\mathcal{L}(\vec{x},\vec{\Phi})\right], \tag{4.10}$$

donde $\mathcal{L} = \mathcal{E}\Big[\log P(\vec{x}|\vec{\Phi})\Big]_{P(\vec{x})}$ es la verosimilitud de los datos. Esta función puede calcularse al marginalizar el producto de la distribución condicional de los datos, conociendo el diccionario y la distribución *a priori* de los coeficientes

$$P(\vec{x}|\vec{\Phi}) = \int_{\mathbb{R}^M} P(\vec{x}|\vec{\Phi},\vec{a})P(\vec{a})\,d\vec{a}, \tag{4.11}$$

con la integral calculada sobre el espacio $M$-dimensional de $\vec{a}$.

Finalmente, la maximización de (4.10) se obtiene aplicando el gradiente ascendente con la siguiente regla de actualización para $\vec{\Phi}$

$$\Delta\vec{\Phi} = \eta\vec{\Lambda}_{\varepsilon}\ \mathcal{E}\left[\vec{\varepsilon}\vec{a}^{T}\right]_{P(\vec{a}|\vec{\Phi},\vec{x})}, \tag{4.12}$$

con tasa de aprendizaje $\eta \in (0,1)$. De esta manera, el diccionario $\vec{\Phi}$ y el conjunto óptimo de coeficientes $\vec{a}$ son obtenidos en un proceso iterativo.

## 4.3 Representación auditiva cortical aproximada

### 4.3.1 Campos receptivos espectro-temporales

En el sistema nervioso, la codificación de los estímulos de diferente naturaleza (visuales, táctiles, etc.) se realiza mediante la generación de trenes de impulsos en las fibras nerviosas, las cuales viajan desde los centros sensitivos hasta la corteza cerebral. En el caso del sistema auditivo, es el oído interno –a través de la cóclea– el encargado de realizar un primer análisis de la señal acústica de entrada obteniendo una descomposición tiempo-frecuencia que va a ser codificada por el nervio auditivo y enviada hacia la corteza cerebral, en este caso la corteza auditiva primaria. Actualmente se dispone de modelos matemáticos que permiten estudiar la representación fisiológica obtenida a este nivel, de los cuales se tomará el modelo desarrollado por Shamma [68].

El *espectrograma auditivo* en una representación coclear interna de los patrones de vibraciones a lo largo de la membrana basilar. Esta representación se implementa por medio de un banco de 128 filtros cocleares que procesan la señal temporal $s(t)$, obteniendo las salidas que asemejan las encontradas en los potenciales de las células ciliadas

$$x^{k}_{\mathrm{ch}}(t,f) = s(t) * h^{k}_{\mathrm{ch}}(t,f), \tag{4.13}$$

donde $h^{k}_{\mathrm{ch}}$ es la respuesta al impulso del $k$-ésimo filtro. Estas salidas se transducen a patrones del nervio auditivo mediante

$$x^{k}_{\mathrm{an}}(t,f) = g_{\mathrm{hc}}\left(\partial_t x^{k}_{\mathrm{ch}}(t,f)\right) * \mu_{\mathrm{hc}}(t), \tag{4.14}$$

donde $\partial_t$ representa el acoplamiento fluido-cilia (efecto de filtro pasa-alto), $g_{\mathrm{hc}}$ es la compresión no lineal de los canales iónicos y $\mu_{\mathrm{hc}}$ representa la pérdida

en la membrana de la célula ciliada (efecto de filtro pasa-bajo). Finalmente, la red de inhibición lateral es aproximada por una derivada de primer orden (con rectificación de media onda) respecto al eje frecuencial como

$$x_{\text{lin}}^{k}(t,f) = \max\left(\partial_f x_{\text{an}}^{k}(t,f), 0\right), \tag{4.15}$$

y la salida final consiste en la integración de esta señal sobre ventanas móviles de tiempo corto.

Por otra parte, las técnicas actuales de generación de imágenes médicas funcionales permiten registrar la localización de la activación correspondiente a grupos de neuronas. Es así que en experimentos fisiológicos se encontraron evidencias de que la localización de las activaciones en la corteza varía para diferentes estímulos, en forma localizada en tiempo y frecuencia. De este hecho surge la idea de que a partir del espectrograma auditivo resultaría de interés la estimación de un diccionario bidimensional de patrones tiempo-frecuencia de activación cortical[69], denominados *campos receptivos espectro-temporales* (STRF, del inglés *spectro-temporal receptive fields*). Los mismos se estudiaron mediante correlación inversa, obteniéndonse la respuesta en la corteza auditiva frente a estímulos de diferente naturaleza: desde patrones simples como tonos puros hasta complejos como ruido modulado, ondas móviles o vocalizaciones naturales [70].

En el contexto de las representaciones ralas, los STRF actuarían como detectores de caraterísticas o rasgos significativos de la señal analizada, obteniendo una sobre-representación a nivel cortical. Así, el cerebro aumentaría la eficiencia de la representación al remover redundancia estadística y aumentar la independencia entre respuestas a estímulos naturales. La Figura 4.3 muestra algunos STRF de ejemplo obtenidos en experimentos fisiológicos con animales. En la misma se utilizó como estímulo una serie de combinaciones lineales de tonos de frecuencia variable, siendo el tiempo medido a partir de la aplicación del éstos. En rojo se observa la alta respuesta con localización específica temporal y frecuencial, correspondiente a una activación altamente rala en la corteza auditiva, al variar los estímulos.

### 4.3.2 Método propuesto

La aproximación propuesta en esta tesis se basa en la obtención de un diccionario de átomos bidimensionales $\vec{\Phi}$ usando (4.10) correspondientes a rasgos en el plano tiempo-frecuencia estimados a partir del espectrograma auditivo de $\vec{x}$.

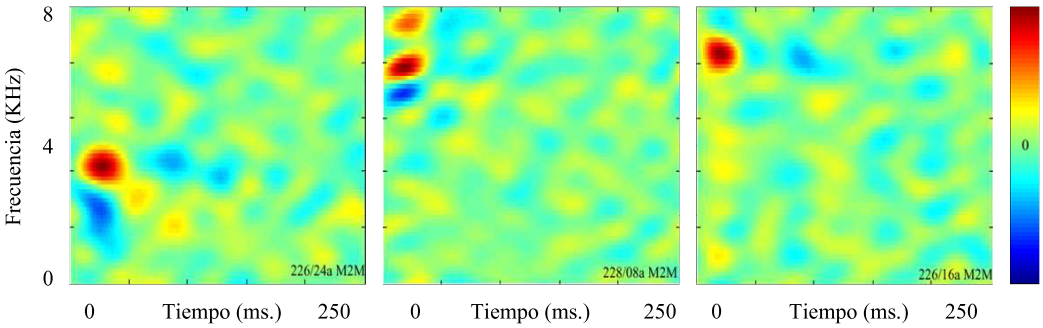



Figura 4.3: Ejemplos de STRF obtenidos en células de la corteza auditiva de hurones por correlación inversa. (Tomado de [71, 70])

Primeramente, dado un conjunto de señales acústicas, se obtiene el espectrograma a nivel auditivo de cada una mediante el modelo de oído descripto previamente. A continuación se estiman el diccionario de "parches" o secciones del espectrograma, tratando de capturar las particularidades de las transiciones presentes, junto con el cálculo iterativo de las activaciones a fin de lograr una representación lo más rala posible. Una vez finalizado este proceso, ante una señal de prueba dada se calcula su espectrograma auditivo y se hallan los coeficientes de la representación usando el diccionario previamente entrenado.

Así, finalmente se obtiene la denominada *representación auditiva cortical aproximada* (AACR, del inglés *approximated auditive cortical response*), donde el nivel de activación de cada neurona o grupo de neuronas puede ser tratado como la activación de los coeficientes $a_\gamma$ en (4.4). La Figura 4.4 muestra un diagrama esquemático del método planteado para obtener la AACR.

A fines ilustrativos, se muestra en la Figura 4.5 una selección de átomos obtenidos de señales de habla correspondientes a un diccionario completo ($\vec{\Phi} \in \mathbb{R}^{256\times256}$). Se observan algunos comportamientos típicos que serían de utilidad para la clasificación, ya que aparecen ciertas características fonéticas significativas como detección de formantes, frecuencias únicas, componentes sordas o ruidosas, entre otros. En rojo se destacan dos átomos que se comparan con STRF obtenidos en experimentos con animales (a la izquierda de la Figura), haciendo evidente las similitudes entre el comportamiento fisiológico y el modelado.

El esquema mostrado será el empleado en la experimentación con señales artificiales para ilustrar la técnica y con señales de habla en el próximo capítulo. Además, se explicarán algunas modificaciones introducidas en el cálculo de las activaciones oportunamente.

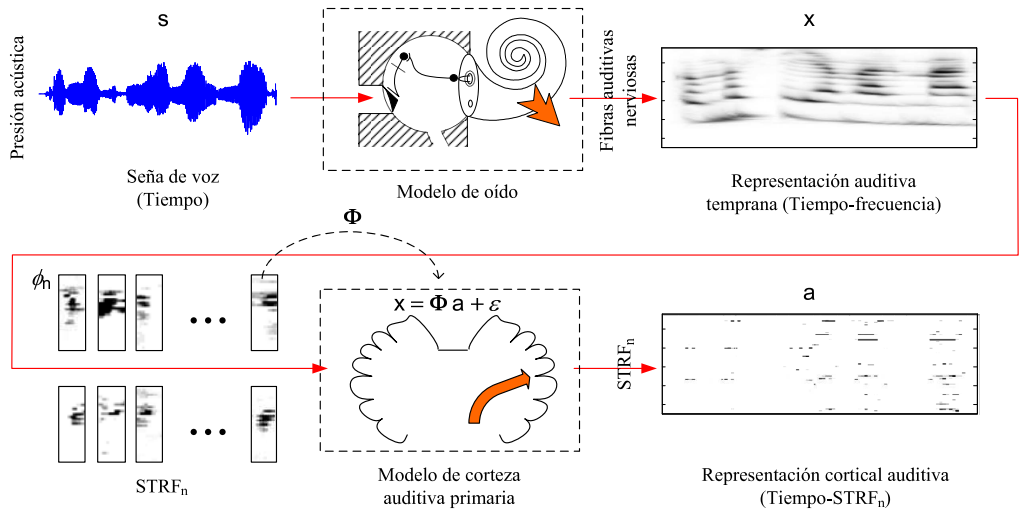



Figura 4.4: Diagrama del método general para la obtención de la representación auditiva cortical aproximada.

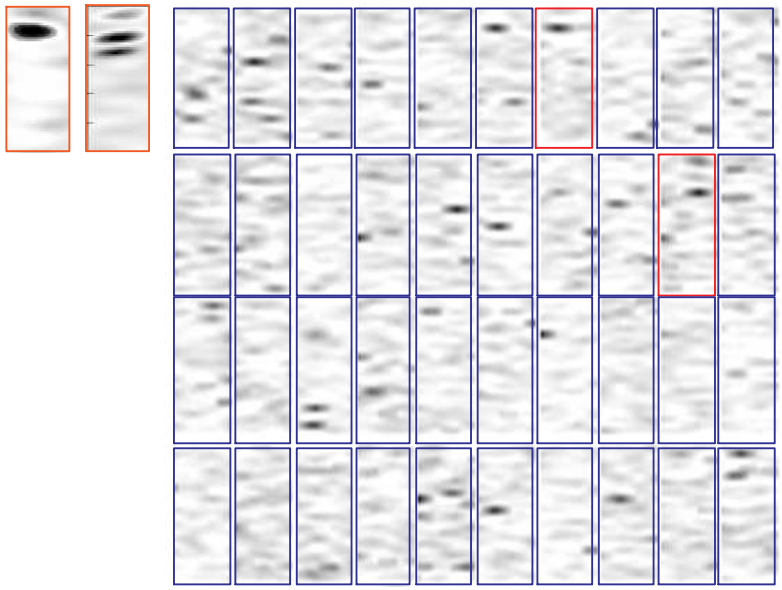


Figura 4.5: Ejemplo de STRF estimados a partir de señales de habla, en particular de segmentos de fonemas. Izquierda: dos STRF encontrados en experimentos fisiológicos con animales, que se comparan (en rojo) con átomos similares encontrados por el método planteado.

### 4.3.3 Resultados con señales artificiales

En esta sección se detallan las condiciones de experimentación con señales temporales artificiales formadas por la combinación de tonos puros, chirps ascendentes y descendentes, para la obtención de representaciones corticales aproximadas. Se estudian aspectos cualitativos y cuantitativos de los diccionarios obtenidos, a fin de obtener conclusiones sobre las características extraídas.

Todos los experimentos fueron realizados en las siguientes condiciones:

- Frec. de muestreo: 8/16 Khz.
- Señales formadas por la concatenación de 7 segmentos de 50 ms c/u, con patrones variables en cada segmento[2].
- 1000 señales en el corpus de entrenamiento.
- Obtención del espectrograma auditivo y entrenamiento del diccionario con 500 iteraciones del algoritmo de aprendizaje.

Como patrones artificiales creados fueron empleados: señales chirp con variación ascendente y/o descendente de la frecuencia y tonos puros. Las frecuencias inicial/final de las chirp, así como la de los tonos, fueron variadas aleatoriamente dentro del rango 0-8 kHz.

Luego del entrenamiento se analizó cualitativamente el diccionario obtenido. La Figura 4.6 muestra un ejemplo del mismo, en donde pueden observarse, ordenados por similitud, cómo los diferentes átomos van capturando las características presentes en los patrones. Para ejemplificar, si se observa la segunda fila puede verse que los primeros átomos capturaron las chirps descendentes en frecuencias centrales, mientras que los últimos de la misma fila corresponden a tonos puros en el mismo rango frecuencial. En la primera fila se observa que los primeros átomos son aquéllos que no lograron especializarse en alguna característica particular, resultando esa dispersión típica de alta frecuencia similar a un ruido.

Un segundo experimento consistió en generar señales artificiales con las siguientes características:

- Se divide en dos franjas al patrón: entre 0 y 1000 Hz, y entre 1000 y 7000 Hz. En cada una se coloca (aleatoriamente) un tono puro o una chirp lineal ascendente o descendente (aleatoriamente).

[2]En el resto de la sección, el término *patrón* designa al STFT del segmento generado.

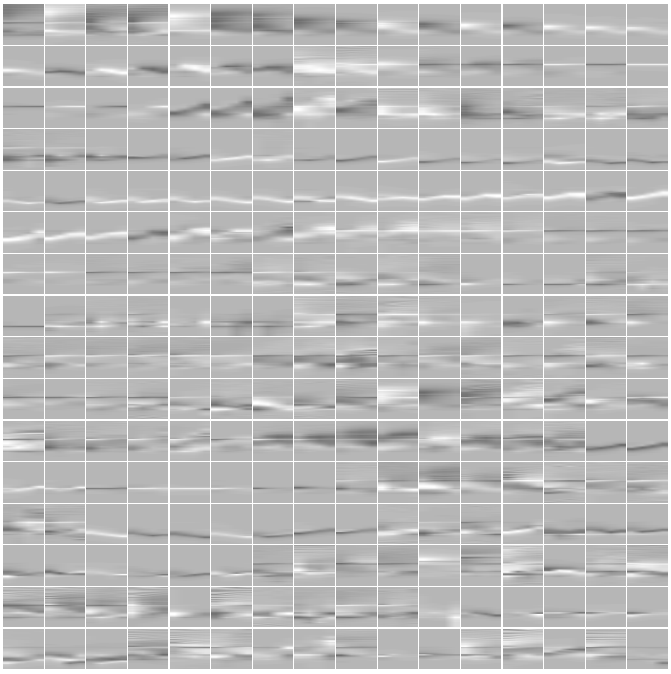


Figura 4.6: Diccionario estimado a partir de tonos puros y chirps con frecuencia máxima 8 kHz.

- Las frecuencias inicial y final de las chirps se eligen al azar en la franja correspondiente, al igual que la frecuencia del tono puro.

La Figura 4.7 muestra el diccionario obtenido a partir de los nuevos patrones construidos. En este caso, y al igual que en el caso anterior, en la primera fila pueden observarse algunos átomos que luego de finalizado el entrenamiento responden sólo esbozando las características temporales de las señales pero sin llegar a ser bien aprendidos. Otros átomos, en cambio, evidencian más definidas ya sea en líneas claras u oscuras –respecto al gris medio de fondo– las características frecuenciales usadas para generar los patrones, por ej: chirp descendente superior y ascendente inferior (cuarta fila, cuarto átomo), chirp descendente superior y tono puro inferior (sexta fila, cuarto átomo).

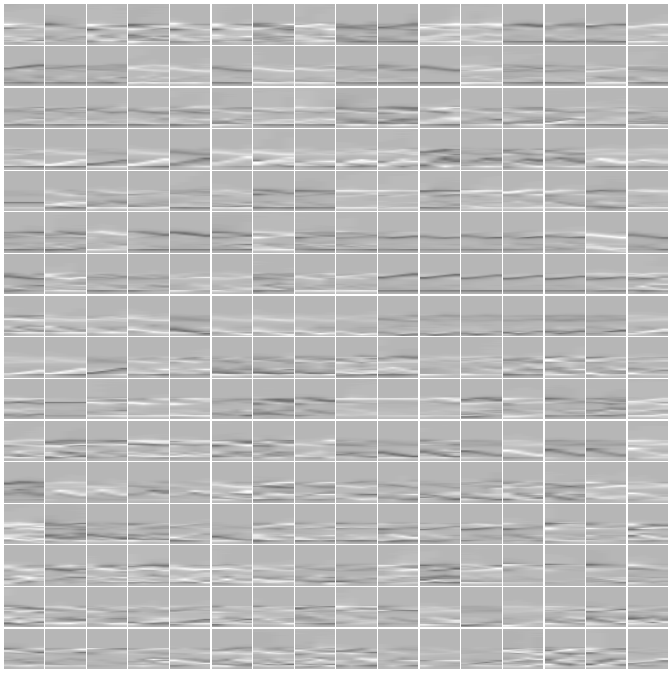


Figura 4.7: Diccionario estimado a partir de la combinación de tonos puros y chirps sobre el mismo segmento temporal.

## 4.4 Representación cortical auditiva no negativa

### 4.4.1 Generalidades de la técnica

En esta sección se propone un método basado en el anterior, pero en un marco de Factorización matricial No Negativa (FNN). Una representación cortical auditiva no negativa es usada con el objeto de proponer un nuevo algoritmo de limpieza de ruido de la señal sonora. La FNN es una familia de técnicas recientemente desarrolladas para encontrar representaciones lineales por tramos en datos no negativos [72], como los espectrogramas auditivos aquí trabajados. Así, los datos son descriptos usando sólo componentes aditivas, esto es, una suma pesada de átomos de STRF sólo positivos. Este nuevo modelo todavía retiene su analogía biológica, a pesar de que los STRF positivos implican la interacción de comportamiento no inhibitorio solamente. Por lo tanto, los coeficientes positivos podrían ser interpretados como impulsos de

disparo de las neuronas corticales excitatorias. El algoritmo de limpieza en el dominio cortical auditivo propuesto toma ventaja del diccionario STRF (sobre)completo que combina los átomos estimados a partir de señales limpias y ruidosas.

### 4.4.2 Algoritmo K-SVD para representación rala no negativa

La representación de una señal $\mathbf{x} \in \mathbb{R}^N$ está dada por la combinación lineal de átomos encontrados por el modelo auditivo, en la forma

$$\mathbf{x} = \mathbf{\Phi}\mathbf{a}, \tag{4.16}$$

donde $\mathbf{\Phi} \in \mathbb{R}^{N \times M}$ es el diccionario de $M$ átomos y $\mathbf{a} \in \mathbb{R}^M$ representa $\mathbf{x}$ en términos de $\mathbf{\Phi}$. La raleza es incluída cuando la solución es restringida a

$$\min_a \|\mathbf{a}\|_0, \tag{4.17}$$

donde $\|\cdot\|_0$ es la norma $l^0$ que cuenta el número de entradas diferentes de cero del vector.

Con el objeto de encontrar la representación requerida, dos problemas deben ser resueltos de manera conjunta: la estimación de la representación rala y la inferencia de un diccionario especializado. Los coeficientes encontrados con métodos tales como Basis Pursuit o MP dan ambos átomos y activaciones con valores positivos y negativos [73, 74]. Sin embargo, en algunas aplicaciones podría ser útil trabajar solamente con valores positivos, dándole al método la capacidad de explicar los datos a partir de la suma controlada de átomos positivos. Este es el objetivo de los métodos de factorización no negativa.

Aharon *et al* introdujeron al K-SVD como una generalización del algoritmo de clustering *k-medias* para resolver el problema de representación, con una aproximación basada en la descomposición en valores singulares –de ahí el nombre del método– [75]. Mas aún, ellos incluyeron una versión no negativa (NN) del algoritmo Basis Pursuit (abreviado como NN-BP), para producir diccionarios no negativos. El método resuelve el problema

$$\min_a \|\mathbf{x} - \mathbf{\Phi}^L\mathbf{a}\| \quad s.t.\ \mathbf{a} \geq 0, \tag{4.18}$$

donde una submatriz $\mathbf{\Phi}^L$ que incluye sólo una selección de los $L$ coeficientes más grandes es usada. El paso de actualización del diccionario fuerza a esta

matriz a ser positiva, mediante el cálculo de

$$\min_{\phi_k, a^k} \|\mathbf{E}^k - \phi_k a^k\| \quad s.t.\ \phi_k, x^k \geq 0, \tag{4.19}$$

donde $\mathbf{E}^k$ es la matriz de error. El algoritmo final fue denominado NN-K-SVD [75].

### 4.4.3 Método propuesto

La idea principal es que las señales de sonido y ruido pueden ser proyectadas a un espacio cortical auditivo aproximado, donde las características significativas de cada tipo de señal puedan ser fácilmente separadas. Las señales bajo análisis pueden ser descompuestas en más de un diccionario (posiblemente sobrecompleto) que contenga una aproximación gruesa a todas las características de interés. Más precisamente, el método propuesto aquí está basado en la descomposición de la señal en dos diccionarios paralelos de STRF, uno de ellos estimado a partir de señales limpias y el otro a partir de señales de ruido. La estimación de ambos diccionarios se lleva a cabo luego de obtener los respectivos espectrogramas auditivos tempranos.

Teniendo en cuenta que este tipo de representación es esencialmente no negativa, una forma natural para obtener tanto el diccionario y las activaciones corticales es utilizar un algoritmo como el NN-K-SVD previamente expuesto. Esto es especialmente cierto en el caso de las aplicaciones de limpieza de ruido, donde forzar a la no negatividad tanto en el diccionario y los coeficientes puede ayudar a encontrar los bloques de construcción de las señales [75].

La Figura 4.8 muestra un diagrama con las dos etapas del método propuesto, donde ambos diccionarios fueron estimados previamente. La etapa *hacia delante* del método produce las activaciones corticales auditivas aproximadas que mejor representan la señal ruidosa (que incluye tanto las activaciones limpias y ruidosas) por medio de la versión no-negativa del algoritmo BP. Luego, en la etapa *hacia atrás*, el espectrograma auditivo se reconstruye tomando la transformada inversa de sólo los coeficientes correspondientes al diccionario de señal, descartando los coeficientes de ruido. De esta manera, la limpieza de ruido de la señal se lleva a cabo en el dominio cortical auditivo aproximado. Por último, se obtiene la señal limpiada en el dominio temporal por el modelo inverso del oído. El método propuesto se denomina *limpieza de ruido cortical no negativa* (NCD, del inglés *non-negative cortical denoising*).

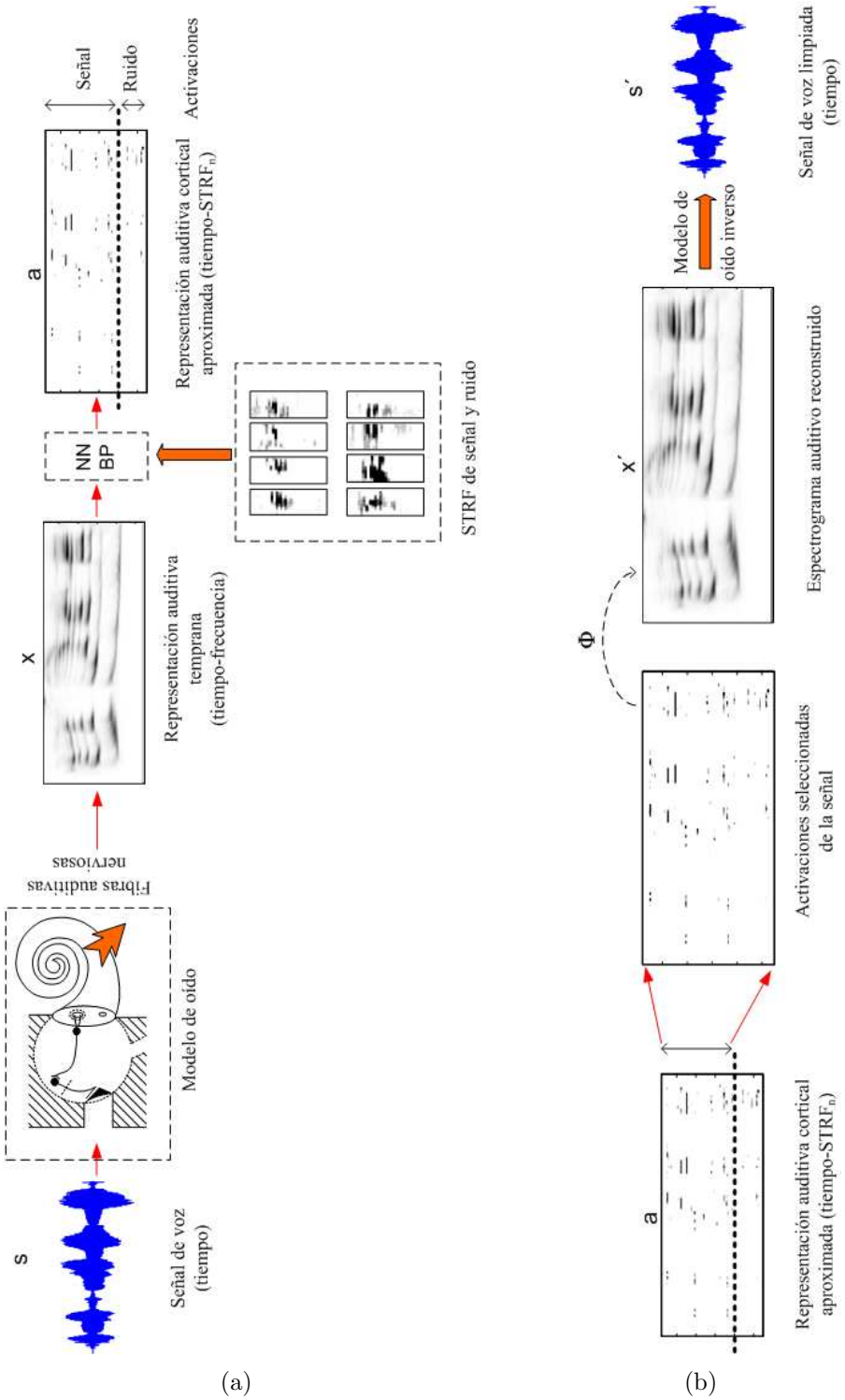



Figura 4.8: Diagrama del método NCD propuesto para limpieza de ruido en el dominio cortical. (a) Etapa hacia adelante: representación cortical. (b) Etapa hacia atrás: reconstrucción limpiada.

La reconstrucción del espectrograma auditivo a partir de la respuesta cortical es directa, ya que sólo consiste en una transformación lineal. Sin embargo, una reconstrucción perfecta de la señal original a partir del espectrograma auditivo es imposible de obtener debido a las operaciones no lineales de la etapa temprana. No obstante, las pruebas de calidad objetiva y subjetiva mostraron que la inteligibilidad resultante no es degradada [76].

La idea de utilizar un modelo cortical para la eliminación de ruido de sonido también fue propuesta por Shamma en un trabajo reciente [77]. Las principales diferencias con el enfoque aquí expuesto son: 1) su representación cortical utiliza el concepto de modulación espectrotemporal en vez de las representaciones ralas, y 2) la manera en que se incorpora la información sobre la señal y el ruido al modelo.

### 4.4.4 Experimentos y resultados

Una serie de pruebas se llevaron a cabo para demostrar las capacidades de la técnica propuesta. Una primera serie de experimentos fue primeramente realizada sobre señales artificiales limpias, generadas mediante una mezcla de señales chirp y tonos puros. Se agregaron ruidos con distribución frecuencial diferente, de forma aditiva y con relaciones señal-ruido (SNR) variable. La técnica propuesta se aplicó para obtener las señales limpiadas y el desempeño fue evaluado por un método objetivo (puntuación PESQ).

#### Señales de prueba y ruido

Un total de 1000 señales artificiales fueron obtenidas mediante la concatenación de 7 segmentos diferentes de 64 ms cada uno, con una frecuencia de muestreo de 8 kHz. Cada segmento consistió en la combinación al azar de chirps ascendentes o descendentes y tonos puros. Con el fin de restringir todas las posibles combinaciones de estas características, de manera que un diccionario relativamente simple fuera capaz de representarlas, el espectrograma se dividió en dos zonas de frecuencia, por debajo y por encima de 1200 Hz. Dentro de cada zona, sólo una de las características podía estar presente. Además, las pendientes de las chirps estaban fijas dentro de cada zona.

Para ensuciar las señales, dos clases de ruido con contenido frecuencial diferente fueron empleados. Por un lado, el ruido blanco tomado de la base de datos Noisex-92 **??**, que presenta un contenido de frecuencia relativamente alta con una distribución no uniforme en el espectrograma auditivo temprano (debido a su escala de frecuencia logarítmica); y por otro lado los ruidos

murmullo de voz de Noisex-92 y ruidos de calle de la base de datos Aurora [78]; ambos con contenido principal de bajas frecuencias en esa representación .

### Representación cortical

Los espectrogramas auditivos de las señales limpias fueron calculados y los datos de entrenamiento para la estimación de los diccionarios fueron extraídos de una serie de ventanas móviles tiempo-frecuencia, sin solapamiento. Las mismas consideraciones se aplican a la estimación de los diccionarios de ruido.

Los diccionarios se generaron con 512 átomos de tamaño de $64 \times 8$ (diccionarios completos). Aquí, los 64 coeficientes corresponden a una versión submuestreada de los 128 coeficientes originales que representan el rango de 0-4 kHz. Las 8 columnas corresponden cada una a una ventana de 8 ms. De cada diccionario, los átomos más activos fueron recolectados y combinados para formar diccionarios con 256 átomos conteniendo características de señal limpia y de ruido.

### Medida de calidad

La puntuación PESQ es una medida objetiva de calidad establecida por la Unión Internacional de Telecomunicaciones (UIT) como un estándar para la evaluación de la calidad de la voz después de la transmision a través de canales de comunicación [79]. La medida utiliza una representación auditiva basada en la escala Bark para comparar las señales de voz original y distorsionada.

La medida se calcula por ventanas, después de la compensación de alineación del desplazamiento de la ventana y de la ganancia, por lo tanto el método no es sensible a retardos variables en el tiempo y al escalado. Después de esta compensación las señales se comparan en el dominio auditivo, utilizando algunos modelos cognitivos para pesar no linealmente las diferencias y por lo tanto producir dos diferencias tiempo-frecuencia ponderadas perceptualmente denominadas “densidad de perturbación” y “densidad de perturbación asimétrica”. El primero tiene en cuenta los umbrales de enmascaramiento del oído humano ($D$), y el segundo la cantidad de contenido frecuencial que se introduce por el método de transmisión ($A$, manifiestado como ruido musical).

Estas dos densidades tiempo-frecuencia son integradas en frecuencia usando diferentes normas $p$, y luego en el tiempo en grupos de 20 ventanas usando una norma $L_6$ y luego otra vez con una norma $L_2$, para producir dos valores individuales, uno para el índice de perturbación $D$ y otro para el índice

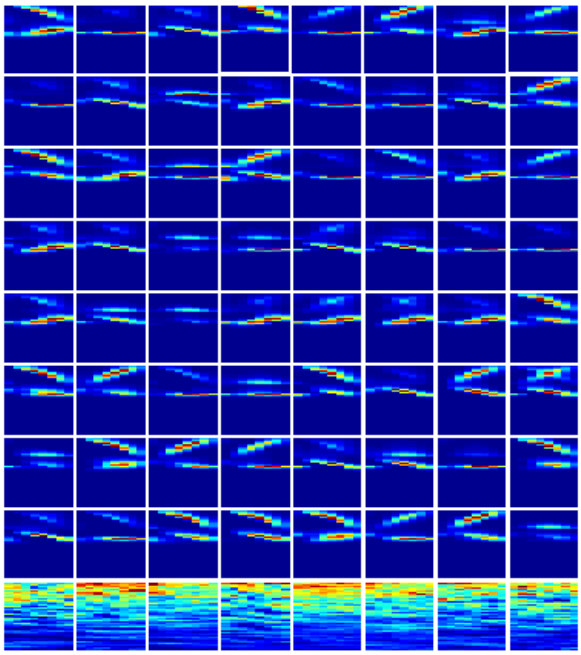


Figura 4.9: Ejemplo de campos receptivos espectro-temporales (STRF) calculados a partir de la representación auditiva de señales artificiales y ruido blanco, mostrando los átomos más activos de cada diccionario. Las 8 filas superiores muestran los 64 STRF más importantes de las señales limpias, mientras que la última fila muestra los respectivos STRF de las señales de ruido.

asimétrico $A$. Estos dos son combinados para producir un solo valor, denominado puntuación PESQ, que se define como $4{,}5 - \alpha D - \beta A$, con $\alpha = 0{,}1$ y $\beta = 0{,}0309$. La medida tiene un valor ideal de 4,5 para señales limpias sin distorsión, y un mínimo de $-0{,}5$ para el peor de los casos de distorsión. Esta medida ha demostrado tener una muy buena correlación con pruebas perceptuales usando MOS [80, 81].

### Diccionarios de átomos no-negativos

La Figura 4.9 muestra una selección de un diccionario, donde los 64 átomos más activos para las señales chirp (8 filas superiores) y 8 átomos de señales de ruido blanco (fila inferior) son mostrados. Se puede ver claramente que las características captadas por el STRF correspondiente a cada diccionario son, respectivamente, la combinación de chirps y tonos puros, y las características del ruido que son más prominentes en las señales de entrenamiento.

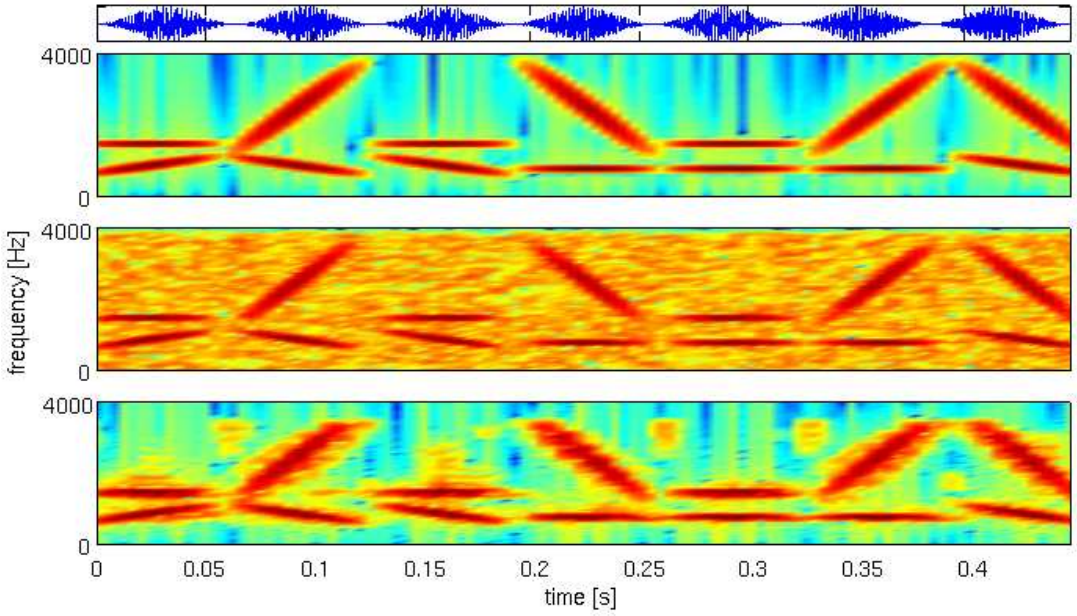



Figura 4.10: Ejemplo de la limpieza de una señal artificial formada por una combinación de 7 segmentos ventaneados de chirps aleatorias y tonos puros. Se muestran los espectrogramas (STFT) de la señal limpia (arriba), una versión ruidosa obtenida mediante la adición de ruido blanco a SNR=0 dB (centro) y la señal limpiada (abajo), con la señal temporal en la parte superior como referencia.

### Limpieza de señales artificiales

El esquema de limpieza de ruido fue aplicado usando esta representación y la reconstrucción del espectrograma auditivo limpiado fue obtenido seleccionando sólo los átomos limpios de los 32 con mayor activación, tal como los obtiene el algoritmo NN-BP. La Figura 4.10 muestra la transformada de Fourier de tiempo corto (STFT) para una señal limpia (arriba), señal ruidosa obtenida al adicionar ruido blanco con SNR=0 dB (centro) y la señal limpiada (abajo). La señal acústica se grafica por encima del espectrograma limpio como referencia temporal. En el espectrograma muestrado en la parte inferior, los efectos de la limpieza de ruido realizada en la representación cortical por el NCD pueden ser vistos, donde las características más importantes son reconstruídas.

La Tabla 4.1 muestra los resultados obtenidos para la puntuación PESQ sobre señales artificiales limpiadas. En todos los casos hubo un aumento en la PESQ cuando el NCD se aplicó a las señales ruidosas. La mejora fue más

Tabla 4.1: Puntuación PESQ obtenida para señales artificiales.

| Ruido | SNR (dB) | Señal | |
|---|---|---|---|
| | | Ruidosa | Limpiada |
| Blanco | 12 | 1,93 | 2,16 |
| | 6 | 1,40 | 2,11 |
| | 0 | 0,69 | 1,99 |
| Murmullo | 12 | 1,82 | 2,05 |
| | 6 | 1,23 | 2,01 |
| | 0 | 0,56 | 1,91 |

marcada cuando la energía del ruido fue mayor (SNR=0 dB) y menor cuando las señales son cada vez más limpias a mayor SNR (baja energía del ruido).

La puntuación PESQ para la señal original (limpia) después de la transformación mediante el modelo auditivo y la reconstrucción posterior al dominio temporal es 2,11. Esta puntuación mide la distorsión de la mejor calidad (PESQ MOS de 4,5) que se introduce por el uso del modelo auditivo temprana, el cual es sólo aproximadamente invertible. Incluso si el ruido es eliminado por completo por el NCD, existe un error intrínseco introducido por el método de análisis auditivo.

## 4.5 Comentarios de cierre de capítulo

En este capítulo se expusieron las nociones básicas de la representación empleada en este trabajo como alternativa a la clásica utilizada en el campo del RAH: las representaciones ralas. En particular, se introdujo el método para el cálculo de la representación aproximada (AACR) basado en un diccionario de átomos espectro-temporales y una versión no-negativa enfocada a la limpieza de ruido. En el Capítulo siguiente se experimenta la técnica con señales de habla, mostrando las capacidades de capturar las pistas acústicas más importantes de fonemas que permitan su clasificación, incluso en condiciones de contaminación con ruido.

CAPÍTULO 5

# Clasificación robusta de fonemas

En este Capítulo se presentan el marco experimental y los resultados obtenidos en la aplicación de la técnica de representación rala propuesta en el Capítulo previo, en una tarea inicial del RAH, la clasificación de fonemas altamente confundibles.

La aproximación propuesta, con todas sus variantes algorítmicas en la obtención de las activaciones a nivel cortical, son evaluadas en condiciones de habla limpia y con señales inmersas en ruido de diverso tipo agregado aditivamente.

## 5.1 Consideraciones preliminares

A fin de validar el método propuesto como técnica intrínsecamente robusta de parametrización de señales, se presenta la experimentación sobre una tarea inicial dentro del reconocimiento automático del habla (RAH). Específicamente, se llevó a cabo la conformación y evaluación de clasificadores de *fonemas*, los cuales son las unidades lingüísticas mínimamente identificables con rasgos distintivos y cuya secuencia forman las palabras [82].

En los últimos años, diversos esfuerzos han sido realizados a fin de proveer robustez al modelado acústico de los fonemas, mediante la propuesta de diferentes aproximaciones en la representación de la voz. En [83, 84, 85], los autores usan el segmento central de las señales acústicas de cada fonema. Ellos mostraron que la divergencia entre las condiciones de habla limpia y ruidosa es mejor manejada por una representación de este tipo que los coeficientes de predicción lineal perceptual, especialmente bajo degradación severa. Recientemente, una técnica de compensación de ruido fue propuesta para suprimir el efecto del ruido aditivo con una estimación de la envolvente del ruido [86]. Estos trabajos, al igual que las técnicas clásicas de parametrización, son llevados a cabo procesando la señal en su dominio temporal y/o frecuencial exclusivamente. La aproximación aquí planteada trata de sortear las limitaciones derivadas de emplear esta señal acústica como entrada al sistema.

Las señales de voz limpia fueron obtenidas del corpus en inglés TIMIT, ampliamente difundido en investigaciones en el área del RAH [87]. Estas señales fueron contaminadas con distintos tipos de ruido mediante un proceso de adición controlada. Las particularidades de estos corpus se detallan en el Apéndice A.

## 5.2 Marco experimental

### 5.2.1 Tarea de clasificación de fonemas

La tarea inicial que se diseñó para la experimentación y validación de la técnica presentada consiste en la clasificación de fonemas. Se eligió un

Tabla 5.1: Distribución de los patrones de los 5 fonemas para entrenamiento y prueba de los clasificadores.

| Fonema | Entrenamiento | | Prueba | |
|---|---|---|---|---|
| | # | (%) | # | (%) |
| */b/* | 211 | (3.26) | 66 | (3.43) |
| */d/* | 417 | (6.45) | 108 | (5.62) |
| */jh/* | 489 | (7.56) | 116 | (6.04) |
| */eh/* | 2753 | (42.58) | 799 | (41.63) |
| */ih/* | 2594 | (40.13) | 830 | (43.25) |
| Total | 6464 | (100.00) | 1919 | (100.00) |

grupo de fonemas en inglés que tienen la particularidad de ser altamente confundibles y que fuera usada en otros trabajos dentro del campo del reconocimiento automático del habla. El grupo se denomina *E-set*, por provenir de palabras correspondientes al alfabeto inglés que tienen como segunda letra a "e" (como "be", "de", "ge", etc.). La baja energía y corta duración de las consonantes en relación a la vocal hacen que sea un conjunto difícil de clasificar. El conjunto tiene los fonemas */b/*, */d/*, */jh/*, */eh/*, */ih/* [88].

La Tabla 5.1 muestra las cantidades relativas de cada fonema del *E-set* encontradas en el corpus TIMIT y su división en conjuntos de entrenamiento y prueba. Puede observarse un desbalance notable en la distribución dada la diferente longitud de cada uno, lo cual puede ser contraproducente a las capacidades de generalización de los clasificadores. Por lo tanto, los conjuntos de entrenamiento y prueba usados en este trabajo fueron balanceados seleccionando el mismo número de patrones para cada clase en ambos conjuntos (211 and 66 patrones, respectivamente).

### 5.2.2 Aspectos de implementación

Las emisiones se encuentran muestreadas a 16 KHz, para cada una de ellas se calculó el espectrograma auditivo mediante un modelo auditivo [89]. Luego, la resolución frecuencial de los datos fue reducida a fin de disminuir la cantidad de dimensiones, obteniéndose espectrogramas auditivos de 64 coeficientes frecuenciales por unidad de tiempo. Finalmente, por medio de una ventana deslizante de 32 ms. desplazada a intervalos de 8 ms. (solapamiento del 75 %), se obtuvo el conjunto de patrones espectro-temporales para la estimación de los diccionarios.

La Figura 5.1 muestra las principales señales obtenidas en el proceso des-

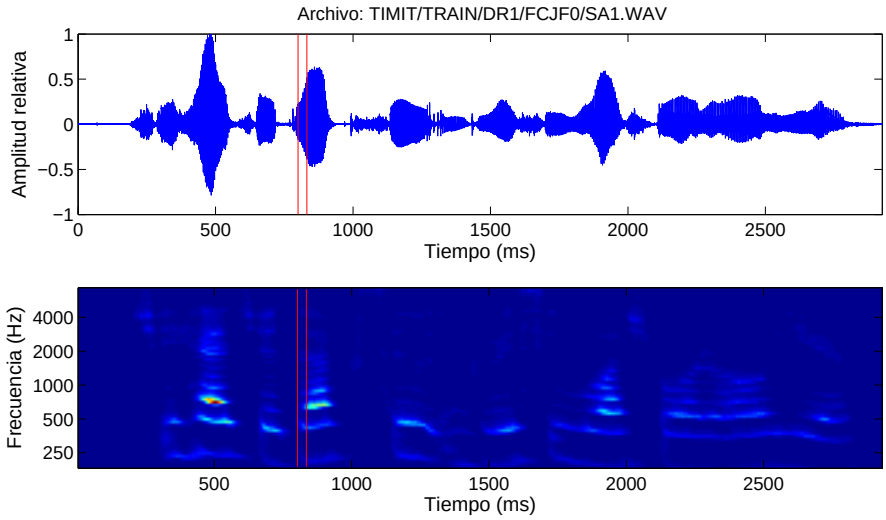



Figura 5.1: Señales principales generadas en el proceso de obtención de los patrones espectro-temporales: sonograma (arriba) y espectrograma submuestreado (abajo). Una sección correspondiente a la ventana móvil de la cual se extraen los patrones espectro-temporales ha sido marcada en rojo.

cripto basado en el algoritmo BP: un sonograma correspondiente a la primer frase del corpus TIMIT junto a su espectrograma auditivo submuestreado por 2 (64 coeficientes frecuenciales). Se resaltan en rojo, además, los límites temporales de una ventana móvil que se utiliza para la extracción de los patrones que luego serán empleados en la obtención de los diccionarios.

Por su parte, en la Figura 5.2 se muestra un extracto de la señal limpia y sus correspondientes espectrogramas auditivos para los 5 fonemas experimentados. Aquí, los fonemas */b/* and */d/* tienen longitud menor que el tiempo requerido para calcular los patrones espectro-temporales, por lo que las señales se rellenan con ceros al comienzo y al final. Los espectrogramas muestran en rojo las amplitudes máximas, resaltando las características de alta frecuencia en los fonemas cortos dada su distribución en todo el rango frecuencial. En los fonemas más largos, por otro lado, se resaltan las características de sonoridad, lo que puede observarse como franjas horizontales a diferente frecuencia indicando la presencia de formantes.

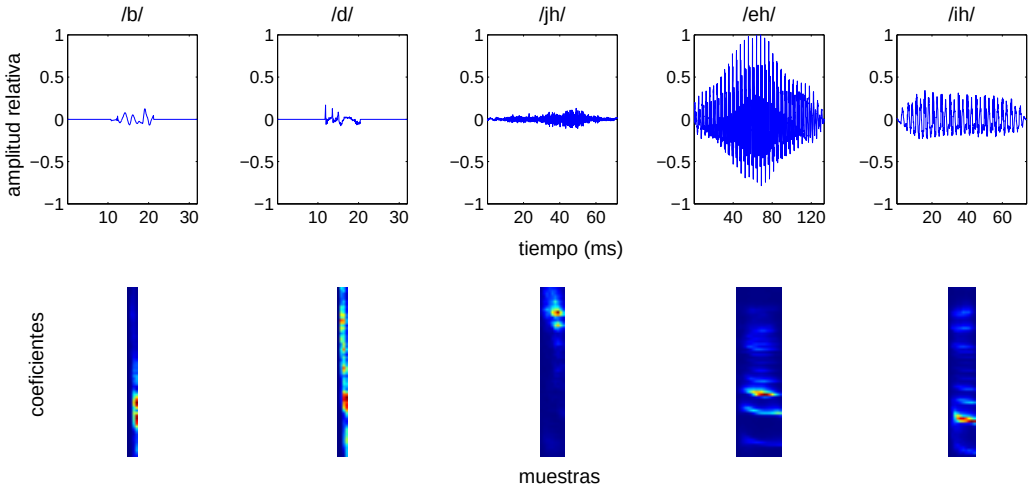



Figura 5.2: Ejemplos de los 5 fonemas usados en los experimentos. Se muestran el sonograma (arriba) y sus respectivos espectrogramas auditivos (abajo). Los espectrogramas tienen 64 coeficientes frecuenciales en altura, desde 0 a 8 KHz y un mínimo de 4 coeficientes en longitud, como puede ser observado en los fonemas más cortos.

## 5.3 Resultados con representación auditiva cortical aproximada

### 5.3.1 Aproximación sub-óptima con Matching Pursuit

En la obtención de la representación rala expuesta en la Sección 4.2.4, el costo computacional de la estimación de $\vec{\hat{a}}$ (ecuación 4.7) es realmente considerable. El algoritmo de *Matching Pursuit* (MP) es un método para aproximar la solución al problema de encontrar una representación rala, una vez que el diccionario está estimado o es provisto [74].

La raleza es incluida directamente mediante la elección de un número apropiado de términos. Dada una aproximación inicial $\vec{x}^{(0)} = \vec{0}$ y un residuo inicial $\vec{R}^{(0)} = \vec{x}$, una secuencia de aproximaciones es iterativamente construida. En el paso $k$, el parámetro $\gamma = \hat{\gamma}$ es seleccionado tal que el átomo $\vec{\phi}_{\hat{\gamma}}^{(k)}$ sea el de máxima correlación con el residuo $\vec{R}^{(k)}$, y un múltiplo de este átomo

es agregado a la aproximación al paso $k-1$, obteniendo

$$\vec{x}^{(k)} = \vec{x}^{(k-1)} + a_{\hat{\gamma}}^{(k)} \vec{\phi}_{\hat{\gamma}}^{(k)}, \tag{5.1}$$

donde $a_{\hat{\gamma}}^{(k)} = \langle \vec{R}^{(k-1)}, \vec{\phi}_{\hat{\gamma}}^{(k)} \rangle$, y $\vec{R}^{(k)} = \vec{x} - \vec{x}^{(k)}$. Luego de $m$ pasos se obtiene una aproximación a (4.4), con residuo $\vec{R} = \vec{R}^{(m)}$. Se dice, entonces, que el MP constituye una solución voraz al problema de la representación rala[1]; por lo tanto ésta comparte las mismas ventajas y desventajas de este tipo de métodos de optimización (rápidos pero generalmente sub-óptimos). No obstante, existen investigaciones que establecen que bajo ciertas condiciones apropiadas estos algoritmos obtienen una solución globalmente óptima [90, 91].

### 5.3.2 Parametrizaciones de referencia

La idoneidad del método AACR propuesto para reconocimiento robusto fue evaluada mediante la comparación de desempeño en clasificación frente a diferentes parametrizaciones usadas en esta área:

- los coeficientes cepstrales en escala Mel (MFCC, *Mel Fourier Cepstral Coefficients*) [92],
- el espectrograma auditivo (AS, *Auditory Spectrogram*) como se obtiene en la primer etapa del método propuesto,
- los coeficientes de predicción lineal perceptual (PLP, *Perceptual Linear Prediction coefficients*),
- la transformación espectral aplicada a éstos últimos coeficientes (RASTA-PLP, *Relative Spectral Transform - Perceptual Linear Prediction coefficients*) [93],
- y un filtrado óptimo probabilístico (POF, *Probabilistic Optimum Filtering*) aplicado a los coeficientes MFCC [94].

A continuación se revisan las ideas básicas de los últimos dos métodos mencionados, por ser los menos difundidos en la literatura.

La técnica RASTA-PLP se basa en el hecho de que la percepción humana tiende a reaccionar más a la amplitud relativa de un estímulo que a su valor

[1] Vorazmente minimiza $\left\| \vec{x} - \vec{\Phi}\vec{a} \right\|_2$.

absoluto. En particular, si al sistema auditivo se presenta un estímulo con variación lenta, la misma tiende a pasar desapercibida por el oyente. Esta idea se implementa en el método RASTA, una transformación que elimina los componentes de cambios lentos del habla por medio de un banco de filtros pasabanda aplicados a cada subbanda de energía. El proceso de RASTA es aplicado a los coeficientes PLP obtenidos por un modelo autorregresivo del espectro del habla, más coherente con la audición humana que los coeficientes de predicción lineal [93].

El análisis POF consiste en un mapeo entre un par de espacios acústicos: las características del habla limpia y ruidosa. Para la tarea de reconocimiento, se supone que un reconocedor entrenado con voz limpia se prueba con una versión ruidosa de los datos, como si hubieran sido adquiridos en un entorno acústico diferente. De estos datos, el mapeo trata de estimar los vectores limpios por medio de una transformación lineal probabilística por tramos. El término POF se refiere al conjunto de filtros que definen la transformación, cuyos resultados se combinan mediante un modelo gaussiano. La estimación de las características limpias $\hat{x}_n$ a partir del vector ruidoso $Y_n$ está dada por

$$\hat{x}_n = \left\{ \sum_{i=0}^{l-1} W_i^T p(g_i|z_n) \right\} Y_n, \tag{5.2}$$

donde $W_i^T$ es la matriz de coeficientes de filtrado y $p(g_i|z_n)$ es la probabilidad de Bayes de que el vector limpio $x_i$ pertenezca al conjunto $g_i$ dado el vector ruidoso $z_n$ [94].

### 5.3.3 Experimentos y resultados

En los siguientes experimentos, la señal acústica es procesada por el modelo de oído, el cual obtiene el espectrograma a nivel auditivo primario. A partir de estas representaciones tiempo-frecuencia, el diccionario de átomos bi-dimensionales es estimado. Finalmente, la representación aproximada con las activaciones de cada átomo para el tramo de señal analizada (el AACR, como fuera introducido en el Capítulo 4) es obtenida por medio del algoritmo de MP. Los patrones resultantes tienen la dimensión correspondiente al caso de representación completa del diccionario (256 coeficientes), pero sólo un subconjunto de estas activaciones son diferente de cero.

Una red neuronal tipo perceptrón multicapa (PMC) fue usada como clasificador. La arquitectura de los PMCs consistió en una capa de entrada de número fijo de unidades, una capa oculta de número ajustable de nodos y una

capa de salida de 5 unidades, una por fonema. La capa de entrada recibe un vector de 256 activaciones por vez, correspondiendo a un STRF de 64x4 coeficientes. El entrenamiento de las redes fue realizado con el algoritmo estándar de retropropagación del error con término de momento, ya expuesto en la sección de conocimientos preliminares [21].

En experimentación exhaustiva sobre habla limpia –detallada en el Apéndice C–, se obtuvieron diferentes diccionarios de átomos mediante el algoritmo de Basis Pursuit, y un gran número de redes neuronales tipo PMC fueron entrenados con patrones espectro-temporales de activación completa con el diccionario. Estas pruebas se realizaron sobre diccionarios completos y sobrecompletos. Los experimentos de clasificación fueron llevados a cabo por medio de un PMC que logró el mejor desempeño con una red de 256 átomos (caso completo). A pesar de que no existe evidencia en la literatura de que ésta sea la mejor elección en habla ruidosa, es la configuración usada en el resto de esta experimentación por consistencia con la configuración empleada en habla limpia.

La primera serie de experimentos, usando habla limpia, fue dedicada a encontrar el número óptimo de coeficientes en el esquema de extracción de características del algoritmo MP. Aquí, la exploración fue llevada a cabo con 4, 8, 16, 32, 64 y 128 coeficientes seleccionados del vector completo en $\mathbb{R}^{256}$. También, la mejor arquitectura para la red fue encontrada mediante la variación del número de nodos ocultos en cantidades crecientes en potencias de 2, desde 4 hasta 512 nodos. Los datos usados fueron dos subconjuntos balanceados de entrenamiento y prueba extraídos del conjunto de entrenamiento DR1 de TIMIT. Los subconjuntos constaban con 100 y 25 patrones de cada fonema, respectivamente. Cada experimento consistió en 3 ejecuciones del entrenamiento con pesos iniciales al azar, reportándose el valor medio de reconocimiento obtenido sobre el subconjunto de prueba.

Los resultados de este ajuste inicial se presentan en la Figura 5.3, donde un comportamiento similar para todas las curvas se observó en general. Estas mostraron que el clasificador obtiene un rendimiento más bajo cuando el tamaño de la capa oculta se redujo, debido a la limitada capacidad del PMC para aprender los aspectos fundamentales de los patrones para los propósitos de la clasificación. Además, el rendimiento alcanza un máximo y luego se aplana cuando el tamaño de la capa oculta se incrementa, debido al mayor número de pesos para entrenar con la misma talla del conjunto de entrenamiento. En cuanto a las diferencias encontradas al variar el número de coeficientes seleccionados, los mejores rendimientos se obtuvieron con pocos coeficientes seleccionados, mostrando las curvas un decaimiento general en el desempeño al aumentar este parámetro. Esta situación puede surgir debido

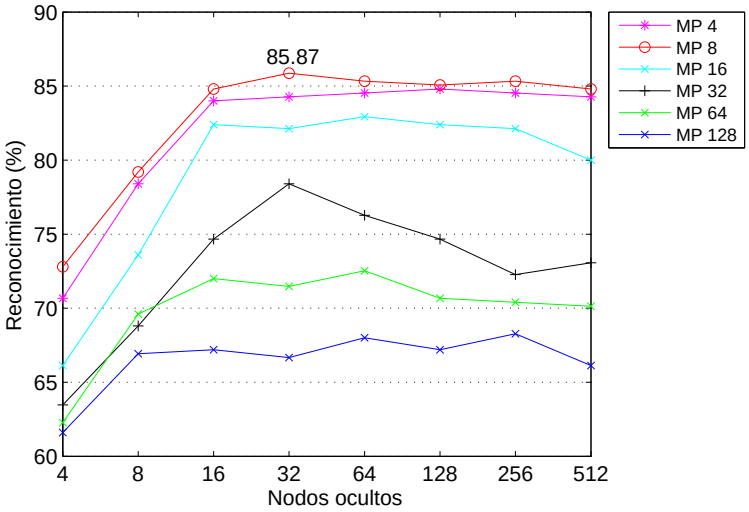



Figura 5.3: Ajuste inicial del número de coeficientes seleccionados en el algoritmo y número de nodos ocultos en la red neuronal. El mejor desempeño se obtiene para 8 coeficientes seleccionados por el algoritmo MP y 32 nodos en la capa oculta del PMC.

al hecho de que los patrones contienen, cada vez, más información que no es relevante para la clasificación.

El mejor desempeño en estas condiciones de habla limpia se obtiene para 8 coeficientes seleccionados por el MP y 32 nodos en la capa oculta del PMC. Por lo tanto, esta configuración del esquema se establece para la siguiente serie de experimentos.

Con el objetivo de evaluar el desempeño de la representación cortical en la presencia de ruido y comparar su robustez contra otras parametrizaciones, los experimentos siguientes consistieron en el entrenamiento del PMC con voz limpia y pruebas en diferentes condiciones de ruido. La versión ruidosa del corpus de habla se obtuvo mezclando aditivamente a diferentes SNRs los datos de TIMIT con ruido blanco tomado de la base de datos NOISEX-92 database [95].

La extracción de características para MFCC, PLP y RASTA-PLP se fijó a 12 coeficientes con los coeficientes delta y energía agregados, dando lugar a patrones de 26 coeficientes. Los patrones obtenidos a partir del AS tienen 256 coeficientes. Para todas estas redes, el número de unidades ocultas se fijó para el mismo número de unidades de entrada, ya que en experimentos preliminares (no mostrados aquí) se constató que es la configuración óptima.

Una proporción diferente en la relación señal-ruido (SNR) se fijó previo

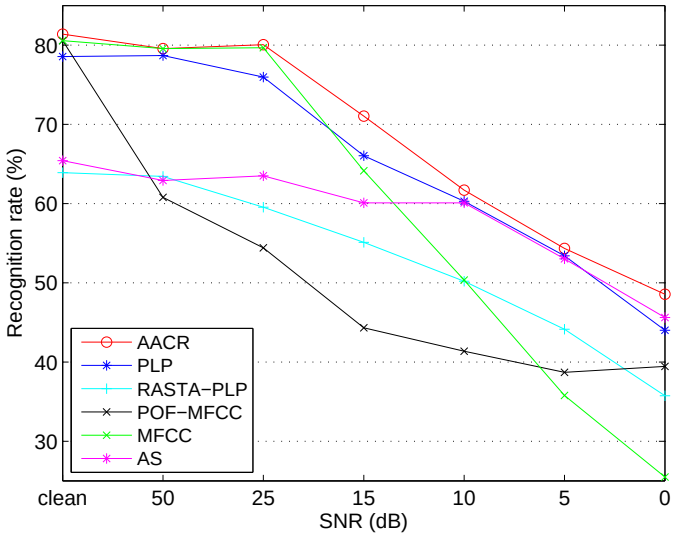



Figura 5.4: Porcentaje de reconocimiento sobre el conjunto de prueba en la clasificación de los 5 fonemas, en presencia de ruido a diferente SNR, desde habla limpia hasta igual energía de habla y ruido (SNR=0 dB).

a cada experimento, desde el habla limpia (SNR infinito) hasta SNR=0 dB (igual nivel de energía del ruido y del habla). Luego, para cada parametrización, se llevó a cabo una serie de 10 ejecuciones del entrenamiento con pesos iniciales al azar. Este método de inicialización parece ser lo suficientemente bueno para nuestros propósitos, dado que el pico de generalización del PMC se alcanza en aproximadamente 10 a 20 iteraciones del algoritmo de retropropagación, en un ciclo de 200 iteraciones máximas.

Los resultados obtenidos se muestran en la Figura 5.4. Las curvas muestran el comportamiento general de los sistemas de RAH en presencia de ruido: logran un buen desempeño en reconocimiento con voz limpia, con una caída en el rendimiento cuando el contenido de ruido en la señal aumenta. El rendimiento de la parametrización menos robusta, el MFCC, rápidamente decae en condiciones severas de ruido (SNR cerca de 0 dB). Todas las otras parametrizaciones obtienen tasas más altas en estas condiciones, como puede verse para 5 y 0 dB. Respecto del comportamiento obtenido por el POF, en [96] los autores mostraron un mejor rendimiento en el POF que el MFCC a niveles altos de ruido también, pero esos resultados se obtuvieron en habla continua y usando modelos ocultos de Markov como clasificadores. Estas condiciones experimentales son muy diferentes de nuestra configuración de reconocimiento de fonemas aislados por un clasificador estático.

Un análisis más profundo de los resultados se presenta en las matrices de confusión mostradas en la Tabla 5.2. Se muestra la tasa media porcentual de reconocimiento para cada fonema, usando la mejor configuración encontrada con el enfoque AACR propuesto: 8 coeficientes en el algoritmo de MP y el PMC con 32 unidades en la capa oculta. Las tasas corresponden a los valores medios sobre el conjunto de prueba para los 10 inicializaciones, con dos diferentes condiciones evaluadas: señales limpias y con ruido añadido a SNR=15 dB. Para cada fonema en la primera columna (salida deseada del PMC), las matrices muestran en las filas los porcentajes de cada fonema dados por las salidas de las redes. Los resultados mostraron que, para el caso de habla limpia, el PMC es capaz de llevar a cabo una clasificación adecuada de casi todas las clases, excepto el fonema */ih/*, cuya clasificación se encuentra dispersa en el resto de clases (última fila). En la condición de ruido, se puede observar que las redes clasifican muy bien a los fonemas */d/*, */jh/* y */eh/*, mientras que los fonemas */b/* y */ih/* son principalmente asignados a las otras clases.

El caso del fonema */b/* es interesante de analizar. En habla limpia se obtiene un buen desempeño (84 %), con una confusión menor con el fonema */d/*. Con la introducción de ruido, aún en cantidades moderadas, la confusión es incrementada: 67 % de las */b/* son reconocidas como */d/*. Este comportamiento puede ser explicado por el hecho de que en habla limpia, estas consonantes sonoras plosivas muestran un contenido de alta energía a bajas frecuencias del AS; pero el fonema */d/* también presenta alta energía a frecuencias mayores, al contrario de la */b/* (véase Fig. 5.2). Cuando se agrega ruido blanco, el AS de la */b/* se asemeja más a aquél de la */d/*, dando lugar al error de clasificación encontrado. Un estudio en línea con esta idea fue presentado en [97], donde los autores mostraron que estos fonemas son fácilmente confundibles dada su alta similaridad acústica (distancia Euclídea entre espectrogramas auditivos promedios).

### 5.3.4 Discusión de resultados

En las condiciones obtenidas para la mejor tasa de reconocimiento, se pudo observar que el óptimo se alcanza con sólo 8 coeficientes en el algoritmo de MP. Así, las pistas importantes de cada patrón se codificarían en aproximadamente 3 % del total de átomos en el diccionario. Por otro lado, esta representación es mejor procesada por un PMC con una baja dimensión en la capa oculta, en este caso 32 nodos. Esto pone de manifiesto la capacidad de generalización de las redes, dado que los patrones portarían sólo la

Tabla 5.2: Matrices de confusión mostrando los porcentajes de clasificación para el AACR propuesto, sobre habla limpia y con ruido a SNR=15 dB. En las filas: salida deseada, en columnas: clasificación obtenida. El porcentaje promedio de reconocimiento es del 83 % (habla limpia) y 71 % (habla con ruido).

| Fonema | Habla limpia | | | | | Habla con ruido a SNR=15 dB | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| real | */b/* | */d/* | */jh/* | */eh/* | */ih/* | */b/* | */d/* | */jh/* | */eh/* | */ih/* |
| */b/* | 84 | 16 | | | | 33 | 67 | | | |
| */d/* | 16 | 70 | 10 | | 4 | 5 | 94 | 1 | | |
| */jh/* | | 1 | 99 | | | | 12 | 88 | | |
| */eh/* | 5 | | | 93 | 2 | 5 | 1 | | 92 | 2 |
| */ih/* | 2 | 2 | 10 | 26 | 60 | 3 | 17 | 15 | 18 | 46 |

información más importante y por lo tanto menos pesos se requieren en el modelo.

Respecto de la robustez frente a ruido, el abordaje AACR aquí propuesto siempre consigue los mayores índices de clasificación comparado a las otras parametrizaciones para todas las SNR evaluadas, incluyendo voz limpia. Este resultado estaría dado por la robustez intrínseca del AACR, donde sólo las activaciones más importantes son retenidas por el algoritmo. Por lo tanto, los coeficientes seleccionados estarían actuando como detectores de pistas fonéticas que capturan las particularidades de cada fonema y hacen posible su caracterización.

La significancia estadística de estos resultados se evaluó teniendo en cuenta la probabilidad de que el error de clasificación $\epsilon$ de un clasificador dado, sea menor que el error $\epsilon_{ref}$ del sistema de referencia. Para hacer esta estimación, se supone la independencia estadística de los errores de cada patrón, y la distribución binomial de los errores fue modelada por medio de una distribución gaussiana (esto es posible debido a que se tiene un número suficientemente grande de patrones de prueba). Por lo tanto, comparando nuestro enfoque contra el segundo mejor resultado (espectrograma auditivo, AS) para el peor de los casos, SNR = 0 dB, se obtuvo una $Pr(\epsilon_{ref} > \epsilon) > 96{,}54\,\%$. El desvío estándar para el AACR varía entre 0,88 (voz limpia) hasta 2,31 (SNR=0 dB), mientras que para los coeficientes PLP el mismo parámetro tiene una variación mayor: de 0,87 hasta 10,71, respectivamente.

## 5.4 Representación cortical auditiva no negativa

Siguiendo los lineamientos de este método expuesto en el Capítulo 4, una serie de experimentos fue desarrollada para llevar a cabo la limpieza de ruido sobre señales de habla correspondientes a frases completas, sin segmentar en fonemas. La voz limpia fue extraída del corpus TIMIT. Los datos usados en esta tesis correspondieron al conjunto de 10 oraciones emitidas por el hablante FCJF0, femenino de la región dialéctica DR1. En las pruebas de limpieza de ruido, los mismos tipos de ruido: blanco, murmullo y calle fueron adicionados con relación señal-ruido variable. El desempeño en la limpieza de ruido fue evaluada mediante la medida PESQ.

### Limpieza de ruido en señales de habla

La Fig. 5.5 muestra un ejemplo de la limpieza de ruido de las señales reales correspondientes a los datos de habla. La señal limpia corresponde a la frase */She had your dark suit in greasy wash water all year/* (que se muestra en el espectrograma de la parte superior). La señal es contaminada con ruido blanco a SNR=0 dB.

Los efectos del ruido se pueden observar en el espectrograma central, donde casi todas las características importantes del habla quedan enmascaradas por el ruido. El esquema de limpieza de ruido, sin embargo, es capaz de recuperar las formantes más importantes y reducir la energía del ruido tal como se muestra en el espectrograma de la parte inferior. Cabe mencionar que este ejemplo corresponde al caso más desfavorable, ruido blanco de igual energía que la señal de interés.

Para la medición de la calidad objetiva, un procedimiento de validación cruzada en 10 bloques fue aplicado con el entrenamiento de un diccionario con 9 señales y la evaluación con la restante. En cada caso, los ruidos blanco y de calle fueron añadidos con SNR de 12, 6 y 0 dB.

Además, se utilizó un filtro de Wiener tiempo-frecuencia con estimaciones de la señal y el ruido sobre la base de la descomposición atómica. Específicamente, se utilizaron las activaciones del diccionario de señal para producir una estimación de la señal limpia $s(t)$, y el diccionario de ruido para producir una estimación del ruido $n(t)$, supuesto aditivo. Luego, después de la

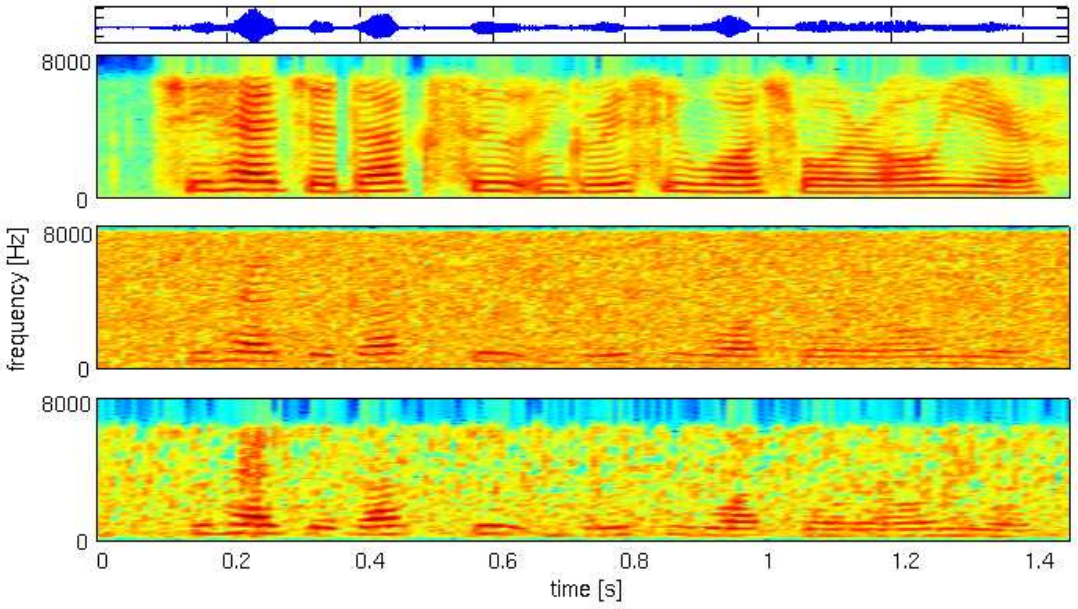



Figura 5.5: Ejemplo del resultado de la limpieza cortical auditiva de una señal de voz contaminada con ruido blanco a SNR=0 dB. Se muestran los espectrogramas (STFT) de la señal limpia (arriba), la señal de ruido (centro) y la señal limpia reconstruida (abajo). La señal acústica en la parte superior de la figura se da como referencia temporal.

aplicación del NCD, se construyó el filtro de Wiener como

$$\frac{|S(\omega,\tau)|^2}{|S(\omega,\tau)|^2+|N(\omega,\tau)|^2} \tag{5.3}$$

donde $S(\omega,\tau)$ y $N(\omega,\tau)$ son las representaciones tiempo-frecuencia de $s(t)$ and $n(t)$, respectivamente [98, 99].

La Tabla 5.3 muestra la puntuación media PESQ obtenida para el esquema de validación cruzada, donde tres diferentes escenarios posibles en la aplicación del NCD son presentados. El "Wiener/ruido NCD" aplica el filtro de Wiener a la señal ruidosa, donde la estimación del ruido está dada por el NCD con sólo STRFs seleccionados del diccionario de ruido. El "habla NCD" corresponde a la reconstrucción NCD a partir de STRFs seleccionadas del diccionario de habla. El "NCD+Wiener" aplica el filtro de Wiener para ambas estimaciones NCD previas, la del ruido y la del habla.

Tabla 5.3: Puntuaciones medias PESQ obtenidas para frases del corpus TIMIT. Los experimentos donde la aplicación del NCD logró una mejora de la calidad con respecto a la señal ruidosa se destacan en negrita.

| Ruido | SNR (dB) | Señal | | | |
|---|---|---|---|---|---|
| | | Ruidosa | Wiener/ruido NCD | habla NCD | NCD+Wiener |
| Blanco | 12 | 2,25 | **2,33** | 2,23 | 2,25 |
| | 6 | 1,92 | **1,99** | **1,93** | **2,04** |
| | 0 | 1,63 | 1,69 | **1,75** | **1,94** |
| Calle | 12 | 2,57 | **2,66** | 2,30 | 2,22 |
| | 6 | 2,21 | **2,31** | 2,14 | 2,18 |
| | 0 | 1,79 | **1,90** | **1,93** | **2,04** |

### 5.4.1 Discusión de resultados

El enfoque planteado emplea una codificación rala no-negativa que se aplica en un algoritmo de limpieza de ruido simple: la reconstrucción de la señal acústica eliminando la activación de átomos de ruido. Este algoritmo explota información *a priori* obtenida de señales limpias y señales contaminadas con ruido.

El método se aplicó a la limpieza de ruido de fonemas en presencia de diferentes tipos de ruidos. Los resultados mostraron que el esquema propuesto puede mejorar una medida objetiva de calidad, sobre todo en las señales severamente degradadas, dadas las diferencias mayores en la puntuación PESQ entre señales ruidosas y limpiadas. Por ejemplo, en el caso de ruido blanco a SNR=0 dB, el método mejora la PESQ de 1,63 hasta 1,94. Por otro lado, algunas puntuaciones más bajas de la PESQ podrían estar señalando que el método se estaría excediendo en la limpieza que efectúa sobre la señal, eliminando no sólo el ruido sino también algunas de las características necesarias para mantener la calidad. Esto puede ser especialmente cierto en el caso de ruido de baja frecuencia, como el de calle o el ruido murmullo, debido a que las características importantes del habla (formantes) se encuentran en el mismo rango de frecuencias del ruido.

## 5.5 Comentarios de cierre de capítulo

En este capítulo se presentó la experimentación de las técnicas propuestas como alternativa a la representación clásica del habla, sobre una tarea inicial de clasificación de fonemas.

Se mostró que el esquema AACR propuesto, junto a variantes en la obtención de las activaciones corticales, logra mejores desempeños tanto en condiciones de habla limpia como de habla inmersa en ruido de diverso tipo respecto a otras parametrizaciones usuales en el área. Esto hace del AACR una alternativa a ser considerada para sistemas de RAH.

El capítulo siguiente presenta las conclusiones finales y los trabajos futuros que pueden continuar las líneas de investigación planteadas en esta tesis.

CAPÍTULO 6

# Conclusiones y desarrollos futuros

## 6.1 Conclusiones generales

En esta tesis se ha presentado un avance en la representación de datos y modelización en el contexto del reconocimiento de patrones. Se propusieron alternativas que mejoran diferentes aspectos respecto a los sistemas actuales: el preproceso, la extracción de características y la modelización, sobre dos aplicaciones de diferente naturaleza.

Las principales contribuciones de esta tesis se resumen en los siguientes puntos:

- Se propusieron nuevas medidas sobre los cromosomas mediante un muestreo local en ambas direcciones, el cual permite capturar las variabilidades de las bandas alternantes claras y oscuras de los mismos.
  En la comparación del desempeño obtenido en clasificación, los sistemas entrenados con estos patrones superaron a aquéllos entrenados con los perfiles clásicos de densidad, gradiente y forma reportados en la literatura, dando cuenta de la bondad de las medidas propuestas.

- Se diseñaron y probaron clasificadores especializados en realizar un análisis por tramos a lo largo del cromosoma, utilizando solamente información obtenida a partir de la variabilidad de intensidades de grises

en las bandas. En particular, se adaptaron dos tipos de redes neuronales y los modelos ocultos de Markov continuos.
Se obtuvo un mejor rendimiento en sistemas con un procesamiento local dedicado, como las redes parcialmente recurrentes de Elman y los modelos de Markov, principalmente por la representación más adecuada de los objetos bajo análisis.

- Se formuló un algoritmo de post-proceso que toma los resultados de la clasificación previa por cromosomas separados y efectúa una reasignación de clases teniendo en cuenta todos los cromosomas de una célula. Mediante la experimentación se demuestra que el algoritmo logra reducir significativamente el error obtenido previamente en la clasificación por cromosomas aislados. Este algoritmo constituye un bloque de clasificación que otorga contexto celular a la tarea, y puede ser empleado en cualquier sistema de clasificación de cromosomas, ya que no es específico del modelo que se utilice para el etiquetado inicial.

- Se propusieron diferentes alternativas a la parametrización de la señal de habla, basadas en un método de extracción de características biológicamente inspirado, las representaciones ralas. Las técnicas planteadas aproximan las características de las activaciones a nivel cortical seleccionando átomos de un diccionario obtenido a partir de los espectrogramas auditivos.
Se mostró el mejor desempeño alcanzado por las activaciones corticales sobre una tarea de clasificación de fonemas respecto de las técnicas clásicas robustas, tanto con señales limpias como en el caso de habla inmersa en ruido.

- Se propuso un nuevo método para limpieza de ruido en el dominio de las activaciones corticales, en un contexto de análisis no-negativo de la combinación de átomos adecuada de diccionarios de señal y de ruido.
Se obtuvo una mejor separación de estos componentes, tanto en la experimentación con señales artificiales como en las pruebas con señales de habla correspondientes a frases completas.

## 6.2 Desarrollos futuros

A fin de mejorar el desempeño general sobre los sistemas propuestos, se comentan a continuación algunos trabajos futuros que surgen como contin-

uación de los avances presentados en esta tesis.

En el reconocimiento de cromosomas, se exploró de manera inicial una técnica paramétrica para obtener el eje, lo cual podría reducir el error en la obtención de perfiles en cromosomas cortos y severamente curvados. La combinación de los perfiles de grises con otros descriptores locales como coeficientes de Fourier, onditas o el patrón canónico de todos los cromosomas propuesto en [100] también podría agregar información útil para la clasificación local no-contextual.

En el reconocimiento robusto del habla, una línea de investigación que surge naturalmente es continuar y extender el trabajo inicial sobre clasificación de fonemas al habla continua, donde los diccionarios deberán especializarse sobre todos los fonemas presentes en el habla. Aquí podrían plantearse, alternativamente a la aproximación descripta en esta tesis, el cálculo de un diccionario por fonema. En este contexto debería explorarse, además, el clasificador óptimo para esta representación, por ejemplo podría diseñarse uno basado en ensamble de clasificadores más pequeños especializados en fonemas particulares.

## 6.3 Publicaciones resultantes

El trabajo de investigación llevado a cabo en esta tesis dio lugar a las siguientes publicaciones.

### Anales de congreso

1. C. Martínez, A. Juan and F. Casacuberta, "Clasificación automática de cromosomas mediante modelos ocultos continuos de Markov". XIV Congreso de la Sociedad Argentina de Bioingeniería (SABI), Córdoba (Argentina), octubre de 2003.

2. C. Martínez, J. Goddard, D. Milone and H. Rufiner "An approach to robust phoneme classification by modeling the auditory cortical representation of speech". XIII Reunión de Trabajo en Procesamiento de la Información y Control (RPIC), Rosario, setiembre de 2009.

### Capítulos de libro

3. C. Martínez, A. Juan and F. Casacuberta, "Using Recurrent Neural Networks for Automatic Chromosome Classification". Lecture Notes in Computer Science 2415, J.R. Dorronsoro (Ed.): ICANN 2002, pp. 565-570, Springer-Verlag Berlin Heidelberg, 2002.

4. C. Martínez, H. García, A. Juan and F. Casacuberta, "Chromosome Classification Using Continuous Hidden Markov Models". Lecture Notes in Computer Science 2652, F.J. Perales et al. (Eds.): IbPRIA 2003, pp. 494-501, Springer-Verlag Berlin Heidelberg, 2003.

5. H. Rufiner, César Martínez, Diego Milone and John Goddard, "Auditory cortical representations of speech signals for phoneme classification". Lecture Notes in Artificial Inteligence 4827, A. Gelbukh and A.F. Kuri Morales (Eds.): MICAI 2007, pp. 1004-1014, Springer-Verlag, Berlin Heidelberg, 2007.

### Revistas de corriente principal

6. H. Rufiner, C. Martínez, D. Milone, J. Goddard, "Extracción de características bioinspirada basada en un Modelo Cortical Auditivo". Anales de la Academia Nacional de Ciencias Exactas, Físicas y Naturales, ISSN: 0365-1185, Tomo 58 (2006): 71-78.

7. C. Martínez, A. Juan and F. Casacuberta, "Iterative Contextual Recurrent Classification of Chromosomes". Neural Processing Letters, vol. 26, number 3, pp. 159-175, Springer Netherlands, 2007.

8. C. Martínez, J. Goddard, D. Milone and H. Rufiner, "Approximated auditory cortical representation for robust speech classification". *Computer Speech & Language*, Elsevier Science Press, 2010 (en revisión).

9. C. Martínez, J. Goddard, L. Di Persia, D. Milone and H. Rufiner, "Denoising of sound signals in the non-negative auditory cortical domain". En preparación.

10. C. Martínez, J. Goddard, D. Milone and H. Rufiner "An approach to robust phoneme classification by modeling the auditory cortical representation of speech". Seleccionado para publicación en Latin American Applied Research, ISSN:0327-0793, 2011.

# Bibliografía


[1] S. Bow. *Pattern Recognition and Image Processing*. Marcel Dekker, Inc., New York, NY, USA, 2002.

[2] R. Duda, P. Hart, and D. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, November 2000.

[3] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: a review. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[4] K.-F. Lee and R. Reddy. *Automatic Speech Recognition: The Development of the Sphinx Recognition System*. Kluwer Academic Publishers, Norwell, MA, USA, 1988.

[5] C.-H. Lee, L. Rabiner, R. Pieraccini, and J. G. Wilpon. Acoustic modeling for large vocabulary speech recognition. *Computer Speech and Language*, 4(2):127–165, 1990.

[6] M. Shirvaikar. Trends in automated visual inspection. *Journal of Real-Time Image Processing*, 1(1):41–43, 2006.

[7] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[8] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003.

[9] J. Dorronsoro, F. Ginel, C. Sánchez, and C. Cruz. Neural fraud detection in credit card operations. *IEEE Trans. on Neural Networks*, 8(4):827–834, 1997.

[10] R. Wheeler and S. Aitken. Multiple algorithms for fraud detection. *Knowledge-Based Systems*, 13(2-3):93–99, 2000.

[11] G. Paliouras, V. Karkaletsis, and C. Spyropoulos, editors. *Machine Learning and Its Applications, Advanced Lectures*, volume 2049 of *Lecture Notes in Computer Science*. Springer, 2001.

[12] J. Mrázek and S. Xie. Pattern locator: a new tool for finding local sequence patterns in genomic dna sequences. *Bioinformatics*, 22(24):3099–3100, 2006.

[13] S. Theodoridis and K. Koutroumbas. *Pattern Recognition, Second Edition*. Academic Press, San Diego, CA, USA, 2003.

[14] I. T. Jolliffe. *Principal Component Analysis, Second Edition*. Springer, New York, 2002.

[15] Taub A. H. (ed.). *John von Neumann: Collected Works*, volume 5:34-79. Pergamon Press, Oxford (UK), 1961.

[16] D. Rumelhart and J. McClelland. *Parallel Distributed Processing*. The MIT Press, 1986.

[17] J. Freeman and D. Skapura. *Neural Networks. Algorithms, Applications and Programming Techniques*. Addison-Wesley Publishing Company, Inc., 1991.

[18] M. Fishler and O. Firschein. *Intelligence, The Eye, The Brain and The Computer*. Addison Wesley, Boston (MA), USA, 1998.

[19] B. Widrow and M. Hoff. Adaptive Switching Circuits. *In IRE WESCON Convention Record, part 4*, pages 96–104, 1960.

[20] M. Hassoun. *Fundamentals of Artificial Neural Networks*. The MIT Press, 1995.

[21] S. Haykin. *Neural Networks: A Comprenhensive Foundation*. Pearson Education, 1999.

[22] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang. Phoneme recognition using time-delay neural networks. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 37(3):328–339, 1999.

[23] G. Dorffner. Neural networks for time series processing. *Neural Network World*, 6(4):447–468, 1996.

[24] L. Gupta, M. McAvoy, and J. Phegley. Classification of temporal sequences via prediction using the simple recurrent neural network. *Pattern Recognition*, 33:1759–1770, 2000.

[25] J. L. Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

[26] X. Huang, A. Acero, and H-W. Hon. *Spoken Language Processing: a guide to theory, algorithm and system development.* Prentice Hall, 2001.

[27] L. R. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition.* Prentice Hall, 1993.

[28] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, V. Valtchev, and P. Woodland. *The HTK Book, v 3.4.* Cambridge University, December 2006.

[29] H. García. Preproceso y extracción de características (sintácticas) para el diseño de clasificadores de cromosomas humanos. Master's thesis, Faculty of Computer Science, Polytechnic University of Valencia, 1999.

[30] D. Hartl and E. Jones. *Genetics: Principles and Analysis.* Gareth Stevens Publishing, 1998.

[31] S. Delshadpour. Reduced size multi layer perceptron neural network for human chromosome classification. In *Proc. of the 25 Annual Int. Conf. of the IEEE EMBS*, volume 3, pages 2249–2252, Cancun, Mexico, 2003.

[32] J. Cho, S. Ryu, and S. Woo. A study for the hierarchical artificial neural network model for giemsa-stained human chromosome classification. In *Proc. of the 26 Annual Int. Conf. of the IEEE EMBS*, pages 4588–4591, San Francisco, CA, USA, September 2004.

[33] P. Biyani, X. Wu, and A. Sinha. Joint classification and pairing of human chromosomes. *IEEE Trans on Computational Biology and Bioinformatics*, 2:102–109, 2005.

[34] X. Wang, S. Li, M. Wood, W. Chen, and B. Zheng. Automated identification of analyzable metaphase chromsomes depicted on microscopic digital images. *Journal of Biomedical Informatics*, 41:264–271, 2008.

[35] M. Sampat, A. Bovik, J. Aggarwal, and K. Castleman. Supervised parametric and non-parametric classification of chromosome images. *Pattern Recognition*, 38:1209–1223, 2005.

[36] M. Moradi and S. Kamaledin Setarehdan. New features for automatic classification of human chromosomes: A feasibility study. *Pattern Recognition Letters*, 27(1):19–28, 2006.

[37] J. H. Kao, J. H. Chuang, and T. P. Wang. Automatic chromosome classification using medial axis approximation and band profile similarity. In *Computer Vision ACCV 2006*, volume 3852 of *Lecture Notes in Computer Science*, pages 274–283. Springer-Verlag, 2006.

[38] Nojun Kwak. Feature extraction for classification problems and its application to face recognition. *Pattern Recognition*, 41(5):1701–1717, 2008.

[39] J. Piper. Variability and bias in experimentally measured classifier error rates. *Patter Recognition Letters*, 13:685–692, 1992.

[40] G. Ritter and K. Gaggermeier. Automatic classification of chromosomes by means of quadratically asymmetric statistical distributions. *Pattern Recognition*, 32:997–1008, 1999.

[41] R. Gonzalez and R. Woods. *Digital image processing*. Prentice-Hall, 2002.

[42] C. Hilditch. Linear skeletons from square cupboards. *Machine Intelligence*, 19:403–420, 1969.

[43] J. Piper and E. Granum. On fully automatic feature measurement for banded chromosome classification. *Cytometry*, 10:242–255, 1989.

[44] S. Yun Ryu, J. Man Cho, and S. Hyo Woo. A study for the feature selection to identify giemsa-stained human cromosomes based on artificial neural network. In *Proc. of the 23rd Annual Int. Conf. of the IEEE EMBS*, pages 691–692, Istanbul, Turkey, October 2001.

[45] S. Hazout, J. Mignot, M. Guiguet, and A. J. Valleron. Rectification of distorted chromosome image: automatic determination of density profiles. *Comput. Biol. Med.*, 14:63–76, 1984.

[46] G. Ritter and G. Schreib. Using dominant points and variants for profile extraction from chromosomes. *Pattern Recognition*, 34:923–938, 2001.

[47] J. Deller, J. Proakis, and J. Hansen. *Discrete Time Processing of Speech Signals*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1993.

[48] R. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Trans. on Image Processing*, 14(3):294–307, 2005.

[49] University of Stuttgart - University of Tübingen. *SNNS Stuttgart Neural Network Simulator*, version 4.2 edition, 1998. User Manual.

[50] R. Pieraccini. Pattern compression in isolated word recognition. *Signal Processing*, 7(1):1–15, 1984.

[51] C. Martínez, A. Juan, and F. Casacuberta. Using recurrent neural networks for automatic chromosome classification. In *Artificial Neural Networks - ICANN 2002: Proceedings*, volume 2415 of *Lecture Notes in Computer Science*, pages 565–570. Springer-Verlag, 2002.

[52] C. Martínez, A. Juan, and F. Casacuberta. Iterative Contextual Recurrent Classification of Chromosomes. *Neural Processing Letters*, 26(3):159–175, 2007.

[53] S. J. Young et al. HTK: Hidden Markov Model Toolkit. Technical report, Entropic Research Laboratories Inc., 1997.

[54] Q. Wu, Z. Liu, T. Chen, Z. Xiong, and K. Castleman. Subspace-based prototyping and classification of chromosome images. *IEEE Trans. on Image Processing*, 14(9):1277–1287, 2005.

[55] H. Toutenburg. *Statistical Analysis of Designed Experiments*. Springer, 2002.

[56] J. Cho. Chromosome classification using backpropagation neural networks. *IEEE Engineering in Medicine and Biology Magazine*, 19(1):28–33, 2000.

[57] A. Carothers and J. Piper. Computer-aided classification of human chromosomes: a review. *Statistics and Computing*, 4:161–171, 1994.

[58] P. Divenyi (editor). *Speech Separation by Humans and Machines*. Springer, 2004.

[59] D. O'Shaughnessy. Invited paper: Automatic speech recognition: History, methods and challenges. *Pattern Recognition*, 41:2965–2979, 2008.

[60] L. Rocha H. Rufiner, J. Goddard and M. Torres. Statistical method for sparse coding of speech including a linear predictive model. *Physica A*, 367:231–251, 2006.

[61] S.G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 2nd edition, September 1999.

[62] B.A. Olshausen and D.J. Field. Emergence of simple cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[63] B.A. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

[64] Marc' A. Ranzato, Christopher Poultney, Sumit Chopra, and Yann Lecun. Efficient learning of sparse representations with an energy-based model. In *NIPS*, 2006.

[65] Hugo L. Rufiner. *Análisis y modelado digital de la voz. Técnicas recientes y aplicaciones*. Ediciones UNL, Santa Fe, 2009.

[66] M.S. Lewicki and T.J. Sejnowski. Learing overcomplete representations. In *Advances in Neural Information Processing 10 (Proc. NIPS'97)*, pages 556–562. MIT Press, 1998.

[67] M.S. Lewicki and B.A. Olshausen. A probabilistic framework for the adaptation and comparison of image codes. *Journal of the Optical Society of America*, 16(7):1587–1601, 1999.

[68] S. A. Shamma. Neural and functional models of the auditory cortex. In M. Arbib, editor, *Handbook of Brain Theory and Neural Networks*, Bradford Books. The MIT Press, 1995.

[69] Konrad P. Kording, Peter Konig, and David J Klein. Learning of sparse auditory receptive fields. In *Proc. of the International Joint Conference on Neural Networks (IJCNN '02)*, volume 2, pages 1103–1108, Honolulu, HI, United States, May 2002.

[70] D.J. Klein, D.A. Depireux, J.Z. Simon, and S.A. Shamma. Robust spectrotemporal reverse correlation for the auditory system: Optimizing stimulus design. *Journal of Computational Neuroscience*, 9:85–111, 1996.

[71] S.A. Shamma. Auditory cortical representation of complex acoustic spectra as inferred from the ripple analysis method. *Comput. in Neural Syst.*, 7:439–476, 1996.

[72] P.O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.

[73] S. Chen, D. Donoho, M. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43(1):129–159, 2001.

[74] S.G. Mallat and Z. Zhang. Matching pursuit with time-frequency dictionaries. *IEEE Trans. in Signal Proc.*, 41:3397–3415, December 1993.

[75] M. Aharon and M. Elad and A.M. Bruckstein. K-SVD and its non-negative variant for dictionary design. In *Proceedings of the SPIE conference wavelets*, volume 5914, 2005.

[76] T. Chiu and P. Ru and S. Shamma. Multiresolution spectrotemporal analysis of complex sounds. *Journal of the Acoustical Society of America*, 118(2):897–906, 2005.

[77] N. Mesgarani and S. Shamma. Denoising in the domain of spectrotemporal modulations. *EURASIP Journal on Audio, Speech and Music Processing*, 2007:8 pages, 2007.

[78] H. Hirsch and D. Pearce. The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *Proceedings of the ISCA ITRW ASR2000*, 2000.

[79] Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. *ITU-T Recommendation P.862*, 2001.

[80] A.W. Rix, J.G. Beerends, M.P. Hollier, and A.P. Hekstra. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 749–752, 2001.

[81] L. Di Persia and D. Milone and H. Rufiner and M. Yanagida. Perceptual evaluation of blind source separation for robust speech recognition. *Signal Processing*, 88(10):2578–2583, 2008.

[82] Antonio Quilis. *Tratado de Fonología y Fonética Españolas*. Biblioteca Románica Hispánica. Editorial Gredos, Madrid, 1993.

[83] J. Yousafzai, Z. Cvetković, P. Sollich and M. Ager. Robustness of phoneme classification using support vector machines: a comparison between plp and acoustic waveform representations. *Proceedings of ICASSP*, 2008.

[84] M. Ager, Z. Cvetković, P. Sollich and B. Yu. Towards robust phoneme classification: augmentation of plp models with acoustic waveforms. *Proceedings of EUSIPCO*, 2008.

[85] J. Yousafzai, Z. Cvetković and P. Sollich. Tuning support vector machines for robust phoneme classification with acoustic waveforms. *Proceedings of INTERSPEECH*, 2009.

[86] S. Ganapathy, S. Thomas and H. Hermansky. Temporal envelope compensation for robust phoneme recognition using modulation spectrum. *J. Acoust. Soc. Am.*, 128(6):3768–3780, 2010.

[87] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren. DARPA TIMIT Acoustic-phonetic continuous speech corpus documentation. Technical report, National Institute of Standards and Technology, 1993.

[88] K. N. Stevens. *Acoustic Phonetics*. MIT Press, 2000.

[89] X. Yang, K. Wang, S. Shamma. Auditory representations of acoustic signals. *IEEE Transactions on Information Theory*, 38:824–839, 1992. Special Issue on Wavelet Transforms and Multiresolution Signal Analysis.

[90] D. Donoho, M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.

[91] D. Donoho, M. Elad, V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6 – 18, 2006.

[92] J. Deller, J. Proakis, J. Hansen. *Discrete Time Processing of Speech Signals*. Macmillan Publishing, New York, 1993.

[93] H. Hermansky. Perceptual linear predictive (PLP) analysis of speech. *Journal of the Acoustic Society of America*, pages 1738–1752, 1990.

[94] L. Neumeyer, M. Weintraub. Probabilistic optimum filtering for robust speech recognition. In *Proceedings of ICASSP*, volume 1, pages 17–20, 1994.

[95] A. Varga, H. Steeneken. Assessment for automatic speech recognition II: NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems. *Speech Communication*, 12(3):247–251, 1993.

[96] L. Neumeyer, M. Weintraub. Robust speech recognition in noise using adaptation and mapping techniques. In *Proceedings of ICASSP*, volume 1, pages 141–144, 1995.

[97] N. Mesgarani, S. David, J. Britz and S. Shamma. Phoneme representation and classification in primary auditory cortex. *J. Acoust. Soc. Am.*, 132(2):899–999, 2008.

[98] Yiteng (Arden) Huang and Jacob Benesty, editors. *Audio Signal Processing for next-generation multimedia communication systems.* Kluwer Academic Press, 2004.

[99] D. Milone and L. Di Persia and M.E. Torres. Denoising and recognition using hidden markov models with observation distributions modeled by hidden markov trees. *Pattern Recognition*, 43(4):1577–1589, 2009.

[100] X. Wu, P. Biyani, and S. Dumitrescu. Globally optimal classification and pairing of human chromosomes. In *Proc. of the 26 Annual Int. Conf. of the IEEE EMBS*, pages 2789–2792, San Francisco, CA, USA, September 2005.

APÉNDICE A

# Corpus empleados

## A.1 Cpa

La base de datos utilizada en los experimentos es la más numerosa de su tipo, una corrección del corpus *Copenhagen* completo. Este corpus consiste de las imágenes de cromosomas pertenecientes a 2804 células humanas en metafase cariotipadas, 1344 de las cuales son femeninas y 1460 son masculinas. Las imágenes se encuentran segmentadas individualmente por cromosomas, con polaridad definida.

El corpus consta en su mayoría de células normales, aquéllas con 46 cromosomas con 2 cromosomas para clases 1 a 22 más 2 cromosomas sexuales (XX para células femeninas o XY para masculinas). Asimismo, se encuentra también un grupo de células con aberraciones de número, producto de constelaciones anormales o artefactos en la preparación o adquisición de las imágenes. Existen 26 células con un cromosoma faltante, que dan lugar a afecciones genéticas como el síndrome de Turner (donde falta un cromosoma sexual, también denominada "monosomía X"), otros casos de faltantes pueden ser debidor a problemas de adquisición (cualquier cromosoma). En 37 células hay un cromosoma extra, dando lugar a trisomías de autosomas como el síndrome de Down (triple 21), síndrome de Edward (triple 18), o formaciones patológicas del par sexual como el síndrome de Klinefelter (indidivuo masculino con un cromosoma X extra) y otras.

Se realizó una corrección sobre el conjunto original, que consistió en el re-

etiquetado por un experto humano de un subconjunto de 200 cromosomas que aparecieron como *outliers* debido a etiquetados erróneos, y a la corrección de 100 polaridades incorrectamente establecidas en el corpus original. El nuevo conjunto etiquetado de imágenes se denominó Cpa.

## A.2 TIMIT

Se trata de un desarrollo conjunto entre Texas Instruments y el Massachusetts Institute of Technology. Consiste en una serie de emisiones de voz grabadas a través de la lectura de diversos textos por un conjunto de hablantes. Esta base ha sido diseñada para la adquisición de conocimiento acústico-fonético a partir de los datos de voz y para el desarrollo y evaluación de sistemas de reconocimiento automático del habla [87].

TIMIT contiene la voz de 630 hablantes representando las 8 mayores divisiones dialécticas del *inglés americano*, cada uno pronunciando 10 oraciones fonéticamente diversas. El corpus incluye la señal de voz correspondiente a cada oración hablada, así como también transcripciones ortográficas, fonéticas y de palabras alineadas temporalmente. Además los datos vienen ya divididos en subconjuntos de entrenamiento y prueba balanceados para cobertura dialéctica y fonética, lo que facilita también la comparación de resultados.

TIMIT contiene un total de 6300 oraciones de aproximadamente 30 s. de duración cada una (5 horas de audio en total). El 70 % de los hablantes son masculinos y 30 % son femeninos. El material de texto consiste de 2 *oraciones de dialecto* (SA), 450 *oraciones fonémicamente compactas* (SX), y 1890 *oraciones fonéticamente diversas* (SI). Cada hablante lee las 2 SA, 5 de las SX y 3 de las SI.

Las grabaciones fueron hechas en una cabina de grabación aislada de ruidos usando un sistema semiautomático para la presentación del texto al hablante y la grabación. Los datos fueron digitalizados a una frecuencia de muestreo de 20 KHz (16 bits) con un filtro anti-alias en 10 KHz. La voz fue filtrada digitalmente, nivelada (debiased) y submuestreada a 16 KHz. A los sujetos se los estimuló con una señal de ruido de fondo de bajo nivel a través de auriculares para suprimir la inusual calidad de voz producida por el efecto de aislación de la cabina. También se les pidió que leyeran el texto con "voz natural".

## A.3 NOISEX-92

Esta base de datos contiene una serie de ruidos generados y registrados en diferentes condiciones, los cuales pueden ser agregados a las señales de habla limpia a fines de evaluar el comportamiento de sistemas automáticos de reconocimiento robusto de habla o algoritmos de limpieza de ruido [95].

El corpus dispone de los siguientes ruidos:

- Conversación (babble), que se grabó mediante un dispositivo DAT equipado con un micrófono tipo condensador. La fuente del "murmullo" fueron 100 personas hablando en una cantina. El radio del cuarto fue de unos dos metros, por lo que las voces individuales son ligeramente audibles. El nivel de sonido durante el proceso de grabación fue de 88 dB SPL.
- Fábrica.
- Ruido blanco y rosa, digitalizados de un generador de ruido analógico de alta calidad (Wandel & Goltermann) a 19,98 kHz y 16 bits, con igual energía en todo el ancho de banda.
- Equipamiento militar: aviones (F16 y Buccaneer), tanques (Leopard, M109), armas, otros.
- Cabina de automóviles: Volvo 340.

## A.4 AURORA

Este corpus dispone de ruidos adquiridos en ambientes reales, donde alguno de ellos tienen características estacionarias a largo plazo, mientras que otros (como los de calle o aeropuerto) contienen segmentos no estacionarios [78].

Los tipos de ruido incluidos son:

- Conversación (*babble*).

- Automóviles.
- Hall de exhibición.
- Restaurant.
- Calle.
- Aeropuerto.
- Estación de trenes.
- Trenes.

APÉNDICE B

# Experimentos adicionales de clasificación de cromosomas

## B.1 Elección de la topología de modelos ocultos de Markov

Dados un conjunto de características particular, por ej. perfiles de 9 muestras por cuerda, y una partición del corpus, existenten diferencias entre las tasas de error obtenidas para redes de Elman y modelos ocultos de Markov (MOM) en favor de las primeras.

En la búsqueda de errores en los MOM resultantes, se revisaron las matrices de transición de los modelos que logran mejores resultados, observándose típicamente para todas las clases una matriz como la del ejemplo dado en la Tabla B.1.

Se observan estados con probabilidad de transición al mismo estado igual a 0. Estos estados solamente emiten un símbolo y descargan toda la responsabilidad de modelar porciones de cromosoma en estados posteriores. Puede deducirse que los modelos necesitan un factor de carga lo más pequeño posible para obtener una cantidad quizás excesiva de estdos para los MOM.

Un problema inherente al factor de carga es que el número de estados que se obtiene una vez fijado $k$ no toma en cuenta la variabilidad de grises en las imágenes, esto es, podemos tener una imagen sólida de un solo tono de gris y un modelo de gran cantidad de estados, cuando se podrían tener solamente unos 2 ó 3 estados que modelaran ese tipo de imágenes. Un ejemplo de la gran

Tabla B.1: Ejemplo de matriz de transición de un MOM.

| | $a_{ii}$ | $a_{ij}$ |
|---|---|---|
| Estado 1 | $0{,}000000e+00$ | $1{,}000000e+00$ |
| Estado 2 | $5{,}142037e-01$ | $4{,}857963e-01$ |
| Estado 3 | $2{,}622389e-01$ | $7{,}377610e-01$ |
| Estado 4 | $0{,}000000e+00$ | $1{,}000000e+00$ |
| Estado 5 | $0{,}000000e+00$ | $1{,}000000e+00$ |
| Estado 6 | $7{,}101122e-01$ | $2{,}898878e-01$ |
| Estado 7 | $7{,}324596e-01$ | $2{,}675404e-01$ |
| Estado 8 | $0{,}000000e+00$ | $1{,}000000e+00$ |
| Estado 9 | $0{,}000000e+00$ | $1{,}000000e+00$ |
| Estado 10 | $1{,}394970e-01$ | $8{,}605030e-01$ |
| ... | ... | |
| Estado $N-1$ | $5{,}589709e-01$ | $4{,}410291e-01$ |
| Estado $N$ | $4{,}867380e-01$ | $5{,}132620e-01$ |

$N$: número de estados emisores para el MOM

cantidad de estados que otorga $k$ se ve en la Figura B.1 para un cromosoma largo, un cromosoma de longitud media, y un cromosoma corto.

Comparando la cantidad de parámetros a estimar en MOM y redes neuronales (RN), vemos que la RN recurrente tiene 53.800 parámetros, mientras que el clasificador con MOM (para $k = 1{,}5$ y 32 densidades gaussianas por estado) posee aprox. 450.000 parámetros, lo cual influye en la tasa de error más alta que obtiene.

Con el objeto de reducir los modelos y, por lo tanto, la cantidad de parámetros libres, se probaron dos aproximaciones iniciales a la búsqueda de la topología óptima (problema totalmente abierto en Reconocimiento de Formas).

La primera aproximación consistió en reducir los modelos de cada estado sustrayendo la cantidad de estados con $a_{ij} = 1{,}0$. Luego se procede a la reestimación mediante Baum-Welch. La segunda aproximación fue la ligadura de parámetros, en este caso se construyeron modelos con una sola matriz de covarianza diagonal para cada MOM a partir de una inicialización plana en donde se calcula la matriz de covarianza global para cada modelo. Ambas aproximaciones no lograron todavía mejorar los resultados con modelos largos, por lo que se debe continuar esta exploración como trabajo futuro.

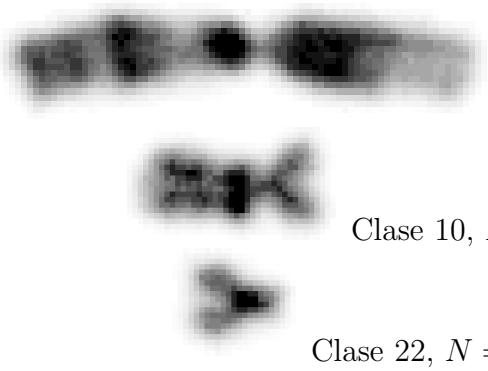



Figura B.1: Número de estados emisores con $k = 1{,}5$.

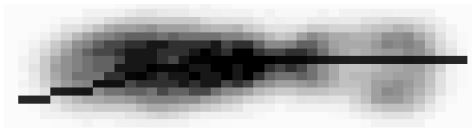


Figura B.2: Desviación debida al filtrado morfológico.

## B.2 Esqueletos paramétricos

Como veremos a continuación con algunos ejemplos, el proceso de esqueletonización conduce a diferentes clases de defectos en los esqueletos obtenidos debido al proceso intrínseco de afinamiento de las imágenes.

### B.2.1 Problemas de la esqueletonización

La Figura B.2 muestra un cromosoma de clase 6 con el esqueleto solapado, donde se observa la desviación típica en el extremo por donde comienza el algoritmo a calcular el esqueleto. Esta desviación hacia un extremo origina que las cuerdas perpendiculares al esqueleto se encuentren mal orientadas.

Otro problema que suele aparecer es el de "ruido" en el esqueleto resultante, lo que conduce a severas perturbaciones en el cálculo de la pendiente de la recta tangente en los puntos de ruido. La Figura B.3 muestra un cromosoma de clase 2 con el esqueleto solapado.

En los cromosomas cortos suele haber problemas de orientación del es-

Figura B.3: "Ruido" en el eje calculado por fallas en la esqueletonización.

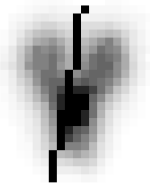


Figura B.4: Mala orientación del esqueleto en cromosomas cortos.

queleto, ya que estos cromosomas tienen aproximadamente la misma dimensión en ancho como en alto, lo que induce errores en el algoritmo de afinamiento. La Figura B.4 muestra un cromosoma de clase 22 con el esqueleto solapado. Obsérvese la desviación del esqueleto obtenido respecto a lo que sería su esqueleto óptimo: un segmento vertical que lo cruzara por su línea media.

Un problema posterior es que el eje medio longitudinal obtenido no es paramétrico, por lo que para calcular la pendiende de la recta tangente en todo punto debe aplicarse una ventana a tramos sucesivos de esqueleto y obtener la tangente por ajuste por valores propios.

Todos estos errores repercuten en el desempeño que pueda lograr el resto del sistema, por lo cual se plantea a continuación un nuevo método de preproceso.

### B.2.2 Esqueletos polinómicos

El método propuesto consiste en obtener un esqueleto paramétrico mediante ajuste por mínimos cuadrados de curvas polinomiales de diferente grado. Con esqueletos paramétricos el cálculo de la pendiente de la recta tangente en todo punto es directo y fiable.

Una vez obtenidos los esqueletos, se desdobla el corpus muestreando todo el cromosoma (con su ancho máximo) sobre las cuerdas perpendiculares. De esta manera, se obtiene un corpus desdoblado que facilita y acelera la extracción de características, ya que cada vez que se necesite extraer una

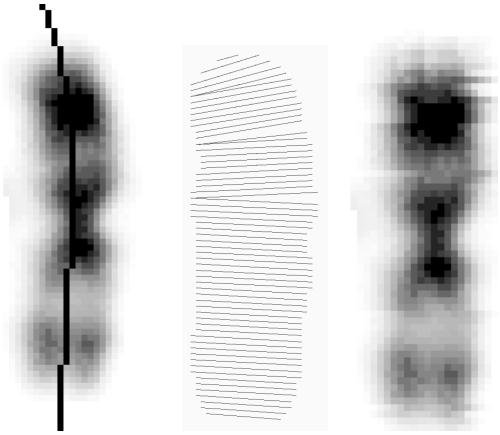


Figura B.5: Cromosoma original con esqueleto (izquierda) - Cuerdas perpendiculares (centro) - Cromosoma desdoblado (derecha).

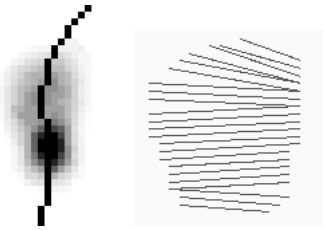


Figura B.6: Cromosoma corto con esqueleto polinómico de grado 3.

característica en particular, bastará con procesar las imágenes por líneas, sabiendo que cada línea corresponde a una cuerda.

La Figura B.5 muestra los resultados de los pasos sucesivos en el desdoblado de cromosomas: ajuste de la curva polinómica, cálculo de las cuerdas y muestreo de grises.

Para el cromosoma de clase 2 de la Figura B.5 se calculó un polimonio de grado 3. Las curvaturas de los esqueletos de grado 3 representan un problema en imágenes de cromosomas cortos, como se ve en la Figura B.6. Heurísticamente se decide que los esqueletos de cromosomas largos serán calculados con polinomios de grado 3, los esqueletos de cromosomas de longitud media serán calculados con polinomios de grado 2, y los esqueletos de cromosomas cortos serán obtenidos mediante ajuste de polinomios de grado 1. La Figura B.7 muestra el mismo cromosoma de clase 22 de la Figura B.6, pero esta vez calculando su esqueleto con polinomio de grado 1. Obsérvese la orientación adecuada del esqueleto en relación al obtenido en la Figura B.4.

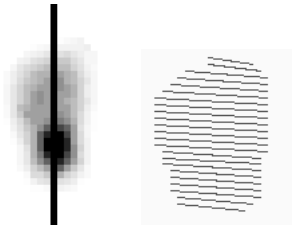


Figura B.7: Cromosoma corto con esqueleto polinómico de grado 1.

Tabla B.2: Comparación de error (en %) sobre patrones calculados con diferente esqueleto.

| Nro. de dist. gauss. por mezcla | Esqueletonización | Polinomios |
|---|---|---|
| 1 | 18 | 14 |
| 2 | 14 | 12 |
| 4 | 12 | 10 |
| 8 | 11 | 9 |

### B.2.3 Resultados iniciales con esqueletos paramétricos

Una serie de primeras pruebas con un conjunto de 9 puntos gaussianos por cuerda, fue realizada utilizando el conjuntos de entrenamiento/prueba de 2400/400 células. El experimento fue realizado empleando MOM de arquitectura fija como clasificador, 9 puntos + derivada horizontal + aceleración horizontal como características, y factor de carga $k = 1{,}5$. Los resultados obtenidos se muestran en la Tabla B.2, donde es posible observar que la nueva aproximación paramétrica al cálculo del eje mejora el desempeño de los esqueletos clásicos.

En estos experimentos se combinaron los grados de los polinomios aproximantes: grado 3 para las clases 1 a 12, grado 2 para las clases 13 a 17 y cromosomas sexuales, y grado 1 para las clases 18 a 22. Claramente, experimentación futura sobre este punto debe ser exhaustivamente realizada a fin de encontrar la combinación óptima de polinomios.

APÉNDICE C

# Experimentos adicionales en clasificación de fonemas

En el ajuste inicial de las redes neuronales y en la búsqueda de una parametrización óptima para el método de representaciones ralas presentado, se realizaron diversos experimentos probando el desempeño en clasificación al utilizar como entrada a la red el espectrograma auditivo, las activaciones a partir de un diccionario completo y de un diccionario 2 veces sobrecompleto. Los átomos correspondieron a la versión submuestreada, con un tamaño de 64 coeficientes frecuenciales y un ancho de 4 ventanas móviles (256 coeficientes). En todos los casos, los fonemas correspondieron al conjunto de 5 fonemas altamente confundibles en inglés, sin ruido adicionado.

La Tabla C.1 muestra los resultados obtenidos, donde los mejores resultados están resaltados en negrita. A los fines comparativos, se dan también las tasas de reconocimiento obtenidas por la parametrización MFCC de referencia, con el coeficiente de energía y deltas agregados.

Como se puede observar, los resultados de la clasificación sobre los datos de entrenamiento y prueba para la representación cortical son mejores que los obtenidos cuando se utiliza la representación auditiva temprana. Para la representación de estos últimos, algunos de los resultados de la clasificación son aparentemente buenos en promedio, sin embargo, al examinar las tasas de clasificación individuales para cada fonema (expuesto en las columnas a la derecha de la Tabla), sólo dos o tres fonemas son, de hecho, clasificados correctamente (ver experimentos Nº 1-8). Este problema surge debido a una solución de error mínimo local que la representación cortical evita (véase el

Tabla C.1: Porcentaje de reconocimiento en clasificación de fonemas al usar patrones construidos del espectrograma auditivo y la representación cortical con dos tamños de diccionario (mejores resultados en negrita). TRN/TST: reconocimiento sobre el conjunto de entrenamiento/prueba.

| Experimento | Red | TRN | TST | */b/* | */d/* | */jh/* | */eh/* | */ih/* |
|---|---|---|---|---|---|---|---|---|
| Auditivo 64x4 | 256/4/5 | 45.84 | 44.76 | 0.00 | 0.00 | 6.90 | 100.00 | 6.27 |
| | 256/8/5 | 44.35 | 43.25 | 0.00 | 0.00 | 4.31 | 100.00 | 3.13 |
| | 256/16/5 | 64.28 | 65.03 | 0.00 | 0.00 | 9.48 | 94.99 | 57.59 |
| | 256/32/5 | 68.92 | 69.67 | 0.00 | 0.00 | 100.00 | 95.87 | 54.82 |
| | 256/64/5 | 70.70 | 72.69 | 0.00 | 0.00 | 83.62 | 72.34 | 86.75 |
| | 256/128/5 | 70.50 | 72.17 | 4.55 | 0.00 | 62.93 | 84.73 | 76.14 |
| | 256/256/5 | 72.15 | 73.74 | 0.00 | 0.00 | 97.41 | 85.23 | 74.82 |
| | 256/512/5 | 69.21 | 71.76 | 0.00 | 0.00 | 100.00 | 94.49 | 60.96 |
| Cortical 64x4 | 256/4/5 | 77.04 | 75.72 | 40.91 | 56.48 | 97.41 | 84.86 | 69.16 |
| | 256/8/5 | 79.64 | **77.64** | 46.97 | 62.96 | 93.97 | 84.86 | 72.77 |
| | 256/16/5 | 75.60 | 76.08 | 65.15 | 51.85 | 97.41 | 89.99 | 63.73 |
| | 256/32/5 | 79.72 | 74.73 | 65.15 | 67.59 | 98.28 | 79.22 | 68.80 |
| | 256/64/5 | 87.27 | 76.86 | 74.24 | 66.67 | 95.69 | 88.24 | 64.82 |
| | 256/128/5 | 100.00 | **78.37** | 72.73 | 70.37 | 96.55 | 78.35 | 77.35 |
| | 256/256/5 | 98.10 | 77.07 | 65.15 | 71.30 | 91.38 | 87.11 | 67.11 |
| | 256/512/5 | 99.92 | **79.16** | 71.21 | 69.44 | 92.24 | 80.35 | 78.07 |
| Cortical 64x4x2 | 512/4/5 | 78.65 | 73.79 | 48.48 | 59.26 | 86.21 | 85.61 | 64.58 |
| | 512/8/5 | 80.62 | 75.51 | 63.64 | 59.26 | 98.28 | 85.36 | 65.90 |
| | 512/16/5 | 78.65 | 74.26 | 54.55 | 53.70 | 99.14 | 82.98 | 66.63 |
| | 512/32/5 | 82.58 | 75.66 | 62.12 | 66.67 | 95.69 | 85.11 | 66.02 |
| | 512/64/5 | 87.27 | 75.87 | 54.55 | 65.74 | 98.28 | 83.48 | 68.43 |
| | 512/128/5 | 84.72 | 75.98 | 65.15 | 56.48 | 95.69 | 84.23 | 68.67 |
| | 512/256/5 | 81.37 | 76.55 | 65.15 | 62.96 | 95.69 | 86.86 | 66.63 |
| | 512/512/5 | 82.64 | 76.32 | 65.15 | 61.11 | 97.41 | 77.97 | 74.70 |
| MFCC+E 14+14 | 28/28/5 | 77.39 | 77.28 | 46.51 | 75.38 | 91.11 | 80.56 | 74.40 |

patrón de la distribución desigual en mostrado en la Tabla 5.1).

Por otra parte, los resultados de la representación cortical son mejores que los obtenidos con la representación clásica MFCC para esta tarea (ver experimentos Nº 16 y 25 en la Tabla C.1). Otro aspecto importante es que el rendimiento es satisfactorio para una red de arquitectura relativamente pequeña en relación a las dimensiones del patrón. Este aspecto corrobora la hipótesis de que las clases están mejor separados en este nuevo espacio de dimensiones superiores, y por lo tanto un clasificador más simple puede completar la tarea con éxito.

La significancia estadística para estos resultados muestra que los mejores resultados de MFCC y la representación cortical se obtiene una probabilidad $Pr(\epsilon_{ref} > \epsilon) > 92\,\%$.