

**UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS  
VILLAS.**

**FACULTAD DE INGENIERÍA ELÉCTRICA.**

**DEPARTAMENTO DE TELECOMUNICACIONES Y  
ELECTRÓNICA.**



**TRABAJO DE MAESTRIA.**

***Separación de consonantes a través de parámetros  
acústicos de energía y sonoridad***

**Autor: Ing. René Antonio Llanes Machado**

**Tutor: Dr. Carlos A. Ferrer Riesgo.**

**Villa Clara  
2012**

## Resumen

Los trastornos motores del lenguaje (TML) son afecciones que producen deficiencia en la producción de la voz. Las disartrias constituyen un grupo principal entre ellos. Debido a que estos trastornos influyen negativamente en la incorporación de los pacientes a la sociedad se le brinda una especial atención a las técnicas que permiten un mejor diagnóstico, seguimiento y/o rehabilitación de las personas con afectaciones en su comunicación oral.

Los autores Darley, Aronson y Brown (DAB) llevaron a cabo una serie de estudios en las Clínicas Mayo donde analizaron las características del habla de grupos de pacientes disártricos desde el punto de vista auditivo. A partir de estos estudios elaboraron una metodología de diagnóstico a través de valoraciones subjetivas de determinados rasgos perceptuales. Entre los resultados más notables se puede mencionar que el rasgo de Consonantes Imprecisas (CI), presentaba la más alta incidencia en 5 de las 7 disartrias analizadas.

Debido a lo anterior, nuestro trabajo pretende mediante técnicas de Procesamiento Digital de Señales, lograr una medida objetiva de imprecisión articulatoria, altamente correlacionada con la valoración del rasgo CI. Se ha observado que la imprecisión se manifiesta generalmente a través sustituciones de las consonantes oclusivas sordas “P”, “T”, y “K” por sus diferentes variaciones en fricación, nasalización y/o sonorización de las tres consonantes. En la caracterización de las consonantes se han empleado índices de sonoridad (ISS) y energía relativa (IES) y varios clasificadores matemáticos para lograr definir regiones de agrupación de las mencionadas consonantes en un plano de Energía vs Sonoridad arribándose a buenos resultados de separación. Este resultado constituye un punto de avance en la determinación objetiva del rasgo CI lo cual puede constituir una herramienta para el diagnóstico y supervisión de pacientes disártricos.

*The motor disorders of language (TML) are conditions that cause deficiency in the production of voice. The dysarthria are a major group among them. Because these disorders have a negative effect on the incorporation of patients to society. A special attention is given to techniques for improved diagnosis, monitoring and / or rehabilitation of people with this affectations in oral communication.*

*The authors Darley, Aronson and Brown (DAB) conducted a series of studies at the Mayo Clinics where they analyzed the speech characteristics of dysarthric patient groups from the standpoint of hearing. From these studies developed a diagnostic methodology through subjective assessments of certain perceptual features. Among the most notable may be mentioned that the trait of imprecise consonants (CI), showed a higher incidence in 5 of the 7 analyzed dysarthria.*

*Because of this, our work aims, through techniques of Digital Signal Processing, to achieve an objective measure of articulatory imprecision, highly correlated with the CI trait assessment. It has been observed that the inaccuracy is usually manifested by substitutions of voiceless consonants "P", "T" and "K" by different variations in frication, nasalization and / or voiced of the three consonants. In the characterization of the consonants were used indices of sonority (ISS) and relative energy (IES) and several mathematical classifiers to achieve defined regions of the aforementioned group of consonants in a plane of energy vs. sonority. Finally obtaining a good separation results. This result is a breakthrough point in the objective determination of CI feature which can be a tool for diagnosis and monitoring of dysarthric patients.*

# Índice

<b>ÍNDICE.....</b>	<b>1</b>
<b>INTRODUCCIÓN.....</b>	<b>4</b>
<b>CAPÍTULO 1. MARCO TEÓRICO REFERENCIAL.....</b>	<b>7</b>
1.1 ALTERACIONES DEL HABLA .....	7
1.2 METODOLOGÍA DAB.....	11
1.3 ANTECEDENTES EN MEDICIÓN DE CONSONANTES IMPRECISAS.....	14
1.3.1 <i>Cuantificación de parámetros de la voz para el diagnóstico de disartrias (2000).</i> .....	14
1.3.2 <i>Detector de Posición de Sílabas Aisladas en Grabaciones de la Señal de Voz de Pacientes con Trastornos Motores del Lenguaje Mediante Técnicas de Procesamiento Digital de Señales (TELEC 2002).</i> ....	15
1.3.3 <i>Estimation of Consonant Imprecision Ratings by using Neural Networks and Acoustic Measures of Energy and Sonority (CIARP 2003).</i> .....	15
1.4 CONCLUSIONES .....	16
<b>CAPÍTULO 2. MATERIALES Y MÉTODOS.....</b>	<b>17</b>
2.1 BASE DE DATOS DE ARCHIVOS DE VOZ.....	17
2.2 PARÁMETROS ACÚSTICOS .....	17
2.2.1 <i>Detector de sílabas.</i> .....	18
2.2.2 <i>Indicador de Energía de la Sílabla (IES).</i> .....	18
2.2.3 <i>Indicador de Sonoridad de la Sílabla (ISS).</i> .....	19
2.3 CLASIFICADORES.....	20
2.3.1 <i>Método de clasificación a través de Sistema de Inferencia Borroso</i> .....	20
2.3.1.1 Funciones de Pertenencia Gaussianas .....	21
2.3.1.2 Funciones de Pertenencia Trapezoidales.....	22
2.3.2 <i>Método de clasificación a través del análisis cuadrático del discriminante.</i> .....	22
2.3.3 <i>Métodos de clasificación a través de un clasificador de naive Bayes.</i> .....	23
2.3.4 <i>Métodos de clasificación a través K Vecinos más cercanos (KNN).</i> .....	24
2.3.5 <i>Método de clasificación a través de árboles de decisión (árboles de clasificación)</i> .....	24
2.4 DESCRIPCIÓN DE LAS PRUEBAS.....	25
2.4.1 <i>Test de normalidad.</i> .....	25
2.4.2 <i>Medidas para la comprobación de la efectividad de los métodos en la clasificación</i> .....	25
2.5 CONCLUSIONES PARCIALES DEL CAPÍTULO.....	27
<b>CAPÍTULO 3. RESULTADOS Y DISCUSIÓN.....</b>	<b>28</b>
3.1 RESULTADOS DE LA DETECCIÓN DE SÍLABAS Y DE LOS TEST DE NORMALIDAD .....	28
3.2 DISTRIBUCIONES DE LAS CONSONANTES EN EL PLANO DE ENERGÍA VS SONORIDAD .....	28

## Índice

3.3	RESULTADOS DE LOS CLASIFICADORES.....	30
3.3.1	<i>Sistema de Inferencia Borroso de tipo Sugeno con función de transferencia trapezoidal (SIST)</i> 30	
3.3.2	<i>Sistema de Inferencia Borroso de tipo Sugeno con función de transferencia gaussiana (SISG) ..</i>	34
3.3.3	<i>Análisis cuadrático del Discriminante (QDA) .....</i>	35
3.3.4	<i>K Vecinos más cercanos (KNN).....</i>	37
3.3.5	<i>Arboles de clasificación.....</i>	39
3.3.6	<i>Clasificador de naive Bayes.....</i>	42
3.4	CONCLUSIONES PARCIALES DEL CAPÍTULO.....	43
	<b>CONCLUSIONES .....</b>	<b>44</b>
	<b>RECOMENDACIONES.....</b>	<b>46</b>
	<b>BIBLIOGRAFÍA .....</b>	<b>47</b>
	<b>ANEXOS.....</b>	<b>50</b>
	A-1 DETECTOR DE SÍLABA UTILIZADO .....	50

## **Introducción**

Dada la importancia que el lenguaje hablado tiene para el hombre, en el mundo se le brinda una especial atención a las técnicas que permiten un mejor diagnóstico, seguimiento y/o rehabilitación de las personas con afectaciones en su comunicación oral. Una de las formas más comunes de alteraciones patológicas del habla la constituyen aquellos trastornos neurológicos que afectan en alguna medida los sistemas involucrados en la producción de la voz. Dentro de estos se encuentran los trastornos motores del lenguaje (TML) donde las disartrias constituyen un grupo principal.

A fines de la década de los 60's y principios de los años 70 Darley, Aronson y Brown (DAB) llevaron a cabo una serie de estudios ([2][3][4][5][6][7][8]) en las Clínicas Mayo donde analizaron las características del habla de grupos de pacientes disártricos desde el punto de vista auditivo. A partir de estos estudios elaboraron una metodología de diagnóstico a partir de las valoraciones subjetivas de determinados rasgos perceptuales los cuales fueron evaluados en escala de 0 a 6 en nivel ascendente de severidad a partir de la realización de tres ejercicios: fonación de una vocal "a" sostenida durante un tiempo no menor de 8 segundos, repetición lo más rápido y uniformemente posible de los monosílabos con consonantes explosivas sordas "Pa", "Ta" y "Ka" y la lectura de un segmento del párrafo "El Abuelo", utilizado con frecuencia en la literatura para el análisis de voces patológicas. Entre los resultados más notables se puede mencionar que el rasgo de Consonantes Imprecisas (CI), perceptible en el ejercicio de pronunciación de los monosílabos, es el de más alta puntuación en 5 de las 7 disartrias analizadas.

Debido a lo anterior, Ferrer y sus colaboradores han desarrollado trabajos ([12][13][11][15]) orientados a obtener, mediante técnicas de Procesamiento Digital de Señales, una medida objetiva de imprecisión articulatoria, altamente correlacionada con la valoración del rasgo CI. En estos trabajos se ha observado que la imprecisión se manifiesta generalmente a través sustituciones de las consonantes oclusivas sordas mencionadas por sus diferentes variaciones en fricación, nasalización y/o sonorización de las tres consonantes. En la caracterización de las consonantes (correcta o incorrectamente pronunciadas) se han empleado índices de sonoridad (ISS) y energía relativa (IES), y alguna manera de combinación de estos (regresiones, redes neuronales, etc.) como estimador de la valoración del rasgo CI. Se ha partido en ellos de la

## Introducción

hipótesis de que los distintos tipos de consonantes son apropiadamente separables en el plano IES, ISS. Sin embargo, esta hipótesis no ha sido evaluada en la práctica, y un conocimiento de la distribución de las consonantes en dicho plano puede favorecer el desarrollo de mejores estimadores del rasgo CI.

Este trabajo se propone determinar las regiones ocupadas por los distintos tipos de consonantes, y evaluar su separabilidad mediante el empleo de sistemas de clasificación.

Esta investigación se desarrolla en un ambiente cuantitativo. Constituye un estudio descriptivo ya que el propósito es describir situaciones y eventos midiendo y evaluando componentes del fenómeno. El tipo de diseño es no experimental ya que se observaran estos fenómenos como ocurren en la realidad, no se recreará una situación artificial para su análisis.

### ***Problema Científico:***

¿Son separables las consonantes P, T, K usadas en el ejercicio de diadocokinesis oral y sus sustituciones más frecuentes según los rasgos acústicos de energía y sonoridad?

### ***Objetivo general:***

Separar las consonantes P, T, K y sus sustituciones más frecuentes usadas en el ejercicio de diadocokinesis oral según los rasgos acústicos de energía y sonoridad

### ***Objetivos específicos:***

- Identificar los referentes teóricos sobre el procesamiento de digital de señales de voz y los parámetros acústicos.
- Obtener una muestra representativa de pares Energía, Sonoridad correspondientes a las consonantes explosivas (P, T y K) y de sus sustituciones más frecuentes (nasalización, fricación y sonorización) en el ejercicio de diadocokinesis oral.
- Utilizar clasificadores para separar las regiones donde se concentran las diferentes consonantes en el plano Energía vs Sonoridad.
- Evaluar a través de sensibilidad, especificidad y tasa de clasificación la separación lograda por los clasificadores analizados.

### ***Hipótesis de Investigación***

- Los indicadores acústicos de energía y sonoridad introducidos a clasificadores permiten distinguir las consonantes del ejercicio de diadocokinesis oral y sus sustituciones más frecuentes.

## Introducción

Definidos el problema, los objetivos e hipótesis de la investigación se pasará a describir un análisis de la literatura referente al tema.

## Capítulo 1 Marco Teórico Referencial

### Capítulo 1. Marco Teórico Referencial

En este capítulo se hace una introducción al tema de las disartrias, su definición, características acústicas y ubicación entre los desórdenes del habla. Se hace una revisión bibliográfica de los trabajos relacionados con aspectos como el diagnóstico diferencial de disartrias basado en la percepción auditiva, diagnóstico de patologías del habla mediante procesamiento de la señal de voz o parámetros extraídos de esta. Se exponen algunos trabajos previos en la obtención, mediante el procesamiento digital de la señal de voz, de parámetros que cuantifican características patológicas de esta. Al final se presentan que existe un problema no resuelto, a cuya solución se pretende contribuir.

### Capítulo 2. Alteraciones del Habla

Bajo el término Trastornos Motores del Lenguaje (TML) se agrupan las diferentes afecciones que producen una alteración en el control sobre los movimientos musculares del lenguaje, como consecuencia de una lesión del sistema nervioso central o periférico [5]. Se reconocen dos clases de TML que son la Disartria y la Apraxia del lenguaje.

La apraxia del lenguaje es sinónimo de disartria a nivel cortical y desintegración fonética. El control muscular está intacto, hay una pérdida del conocimiento de la articulación de la palabra para producir fonemas y para programar palabras. En contraste con la disartria no hay distorsión del sonido del lenguaje sino sustituciones fonéticas, omisiones, adiciones, o inversiones de los sonidos. Uno de los caracteres distintivos de la apraxia del lenguaje es que el lenguaje automático o de reacción es frecuentemente hablado sin errores mientras que el lenguaje volitivo o con propósito contiene los fenómenos ya descritos.

La Disartria se define como cualquier combinación de trastornos neurológicos coexistentes de algunos o todos los procesos básicos del habla: respiración, fonación, articulación, resonancia y prosodia. Es causada por disturbios en el control muscular de estos procesos, como debilidad, lentitud, descoordinación o alteración del tono muscular, debidos al daño del sistema nervioso central o periférico que provoca este TML. Como causas de este daño al sistema nervioso central o periférico se pueden mencionar los accidentes vasculares y traumas cerebrales, enfermedades degenerativas del cerebelo, el mal de Parkinson, la desmielinización de la sustancia blanca del sistema nervioso central, algunas infecciones como el estafilococo causante de la Corea, y causas tóxicas como la acumulación de cobre en distintas zonas del cerebro por un trastorno metabólico en la enfermedad de Wilson. La voz

## Capítulo 2 Alteraciones del Habla

defectuosa durante la articulación se debe a descoordinación del control muscular pero el centro del control del lenguaje no está afectado y el conocimiento del lenguaje es normal.

Las disartrias constituyen una familia de trastornos motores del lenguaje, cada una con características acústicas distintivas, debido a que se originan en una zona motora diferente del sistema nervioso y por tanto, afectan de forma distinta el proceso del habla. Entre ellas se pueden mencionar:

- Disfonía Flácida por parálisis unilateral de las cuerdas vocales o lesión unilateral de cuerdas vocales o lesión unilateral del nervio laríngeo recurrente.
- Disartria Flácida causada por parálisis de la musculatura fonatoria-articulatoria o por parálisis múltiple bilateral de pares de nervios craneales.
- Disartria Espástica resultante de espasticidad de la musculatura periférica del lenguaje debido a lesión bilateral de la primera neurona motora.
- Disartria Atáxica debido a descoordinación muscular por lesión del cerebelo.
- Disartria Hipocinética en parkinsonismo por rigidez de la musculatura periférica del lenguaje causado por una lesión en los ganglios basales.
- Disartria Hipercinética asociada a la Corea y caracterizada por movimientos rápidos, descoordinados y espontáneos del sistema nervioso
- Disartria Hipercinética asociada a la Distonía, consiste en movimientos lentos, continuos y descoordinados, producida también por lesión en los ganglios basales.
- Disartria Hipercinética llamada temblor orgánico de la voz, causada por movimientos espontáneos rítmicos de la musculatura laríngea y producida por una lesión de origen desconocido.
- Disartria Hipercinética conocida como Palato-faríngeo-laríngeo-mioclono, caracterizada por movimientos espontáneos de la laringe, faringe, y el paladar blando, producida por una lesión en el tallo cerebral.
- Disartria Hipercinética encontrada en el Síndrome Gill de la Tourette. Es un raro desorden caracterizado por ladridos incontrolados y coprolalia, debido a una lesión de origen desconocido.
- Disartria Mixta encontrada en la Esclerosis Lateral Amiotrófica, en la cual aparecen componentes flácidos y espásticos simultáneamente debido a un trastorno genético de las neuronas motoras central y periférica.

## Capítulo 2 Alteraciones del Habla

- Disartria Mixta encontrada en la Esclerosis Múltiple en la cual pueden encontrarse simultáneamente componentes espásticos y atáxicos debido a la desmielinización del cerebelo y la primera motoneurona.
- Disartria Mixta encontrada en la enfermedad de Wilson en la cual pueden aparecer espástica, atáxica, hipercinética o combinación de estas en dependencia del cúmulo excesivo de cobre en neuronas del cerebelo, ganglios basales y sistema piramidal.

Darley, Aronson y Brown [5] proporcionaron las siguientes descripciones perceptuales de algunas disartrias:

- Parálisis Pseudobulbar (Disartria Espástica) típicamente el habla es lenta y laboriosa, la articulación es más bien imprecisa con más incidencia en los grupos consonánticos complejos. El tono de la voz es bajo y monótono. La calidad de la voz es áspera (chillona) y forzada o con sonidos estrangulados. Puede haber una hipernasalidad considerable, pero la emisión nasal no es audible en general.
- Parálisis Bulbar (Disartria Flácida) la característica más prominente del habla es la hipernasalidad asociada con emisión nasal de aire. La inspiración audible y la espiración jadeante se presentan a menudo. La pérdida de aire también se manifiesta como frases cortas. La articulación es imprecisa por dos razones:
  - Las consonantes pueden ser débiles debido a la insuficiente presión de aire intraoral causada por la incompetencia velofaríngea.
  - Inmovilidad de la lengua y los labios debido a lesiones de los nervios hipoglosal y faciales impidiendo una producción normal de vocales y consonantes.
- Esclerosis Lateral Amiotrófica (Disartria Flácida–Espástica) tiene un efecto progresivo en el lenguaje. En una etapa temprana predominan características de una de las dos disartrias o de ninguna. En una etapa avanzada ambos conjuntos de las características ya descritas coexisten: Razón lenta, tono bajo, calidad de la voz ronca y forzada-estrangulada, articulación muy defectuosa, marcada hipernasalidad y emisión nasal, todo lo cual contribuye a producir frases cortas escasamente inteligibles.
- Trastornos Cerebelares (Disartria Atáxica) usualmente produce uno de dos patrones de disturbios del lenguaje, estos aparecen concurrentemente:
  - Desintegración intermitente de la articulación, junto con alteraciones en el ritmo e irregularidades del tono y del nivel.

## Capítulo 2 Alteraciones del Habla

- Prosodia alterada conllevando a prolongación de sonidos, inadecuada acentuación de sílabas y palabras, intervalos prolongados entre sílabas y palabras, el habla es artificialmente plana con un paso medido.
- Parkinsonismo (Disartria Hipocinética) énfasis vocal reducido, los picos y valles del tono y las variaciones del nivel son aplanados monótonamente. Ráfagas cortas del habla son separadas por pausas inapropiadas. La razón puede ser variable, a menudo acelerada. La articulación de consonantes en el habla contextual y la repetición de sílabas es borrosa como si los músculos fallaran antes de realizar su movimiento completo. Se muestra dificultad en el inicio de la articulación en la repetición de sonidos y silencios inapropiados. La voz es jadeante a veces y el nivel reducido, casi inaudible.
- Distonía (Disartria Hipercinética) como consecuencia de movimientos involuntarios corporales y faciales, aparecen paradas impredecibles en el flujo del habla, desintegración de la articulación, excesivas variaciones en el nivel y vocales distorsionadas. Para compensar estas alteraciones, la prosodia normal se altera haciendo la razón más lenta, se reducen las variaciones del tono y el nivel, aparecen intervalos prolongados entre las palabras y se intercalan silencios inapropiados.
- Corea (Disartria Hipercinética) produce movimientos involuntarios que provocan una alteración del ciclo normal de la respiración, la espiración es en forma de ráfaga rápida; explosión del nivel con un incremento en la alocución, elevación del tono y desintegración de la articulación. Para anticiparse a las rupturas se varía la razón introduciendo y prolongando las pausas e igualando la acentuación en sílabas y palabras.

Las disartrias son trastornos bastante frecuentes en la población, de forma particular en edades avanzadas, pues generalmente son provocadas por enfermedades crónicas degenerativas del sistema nervioso.

Los trabajos encontrados en la bibliografía sobre el tema de las disartrias han tratado siempre con gran número de pacientes: 212 casos en [4][5][6], 168 casos en [8], 20 casos [2][3], 79 casos [10], 107 casos [9]. Esto constituye un indicador de la incidencia cuantitativa de las disartrias. Desde el punto de vista cualitativo tampoco es despreciable la influencia que tienen, pues los desórdenes en el habla atentan contra la plena y normal incorporación de los

## Capítulo 2 Alteraciones del Habla

afectados a la sociedad, de aquí el interés que se le presta en países desarrollados a las investigaciones sobre el tema [18].

### 2.1 Metodología DAB

En una serie de estudios llevados a cabo en las Clínicas Mayo en los finales de la década de los 60's y principios de los 70, Darley, Aronson y Brown examinaron las características perceptuales del habla de grupos de disártricos. Los resultados de estos estudios y de trabajos relacionados estrechamente con ellos, se consideran aún hoy las bases del diagnóstico diferencial clínico de las disartrias [32]. En la parte más abarcadora del mencionado estudio ([4][5][6]), todos los pacientes fueron unívocamente diagnosticados como representante de una de las siete categorías mencionadas. Formaron la muestra 30 pacientes de cada categoría con excepción del Parkinson donde la muestra fue de 32 pacientes.

Cada paciente ejecutó tres actividades habladas que clínicamente han demostrado (según las Clínicas Mayo) que proveen un máximo de información en el menor tiempo posible:

Una muestra del lenguaje contextual obtenida poniendo al paciente a leer en voz alta párrafos estándar y permitiendo al evaluador medir la forma en que el paciente integra las variadas características fonatorias, resonatorias y prosódicas del lenguaje o discurso contextual.

Prolongación de la vocal /a/ con el propósito de permitir al evaluador de juzgar la calidad, amplitud, duración y persistencia del control fonatorio.

Repeticiones de los sonidos “Pa”, “Ta”, “Ka”, tan rápido y parejo como sea posible con el objetivo de proveer información acerca del ritmo, regularidad, y duración de todo tipo de movimientos articulatorios.

Como párrafo estándar se empleó la lectura de un segmento del párrafo del abuelo:

*“Do you wish to know all about my grandfather? Well, He is nearly 93 years old and he still keeps his mind as swiftly as ever. He wears an old black frac coat, usually several buttons missing. His long beard clings from his chin, giving those who observe him a profound feeling of the outmost respect”*

Basándose en estos tres ejercicios, un panel de tres especialistas evaluó auditivamente la percepción de 38 dimensiones subjetivas, en una escala de 7 puntos igualmente espaciados (0 no perceptible, 6 extremadamente notable). Las dimensiones evaluadas se listan a continuación:

## Capítulo 2 Alteraciones del Habla

### Características del tono

- Nivel del período.
- Rupturas del período.
- Monotonía.
- Temblor de la voz.

### Asociadas al nivel

- Nivel monótono.
- Excesivas variaciones de nivel.
- Decadencia del nivel.
- Nivel alterno.
- Descripción general del nivel.

### Calidad de la voz

- Voz áspera, tosca.
- Voz ronca o húmeda.
- Voz resollante (continua).
- Voz resollante (transitoria).
- Voz forzada - estrangulada.
- Interrupciones de la voz.
- Hipernasalidad.
- Hiponasalidad.
- Emisión nasal.

### Relativas a la Respiración

- Inspiración - espiración audible.
- Inspiración audible.
- Gruñido al final de la espiración.

### Prosodia

- Razón.
- Frases cortas.
- Incremento de la razón segmentos.
- Incremento de la razón en general.
- Énfasis reducido.
- Razón variable.
- Intervalos prolongados.
- Silencios inapropiados.
- Ráfagas cortas del habla.
- Acentuación igual y excesiva.

### Articulación

- Consonantes imprecisas.
- Fonemas prolongados.
- Fonemas repetidos.
- Caídas irregulares de la articulación.
- Vocales distorsionadas.

### Impresión general

- Inteligibilidad.
- Habla rara, grotesca.

La evaluación de los tres especialistas para cada dimensión se promedia, obteniéndose de esta forma el valor asignado a esa característica acústica. Al efectuar este procedimiento a los distintos grupos de disartrias se encontró que eran notables (con un valor medio mayor de 2) las dimensiones por disartria que se muestran a continuación (junto a la dimensión se muestra entre paréntesis el valor medio obtenido).

## Capítulo 2 Alteraciones del Habla

### Parálisis Pseudobulbar

- Consonantes imprecisas (3.98)
- Tono Monótono (3.72)
- Acento reducido (3.32)
- Voz áspera, tosca (3.23)
- Nivel monótono (2.98)
- Tono bajo (2.82)
- Razón lenta (2.66)
- Hipernasalidad (2.64)
- Voz estrangulada – forzada (2.49)
- Frases Cortas (2.41)

### Parálisis Bulbar

- Hipernasalidad (3.61)
- Consonantes Imprecisas (2.91)
- Voz jadeante (Continua) (2.28)
- Tono Monótono (2.09)

### Esclerosis Lateral Amiotrópica

- Consonantes Imprecisas(4.39)
- Hipernasalidad (3.14)
- Voz áspera, tosca (3.00)
- Razón lenta (2.89)
- Tono monótono (2.77)
- Frases Cortas (2.69)
- Vocales distorsionadas (2.60)
- Tono bajo (2.59)
- Nivel Monótono (2.51)
- Acentuación igual y excesiva (2.33)
- Intervalos prolongados (2.22)

### Lesiones Cerebrales

- Consonantes imprecisas (3.19)
- Acentuación igual y excesiva (2.69)

- Caídas articulatorias irregulares (2.59)
- Vocales distorsionadas (2.14)
- Voz áspera, tosca (2.10)

### Parkinsonismo

- Tono monótono (4.64)
- Acento reducido (4.44)
- Nivel monótono (4.26)
- Consonantes imprecisas (3.59)
- Silencios inapropiados (2.40)
- Ráfagas Cortas (2.22)
- Voz áspera, tosca (2.08)
- Voz jadeante (continua) (2.04)

### Distonía

- Consonantes imprecisas (3.82)
- Vocales distorsionadas (2.41)
- Voz áspera, tosca (2.40)
- Caídas articulatorias irregulares (2.28)
- Voz estrangulada- forzada (2.14)
- Tono monótono (2.14)
- Nivel monótono (2.01)

### Corea

- Consonantes imprecisas (2.93)
- Intervalos prolongados (2.56)
- Razón variable (2.29)
- Tono monótono (2.23)
- Voz áspera, tosca (2.20)
- Silencios inapropiados (2.17)
- Vocales distorsionadas (2.13)

## Capítulo 2 Alteraciones del Habla

- Variaciones excesivas del nivel (2.04)

Una revisión sugiere que existen similitudes entre las diferentes disartrias. De hecho, cuatro magnitudes (tono monótono, nivel monótono, voz áspera y consonantes imprecisas, en orden ascendente de valoración) recibieron un valor medio en la escala de 1.5 o superior en todos los grupos neurológicos y tres magnitudes (frases cortas, acento reducido y vocales distorsionadas) lo obtuvieron en cinco de los siete grupos [22].

Como resumen de la metodología DAB, se plantea que las disartrias son una familia de trastornos motores del lenguaje cada una caracterizada por caracteres acústicos distintivos. Un tipo acústico diferente de disartria se origina en una zona motora distinta del sistema nervioso. Estas diferencias pueden apreciarse por el personal entrenado y por lo tanto al oírse un trastorno del lenguaje se puede, siguiendo la metodología desarrollada en las Clínicas Mayo, realizar un diagnóstico diferencial e inferir la topología de la lesión en el sistema nervioso[17][21].

En cuanto a los intereses que motivan este trabajo, el resultado a destacar es que el rasgo de consonantes imprecisas (CI) es el más alterado en general, para todas las disartrias analizadas. Esto ha provocado que se hayan dedicado esfuerzos a la obtención de índices objetivos que permitan cuantificar este rasgo[23][24][27][29][28][30][31][34][35][36], por la posible utilidad en la evaluación del estado del paciente, la efectividad del tratamiento aplicado, o sencillamente la documentación de su evolución[33].

### 2.2 Antecedentes en medición de Consonantes Imprecisas

A continuación se reseñarán un conjunto de trabajos que han precedido el presente y han tenido incidencia en el.

#### 2.2.1 Cuantificación de parámetros de la voz para el diagnóstico de disartrias (2000).

En dicho trabajo [12] sale a relucir la creación un detector de sílabas, por la necesidad de extraer parámetros de articulación. Se implementan indicadores con el fin de satisfacer la necesidad de establecer una medida de la perceptibilidad de la explosión relacionada con la rápida transición del silencio a la liberación de la energía acumulada en la oclusión del tracto. Teniendo en cuenta que las CI parten del hecho, de que en el ejercicio “Pa-Ta-Ka” las consonantes empleadas son explosivas, por lo que en la curva de Intensidad vs. Tiempo pueden apreciarse las zonas de silencio previas a la explosión seguidas de otra zona con intensidad considerable, correspondientes a la explosión y la vocal “a” contigua, respectivamente.

## Capítulo 2 Alteraciones del Habla

### 2.2.2 Detector de Posición de Sílabas Aisladas en Grabaciones de la Señal de Voz de Pacientes con Trastornos Motores del Lenguaje Mediante Técnicas de Procesamiento Digital de Señales (TELEC 2002).

En este trabajo [14] se hace una comparación entre el detector creado en [12] y un método clásico basado en la máquina de estados, resultando el detector creado muy superior, ya que los errores cometidos por el método de la máquina de estados duplican los del método propuesto. Este método obtuvo un 89% de detección correcta de sílabas.

### 2.2.3 Estimación de índice de imprecisión de consonantes usando una red neuronal y parámetros acústicos de energía y sonoridad. (CIARP 2003).

Aquí [15] se emplean Perceptrones Multicapa (MLP) para predecir la severidad del rasgo CI, partiendo de una suposición de que las explosivas sordas y las tres sustituciones posibles (sonorización, fricación y nasalización) son separables en el plano CIE/CIS. Se obtienen resultados aceptables de correlación (0.73) entre las valoraciones emitidas por un panel de jueces, y la predicción producida por un MLP relativamente sencillo (tres capas, con 2/2/1 neuronas en cada capa). El valor de correlación está en el orden de las correlaciones entre jueces (0.75). Finalmente se plantea la hipótesis de la distribución de las consonantes en un plano de  $EvsS$  como se muestra en Fig. 1 y Tabla 1 [20][1].

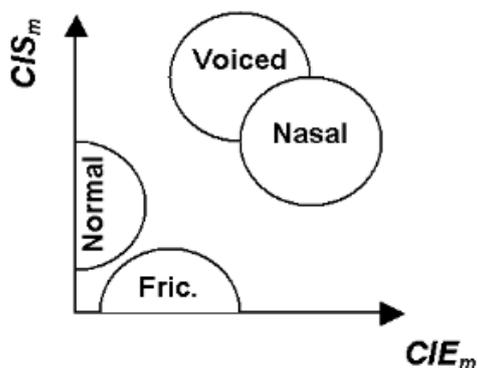


Fig. 1. Distribución de las consonantes en el plano CIEvsCIS. Tomado de CIARP 2003.

## Capítulo 2 Alteraciones del Habla

Tabla 1. Características de energía y sonoridad de las consonantes relacionadas al ejercicio de diadocokinesis oral

Grupos de Consonantes	Energía relativa a la vocal	Sonoridad
PTK (Oclusivas sordas)	BAJA	MEDIA
BDG (Oclusivas sonoras)	MEDIA	ALTA
FSJ (Fricativas)	BAJA	BAJA
MN (Nasales)	ALTA	ALTA

### 2.3 Conclusiones

El rasgo de Consonantes Imprecisas es una alteración común en muchos trastornos motores del lenguaje. De hecho, es el rasgo más prominente en 5 de las 7 disartrias analizadas en los estudios de las clínicas Mayo. Por este motivo resulta de marcado interés su medición desde el punto de vista objetivo. Trabajos anteriores han abordado esta tarea en el ejercicio de la repetición de sílabas explosivas sordas “P”, “T” y “K”, típico para evaluar alteraciones de la articulación. Las alteraciones más frecuentes reportadas en el mismo son las sustituciones de estas consonantes explosivas por sus variantes africadas, sonorizadas o nasalizadas. Este trabajo aborda la problemática de encontrar un método apropiado de separación de las consonantes explosivas sordas de sus distintas desviaciones en el plano *CIE vs CIS*.

### Capítulo 3. Materiales y Métodos

En este capítulo se describirá el conjunto de señales que se han utilizado para realizar los experimentos. Se detallarán los algoritmos empleados en la determinación de los indicadores de interés. También se describirán los métodos de clasificación empleados para la conformación de los grupos de consonantes, las originales “P”, “T” y “K”, así como sus desviaciones más comunes hacia la sonorización, fricación y nasalización. Por último se realiza un análisis de los experimentos llevados a cabo.

#### 3.1 Base de datos de archivos de voz

Para la determinación de los indicadores que nos permitirán definir las regiones en las cuales se agrupan las consonantes se conto con la colaboración de profesores y estudiantes de la Facultad de Ingeniería Eléctrica. Se escogieron personas siguiendo un muestreo simple aleatorio (MÁS). Para realizar este muestreo se tuvieron el cuenta tres aspectos fundamentalmente, voluntad para realizar el estudio, no tener patologías del habla registradas y por último el sexo de la persona. La mitad de estas señales corresponden a hombres y la otra mitad a mujeres, todos sin patologías del habla registradas y entre 20 y 35 años de edad. Se grabaron un total de 20 personas y 11 señales por cada persona. Era necesario obtener una base de datos de sujetos sanos para que los valores de energía y sonoridad extraídos de ellas tipificaran el comportamiento de las consonantes en un plano de  $EvsS$ . Las grabaciones contienen las consonantes del ejercicio original “Pa”, “Ta” y “Ka” pertenecientes a consonantes sordas, así como sus desviaciones más comunes, es decir, “Ba”, “Da” y “Ga” (sonorización); “Fa”, “Sa” y “Ja” (fricación); “Ma” y “Na” (nasalización). Estas señales fueron registradas utilizando un micrófono Hama, modelo CS-460. La frecuencia de muestreo utilizada en su adquisición fue de 22050 Hz, con 16 bits de resolución, y fueron guardadas utilizando el programa *Sound Forge* de la Compañía Sony y el programa *ECAH* desarrollado por el Departamento de Voz del Centro de Estudios de la Electrónica y Tecnologías de la Información (CEETI).

#### 3.2 Parámetros acústicos

Antes de comenzar a explicar de forma general el funcionamiento de los clasificadores utilizados, es necesario describir como se realizo la extracción de los parámetros introducidos a dichos clasificadores.

## Capítulo 3 Materiales y Métodos

### 3.2.1 Detector de sílabas

Para la implementación del algoritmo se utilizó el software matemático Matlab, que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio, disponible para Windows.

Antes de efectuar el cálculo de los indicadores se requiere ubicar las sílabas dentro de la grabación, y en particular, la transición de la consonante a la vocal. Con este propósito se emplea el detector de sílabas reportado en [12], descrito sintéticamente a continuación.

El detector se basa en la detección de máximos en la envolvente de la señal de energía de la señal, filtrada pasabajos a 50 Hz. Los máximos que no satisfacen ciertas condiciones para ser considerados como una sílaba son eliminados y el resto son considerados como centro de sílabas. Las condiciones usadas para definir un máximo son las siguientes:

- Amplitud mayor que veinte veces el menor valor de la envolvente de energía.
- Separación entre máximos mayor que 100 ms (cuando dos máximos están más cercanos que este valor el que posee menor amplitud no se tomará en cuenta).
- La presencia de un mínimo de menos del 75 % de la máxima amplitud entre él y los máximos posteriores y anteriores.
- Separación entre los mencionados mínimos de más de 50 ms.

Una vez que las sílabas son detectadas el instante de liberación es determinado como el punto donde la envolvente de la energía posee su mayor valor de pendiente positiva en el segmento de tiempo entre el centro de esta sílaba y el centro de la anterior. Este método de determinación de la sílaba, tiene un 89% de detección correcta en un conjunto de 3750 sílabas de pacientes disártricos. El programa con la implementación del detector de sílabas se muestra en el Anexo 1.

### 3.2.2 Indicador de Energía de la Sílabas (IES)

El IES (Indicador de Energía de la Sílabas) [13] se define a partir de la relación entre el área de la envolvente de energía  $E(t)$  antes y después de la liberación de la constricción (1):

$$IES = \frac{A_b}{A_a} = \frac{\sum_{T_{on}-T_r}^{T_{on}-T_r} E(t)}{\sum_{T_{on}+T_r}^{T_{on}+T_r} E(t)} \quad (1)$$

## Capítulo 3 Materiales y Métodos

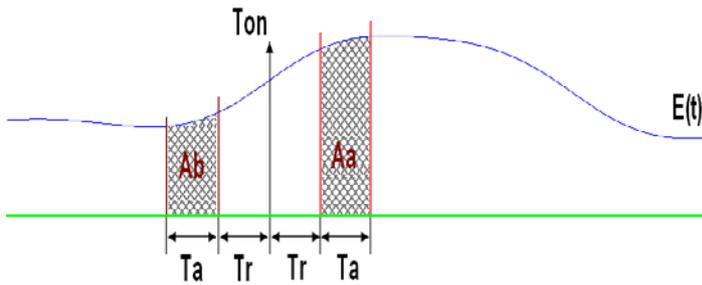


Fig. 2. Representación gráfica de las áreas y los instantes de tiempo utilizados en el cálculo de IES.

Los elementos que se tienen en cuenta en el cálculo de IES se muestran en la Fig. 2. El instante de liberación de la consonante, o lo que es lo mismo el inicio de la vocal se denota como  $T_{on}$ . Los intervalos de tiempo  $T_r$  se refieren al tiempo de guarda durante el cual se asume que está ocurriendo la transición y no hay información de interés. Por otra parte  $T_a$  definido como tiempo de análisis se encuentra escogido en el segmento temporal donde se encuentran las porciones de menor y mayor área bajo la curva o energía de la sílaba. Los valores de estos elementos fueron obtenidos a través de una búsqueda exhaustiva logrando los máximos valores de detección de sílabas en [19].

El tiempo de guarda para el cálculo del indicador de energía es de 10 ms. El valor del tiempo de análisis es de 26 ms.

### 3.2.3 Indicador de Sonoridad de la Sílabas (ISS)

El ISS (Indicador de Sonoridad de la Sílabas) descrito en [13] se calcula como la magnitud del máximo de la función de autocorrelación en el intervalo correspondiente a los valores posibles de período fundamental (2 ms a 14.3 ms, 70 Hz a 500 Hz). La autocorrelación se halla para un segmento de 30 ms, previos a  $T_{on} - T_r$ . En este trabajo se introduce una corrección al valor calculado según el procedimiento descrito anteriormente. La corrección compensa la reducción del pico de autocorrelación causada por el tamaño finito del segmento de análisis. En la Fig. 3 se muestran los elementos involucrados en el cálculo tanto del IES original como de la corrección introducida en este trabajo, y la fórmula se brinda en (2), donde  $W_L$  es el tamaño de la ventana de análisis, 30 milisegundos. La línea recta roja representa el valor máximo que puede alcanzar la función de autocorrelación a causa de la duración finita del segmento. Al igual que en el caso del IES estos elementos fueron recalculados de forma exhaustiva logrando los máximos valores de detección de sílabas.

El tiempo de guarda final utilizado para el cálculo del indicador de energía es de 16 ms. El valor del tiempo de análisis final es de 25 ms.

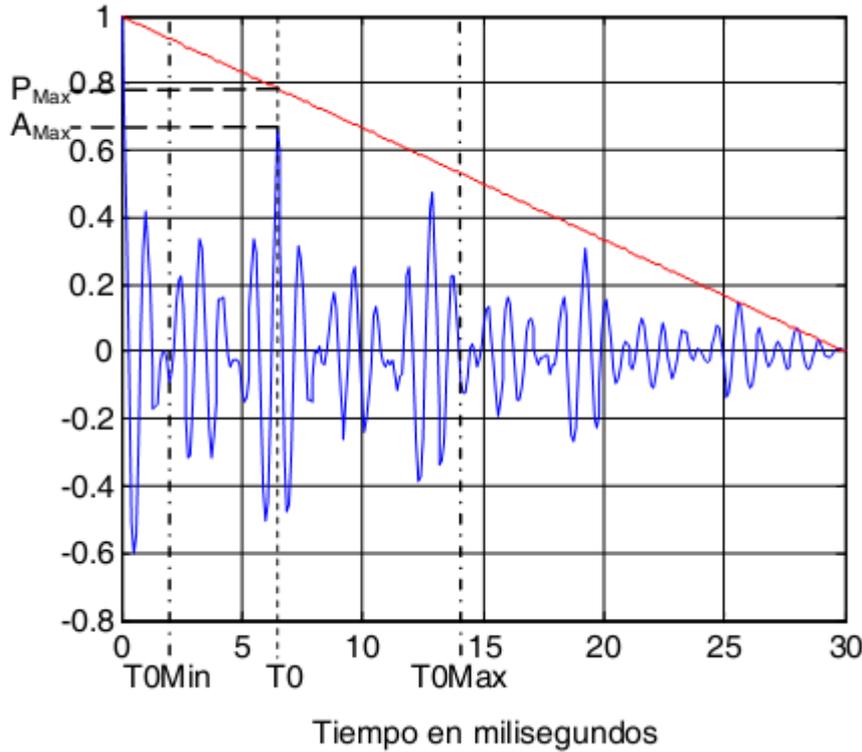


Fig. 2. Representación gráfica de los elementos utilizados en el cálculo de ISS.

$$ISS_{Orig} = A_{Máx}$$

$$ISS = \frac{A_{Máx}}{P_{Máx}} = \frac{A_{Máx} W_L}{W_L - T_0} \quad (2)$$

### 3.3 Clasificadores

A continuación se explicara brevemente el funcionamiento de los clasificadores utilizados.

#### 3.3.1 Método de clasificación a través de Sistema de Inferencia Borroso

Los sistemas de inferencia borrosos pueden verse como una generalización de los sistemas de lógica binaria booleana, donde a partir de valores de entrada se definen valores de salida. Una descripción de estos sistemas puede encontrarse en [25], donde se lee (pág. xxxiii).

*Esquemáticamente, en el desarrollo de un sistema borroso en primer lugar se proponen las variables lingüísticas (borrosas) que definen el sistema, junto a sus posibles estados y funciones de pertenencia, para formular entonces un conjunto de reglas que definan la operación (en ocasiones bastan unas pocas reglas), de cuya aplicación se infiere una respuesta. El sistema resultante suele ser tan sencillo, y los recursos de cálculo tan*

### Capítulo 3 Materiales y Métodos

reducidos, que todo puede incorporarse en un pequeñísimo y barato sistema microprocesador. De esta manera, la lógica borrosa permite incorporar de una manera relativamente sencilla y directa el conocimiento de un experto (o, simplemente, conocimientos intuitivos) en un determinado campo, y aplicarlas ejecutando un tipo de razonamiento aproximado a partir de la información imprecisa que suministra el entorno.

Este tipo de inferencia es atractivo en el problema abordado en este trabajo, pues los términos ligísticos de energía y sonoridad pueden tener estados altos, medios y bajos (por ejemplo) y definirse funciones de pertenencia que puedan combinarse en reglas diferentes para las posibles respuestas (la sílaba fue sorda, sonora, nasal, o fricativa).

Se decidió emplear un sistema de tipo Sugeno (salida en forma de suma ponderada o promediada del resultado de las reglas) y no Mamdani (salida como combinación de funciones de pertenencia de las reglas) por la posibilidad del primero de ser autoajustado mediante entrenamiento.

#### 3.3.1.1 Funciones de Pertenencia Gaussianas

En nuestro trabajo emplearemos como función de pertenencia una campana gaussiana de dos dimensiones, quedando según (3) la función de pertenencia.

$$F_{(IES, ISS)} = \frac{1}{2\pi\sigma_{IES}\sigma_{ISS}} e^{-\left[\frac{(IES-\mu_{IES})^2}{2\sigma_{IES}^2} + \frac{(ISS-\mu_{ISS})^2}{2\sigma_{ISS}^2}\right]} \quad (3)$$

Después de definida la función de transferencia se calcula para cada una de las consonantes sustituyendo en ella los valores de  $\mu_{IES}$ ,  $\sigma_{IES}$ ,  $\mu_{ISS}$ ,  $\sigma_{ISS}$ . Estos valores son calculados como el promedio y la variación estándar respectivamente de los valores de IES e ISS de todas las sílabas de los sujetos sanos. En la Tabla 2 valores para los diferentes grupos son los siguientes.

Tabla 2: Resumen de los valores de media y varianza para cada uno de los indicadores

	$\mu_{IES}$	$\sigma_{IES}$	$\mu_{ISS}$	$\sigma_{ISS}$
PTK	0.11	0.09	0.45	0.16
BDG	0.16	0.10	0.74	0.12
FSJ	0.29	0.17	0.45	0.22
MN	0.59	0.11	0.97	0.032

## Capítulo 3 Materiales y Métodos

### 3.3.1.2 Funciones de Pertenencia Trapezoidales

Las funciones de transferencia trapezoidales se configuran de manera que las fronteras de los trapecios sigan una distribución semejante a la supuesta en la Tabla 1 de la sección 1.3.3. Posteriormente se optimiza la ubicación de los umbrales manera que se maximice la TCC (ver definición de esta en la sección 2.4.2). La configuración final se muestra en el epígrafe 3.3.1.

### 3.3.2 Método de clasificación a través del análisis cuadrático del discriminante

El análisis del discriminante (DA) es un método estadístico utilizado en reconocimiento de patrones y aprendizaje automatizado (*machine learning*) con el objetivo de encontrar una combinación lineal de rasgos que caracterizan o separen dos o más clases de objetos. La combinación resultante puede ser utilizada como un clasificador lineal (LDA), cuadrático (QDA) o cualquier otro dependiendo del tipo de superficie que va a separar las clases. En el caso de solo dos clases LDA enfoca el problema asumiendo que las funciones de densidad probabilísticas condicionales están normalmente distribuidas con parámetro de media y covarianza. Bajo esta suposición la solución óptima de Bayes es predecir que hay puntos que pertenecen a la segunda clase si la razón entre el logaritmo de las pertenencias se encuentra por debajo de cierto umbral.

El LDA está cercanamente relacionado a los análisis de varianza (ANOVA) los cuales también a su vez tratan de expresar una variable dependiente como combinación lineal de otros rasgos y características. La principal diferencia consiste en que en estos métodos la variable dependiente es una entidad numérica, mientras que para el LDA es una variable categórica, como por ejemplo un nombre de una clase. Otra diferencia fundamental entre LDA y los demás métodos se manifiesta en que LDA no es una técnica independiente, se necesita para realizarla hacer una separación entre las variables dependientes y las independientes. LDA funciona cuando las mediciones de las variables independientes son continuas. Cuando se tienen variables independientes categóricas, se debe utilizar el análisis de correspondencia del discriminante.

En el presente trabajo se utiliza la función *classify* implementada en el programa de ingeniería Matlab. En la utilización de esta función se pueden definir el parámetro del tipo de análisis multivariado a utilizar. El tipo *lineal* donde se asigna una densidad multivariada de tipo normal a cada grupo a clasificar. El tipo *cuadrático* que establece una densidad multivariada de tipo normal con estimados de covarianzas estratificados para cada grupo. Por último se encuentra el tipo *mahalanobis* que utiliza distancias Mahalanobis con estimados de

## Capítulo 3 Materiales y Métodos

covarianzas estratificados. Se realizó la clasificación teniendo en cuenta todos los posibles análisis mencionados y se compararon los resultados de clasificación arribándose a la conclusión de que se usaría el QDA.

### 3.3.3 Métodos de clasificación a través de un clasificador naive Bayes

El clasificador bayesiano combina el modelo de probabilidad Bayesiano con una regla de decisión.

Usando el Teorema de Bayes se obtiene:

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}. \quad (4)$$

Donde:

C: Clases

$F_1, \dots, F_n$ : Rasgos.

Además se asume que dada una variable de clase la presencia o ausencia de un rasgo particular de una clase no está relacionado con a la presencia o ausencia de cualquier otro rasgo.

Por último la regla de decisión más común es escoger la hipótesis que es más probable, esto se conoce como el máximo a posteriori (MAP).

Dependiendo de la naturaleza del modelo probabilístico los clasificadores naive Bayes pueden ser entrenados de forma eficiente bajo condiciones de aprendizaje supervisado. En muchas aplicaciones prácticas, la estimación de parámetros en el modelo bayesiano utiliza el método de máxima similitud; esto implica que el modelo bayesiano puede ser utilizado incluso cuando no se utiliza específicamente probabilidades bayesianas.

A pesar de poseer principios de funcionamiento sencillos el clasificador bayesiano funciona bien enfrentando muchos problemas del mundo real.

Una ventaja que posee este tipo de clasificador es que solo necesita una pequeña cantidad de datos de entrenamiento para estimar los parámetros (media y varianza de las variables) necesarios para la clasificación. Ya que se asume la independencia de las variables solo las varianzas de cada variable para cada clase necesitan ser determinadas y no la matriz de covarianza completa. El clasificador bayesiano utilizado se encuentra implementado en el programa MATLAB.

## Capítulo 3 Materiales y Métodos

### 3.3.4 Métodos de clasificación a través K Vecinos más cercanos (KNN)

En reconocimiento de patrones el algoritmo de  $k$  vecinos más cercanos ( $knn$ ) es un método de clasificación basado en la cercanía de conjuntos de entrenamiento en el espacio de rasgos. Este algoritmo es uno de los más sencillos dentro del campo de aprendizaje automatizado (*machine learning*) un objeto es clasificado según el tipo de clase preponderante de los objetos que lo rodean.

El mismo método puede ser utilizado para regresión, asignando como valor de clasificación el promedio de los valores de clasificación de los vecinos más cercanos. También se puede ponderar la contribución de forma que los más cercanos contribuyan más al promedio que los más lejanos. Una forma usual de hacer esto es asignándole a cada vecino un peso de  $1/\text{distancia}$  hasta el vecino.

Los ejemplos de entrenamiento son vectores en un espacio de rasgos multidimensional cada uno con una clase asignada. La fase de entrenamiento del algoritmo consiste en almacenar el vector de rasgos y las etiquetas de clases de las muestras de entrenamiento.

En la fase de clasificación, el valor  $k$  es definido por el usuario y un vector no definido (punto de prueba) es clasificado asignándole la clase que es más frecuente entre los  $k$  muestras de entrenamiento cercanas a ese punto de prueba.

Usualmente la distancia Euclidiana es utilizada como métrica, sin embargo esto solo es aplicable cuando las variables son continuas. En otros casos como son las clasificaciones de textos se utiliza la distancia de *Hamming*. Por otro lado la eficiencia de este clasificador puede ser aumentada considerablemente si el algoritmo de calcular las distancias es mejorado por ejemplo utilizando Vecino más cercano de margen largo (LMNN) o análisis de componentes del vecindario.

Un problema habitual presentado por este clasificador es que el algoritmo de votar por mayoría favorece a las clases que tengan más elementos en el conjunto de entrenamiento. Una forma de solapar este problema es pesando la clasificación teniendo en cuenta la distancia desde el punto de prueba hasta cada uno de sus  $k$  vecinos. En el caso de este trabajo se utilizó la implementación del algoritmo incluida en el programa MATLAB.

### 3.3.5 Método de clasificación a través de árboles de decisión (árboles de clasificación)

El aprendizaje a través de árboles de decisión utiliza este modelo como modelo predictivo para relacionar observaciones de un elemento con valores del elemento. Estos modelos cuando son aplicados a problemas de clasificación se les suele denominar arboles de clasificación o arboles de regresión. La diferencia entre ellos se encuentra en que los árboles

## Capítulo 3 Materiales y Métodos

de clasificación brindan como resultado un valor de clase, mientras que los de regresión brindan un número. En cualquiera de estas estructuras las hojas del árbol representan clases mientras que las ramas son conjunciones de rasgos que resultan en esas clases.

### 3.4 Descripción de las pruebas

A continuación se describirá la forma en que se prepararon los datos para la realización de las pruebas así como las medidas que se van a utilizar para comprobar los resultados. Es importante destacar que los clasificadores fueron entrenados y evaluados utilizando el conjunto total de pares (*IES*, *ISS*), o sea, el conjunto de control y entrenamiento son iguales y abarcan todas las muestras.

#### 3.4.1 Test de normalidad

El primer paso de procesamiento de los datos fue realizar un test de normalidad a las variables independientes *IES* e *ISS*. En el caso de la normalidad para cada variable por separado se realizó con el test de Kolmogorov-Smirnov. Para el caso de la prueba de normalidad asumiendo 2 dimensiones compuestas por las variables se utilizó el test de Mardia. Esta prueba está basada en las extensiones multivariadas de las medidas de (sesgo) skewness y kurtosis. Por ejemplo se tiene  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  de vectores  $p$ -dimensionales.

$$A = \frac{1}{6n} \sum_{i=1}^n \sum_{j=1}^n \left[ (\mathbf{x}_i - \bar{\mathbf{x}})' \hat{\Sigma}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}) \right]^3 \quad (5)$$

$$B = \frac{\sqrt{n}}{\sqrt{8p(p+2)}} \left[ \frac{1}{n} \sum_{i=1}^n \left[ (\mathbf{x}_i - \bar{\mathbf{x}})' \hat{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \right]^2 - p(p+2) \right] \quad (6)$$

Bajo la hipótesis nula de normalidad multivariada, la estadística  $A$  (5) se tendrá una distribución chi-cuadrado con  $1 p(p+1)(p+2)$  grados de libertad, mientras que  $B$  (6) será aproximadamente una distribución normal  $N(0,1)$ .

La kurtosis de Mardia es sesgada y converge muy lentamente hacia el límite de la distribución normal. Para muestras medianas ( $50 < n < 400$ ), los parámetros de las distribuciones asintóticas de la estadística de kurtosis son modificadas. Para muestras pequeñas ( $n < 50$ ) son utilizados valores críticos empíricos.

#### 3.4.2 Medidas para la comprobación de la efectividad de los métodos en la clasificación

La sensibilidad y la especificidad son medidas estadísticas del rendimiento de una prueba de clasificación lineal o función de clasificación. La sensibilidad mide la proporción de

### Capítulo 3 Materiales y Métodos

positivos que realmente fueron identificados como tal. Mientras que la especificidad mide la proporción de negativos que fueron identificados como tal. Estas dos medidas están cercanamente relacionadas con los errores de Tipo 1 y Tipo 2 definidos en la Tabla 3. En teoría un clasificador perfecto tendría 100% de sensibilidad y el mismo valor de especificidad sin embargo teóricamente todos deben tener un margen de error conocido como la tasa de error de Bayes.

Para cualquier prueba existe usualmente una relación de compromiso entre las medidas.

Un ejemplo sencillo que demuestra estos conceptos puede concebirse con el análisis de personas para determinar si poseen una enfermedad determinada. El resultado de la prueba puede ser positivo si la persona tiene la enfermedad y se diagnostica enfermo o negativo si la persona está sana y se le diagnostica sana. Sin embargo el resultado de la prueba puede que no coincida con el estatus real del paciente, es decir:

Verdadero positivo (TP): Personas enfermas diagnosticadas como tal. Es un acierto.

Falso positivo (FP): Personas sanas identificadas como enfermas. Es una falsa alarma.

Verdadero negativo (TN): Personas sanas identificadas como tal. Es un descarte.

Falso negativo (FN): Personas enfermas identificadas como sanas. Es un fallo.

La sensibilidad se relaciona con la habilidad de una prueba de identificar resultados que son positivos según (7):

$$sensibilidad = \frac{totalTP}{totalTP + totalFN} \quad (7)$$

En el ejemplo de la enfermedad, si una prueba posee alta sensibilidad entonces un resultado negativo equivaldría a la ausencia de la enfermedad. Es decir, una sensibilidad del 100% significa que una prueba reconoce todos los positivos. Una prueba con alta sensibilidad posee una baja tasa de error de Tipo 2.

Por otra parte la especificidad se relaciona con la habilidad de identificar resultados negativos según (8):

$$especificidad = \frac{totalTN}{totalTN + totalFP} \quad (8)$$

Una prueba con alto grado de especificidad tiene un bajo nivel de error Tipo 1. Tabla 3 se muestra como se consideran los tipos de errores.

Tabla 3: Errores posibles en la clasificación en la matriz de confusión.

	Condición Positiva	Condición Negativa
--	--------------------	--------------------

### Capítulo 3 Materiales y Métodos

Resultado Positivo de la prueba	TP	FP (Error Tipo 1)
Resultado Negativo de la prueba	FN (Error Tipo 2)	TN

Por último introducimos el concepto de tasa de clasificación correcta (9).

$$TC = \frac{totalTP + totalTN}{totalTN + totalFP + totalTP + totalFN} \quad (9)$$

La tasa de clasificación correcta o exactitud describe la capacidad de un clasificador de detectar correctamente, los positivos como positivos y los negativos como tal.

Existen otras medidas que pueden ser extraídas de la matriz de confusión (Media Armónica F, por ejemplo) pero solo utilizaremos las mencionadas anteriormente.

#### 3.5 Conclusiones Parciales del Capítulo

Se grabaron las pronunciaciones de las 11 consonantes por parte de 20 sujetos sin historial de patologías del habla. Se seleccionó un conjunto de clasificadores de los existentes en la versión R2010a de MATLAB, y las medidas de efectividad a emplear en la evaluación de la separación que los mismos logren entre las consonantes involucradas.

### Capítulo 4. Resultados y Discusión

En este capítulo se mostraran los resultados de los clasificadores. Además se realizará una comparación entre ellos y se arribará a conclusiones sobre la comprobación de la hipótesis de investigación.

#### 4.1 Resultados de la detección de sílabas y de los test de normalidad

El detector de sílabas utilizado fue capaz de reconocer 3098 sílabas en 220 señales. No se midió la efectividad de este ya que no constituía objetivo de este trabajo solo se dio por sentado su efectividad basándose en los resultados reportados. De estas sílabas 1096 son oclusivas sordas, 823 oclusivas sonoras, 649 fricativas y 530 nasales. Se comprobó a través de la media y la desviación estándar de estos valores que no existe desbalance entre las poblaciones de estas consonantes, ya que el valor de  $\sigma/\mu$  en porciento es de 31.7.

Por otra parte los resultados de los test de normalidad (*Kolmogorov-Smirnov* y *Mardia Test*) de una y dos dimensiones fueron negativos para una significación de 95%.

#### 4.2 Distribuciones de las consonantes en el plano de Energía vs Sonoridad

Los valores individuales de *IES* e *ISS* para cada sílaba correspondientes a las señales de sujetos sanos fueron promediados, y se obtuvieron además sus desviaciones estándar. En la Fig. 4 se muestran las aglomeraciones de puntos coordinados en el plano de *EvsS*, así se puede apreciar la complejidad del problema, se representan los puntos correspondientes a las consonantes “PTK” como cruces, las consonantes “BDG” son rombos, las consonantes “FSJ” se representan como cuadrados y por último las “MN” se representan como círculos. En la Fig. 5 se muestran elipses que representan las regiones correspondientes a cada grupo de consonantes en un plano (*IES*, *ISS*). Cada elipse (región) está determinada por su centro (valores de *IES* e *ISS* promedios del grupo de consonantes correspondiente), su radio horizontal (desviación estándar de *IES* para el grupo en cuestión) y su radio vertical (desviación estándar de *ISS* en el grupo). Puede apreciarse en la figura que las regiones obtenidas describen adecuadamente las características esperadas para los distintos grupos de consonantes, como la alta sonoridad de nasales y sonoras, o la menor energía de las sordas. En cuanto a la capacidad de los índices para separar las regiones, se observa que la mayor dificultad surge en el caso de las explosivas sordas de las fricativas. Estas últimas muestran una dispersión de los posibles valores de *IES* muy elevada que produce un solapamiento notable de las regiones. El hecho de que el *ISS* de ambos grupos sea semejante, por demás

## Capítulo 4 Resultados y Discusión

esperado ya que ambos grupos son de consonantes sordas, no contribuye a eliminar este efecto negativo.

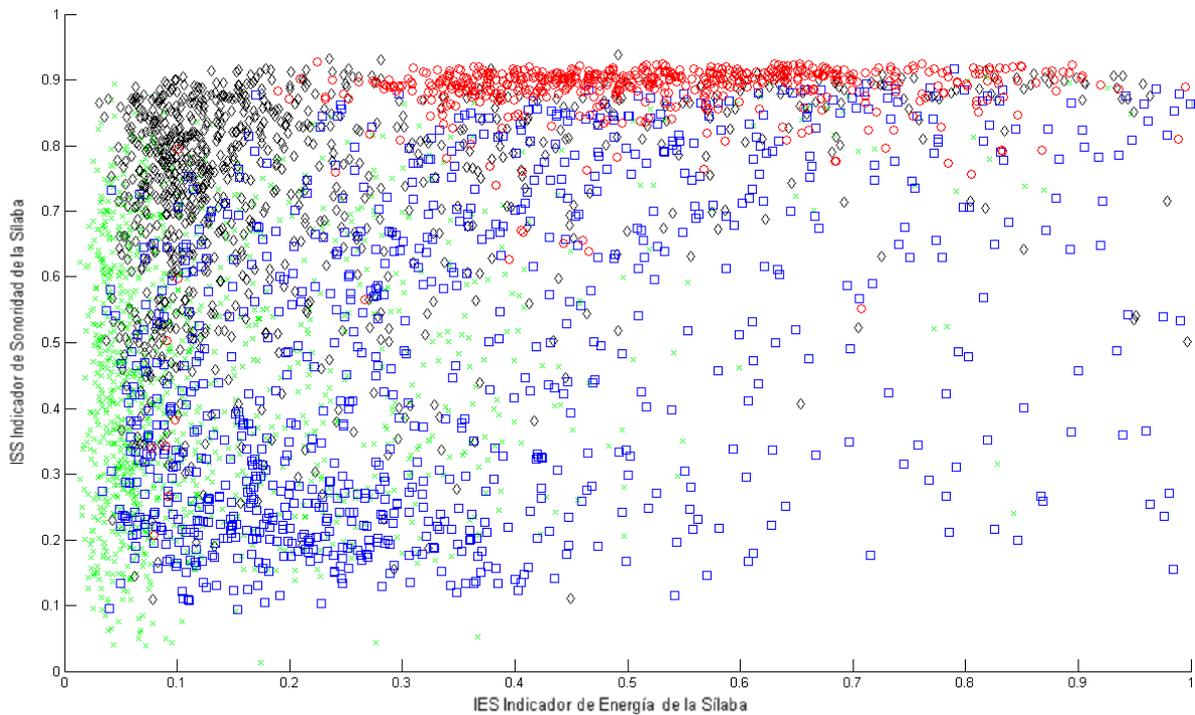


Fig. 4. Distribución de los pares coordenados en el plano de EvsS.

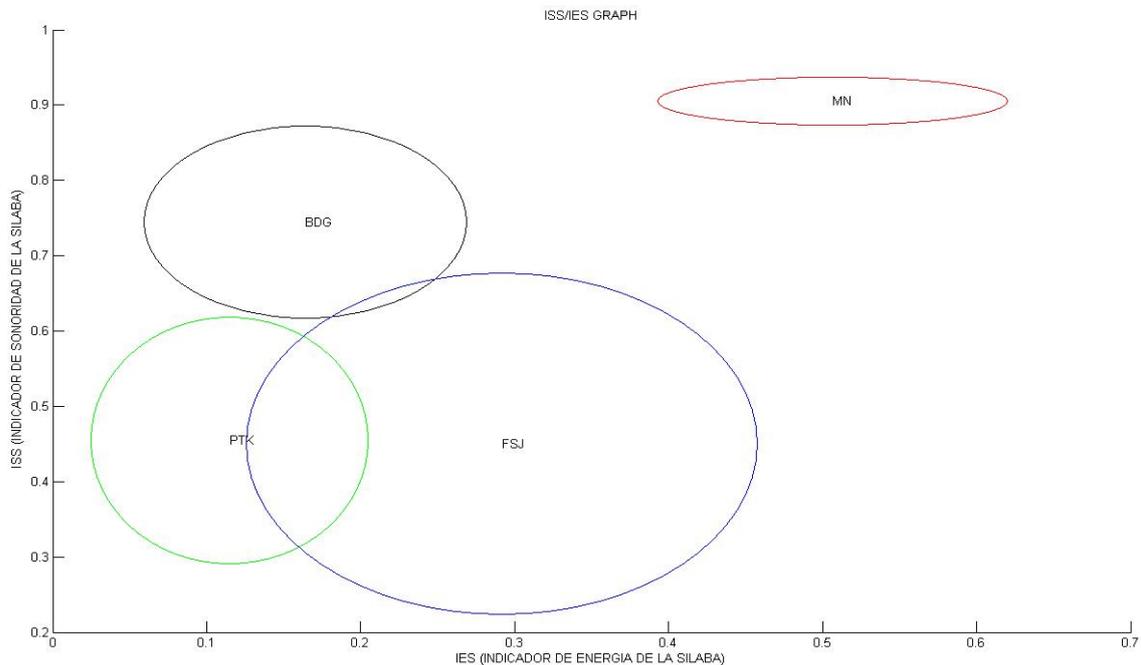


Fig. 5. Regiones de agrupación de las consonantes.

Teniendo en cuenta que se demostró que las variables de entrada no poseen una distribución de tipo normal, la Fig. 5 aunque es representativa del resultado alcanzado no es estadísticamente correcto. Debido a esto se estimó como medida de tendencia central la

## Capítulo 4 Resultados y Discusión

mediana de las variables y se calculó el valor del 25 y 75 percentil. A partir de estos valores se elaboró un gráfico de caja (*Boxplot*), el cual se muestra en la Fig.6.

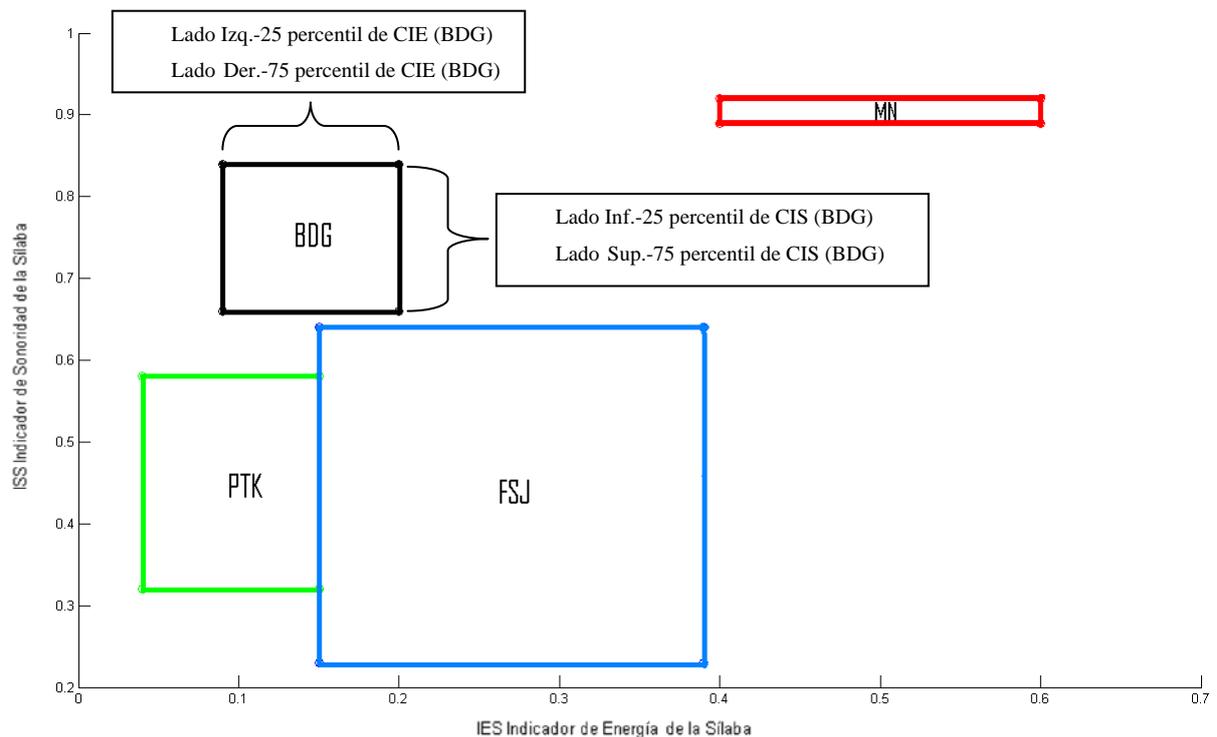


Fig. 6. Distribución de las regiones a través de la mediana y los valores correspondientes al 25 y 75 percentil.

### 4.3 Resultados de los clasificadores

#### 4.3.1 Sistema de Inferencia Borroso de tipo Sugeno con función de transferencia trapezoidal (SIST)

El sistema no pudo ser entrenado para que se autoajustara por limitaciones del algoritmo de entrenamiento (función ANFIS del MATLAB):

1. No pueden entrenarse sistemas de más de una salida
2. No puede emplearse un criterio diferente al de error de clasificación.

Por este motivo se fijaron los umbrales de las funciones trapezoidales de manera manual, y se ajustaron efectuando corrimientos de 0.01 de manera que se redujera el error de clasificación (empleando ANFIS). Cuando el error de clasificación no se redujo más se midió la efectividad del clasificador así configurado.

## Capítulo 4 Resultados y Discusión

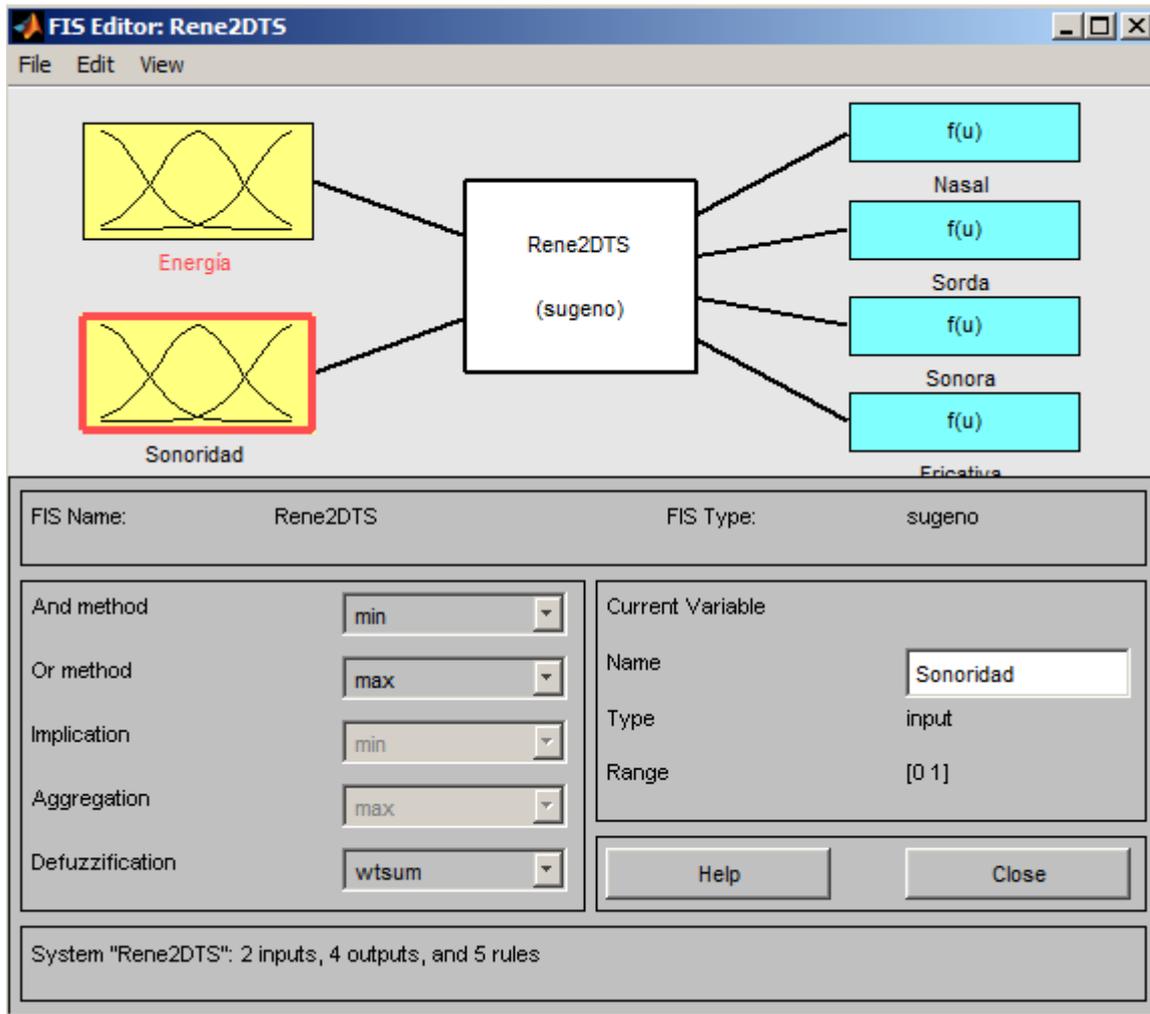


Fig. 7. Sistema de inferencia desarrollado.

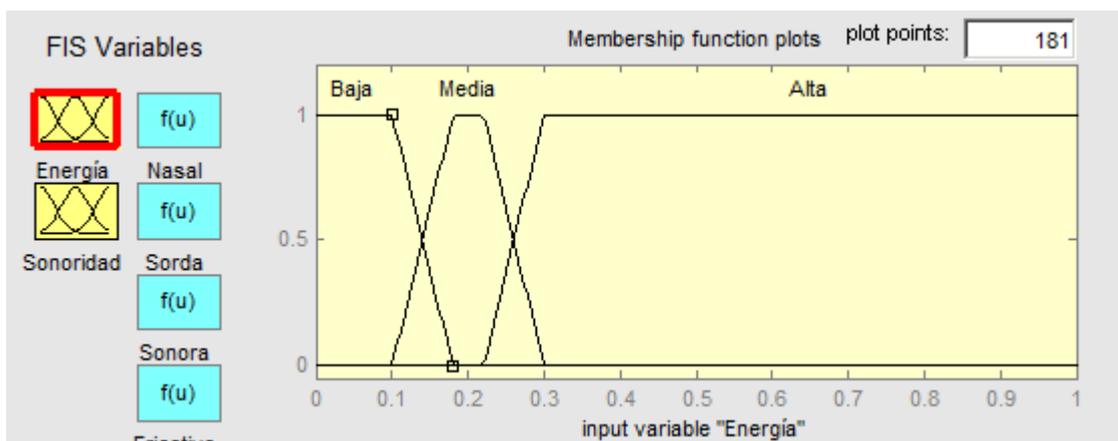


Fig.8. Funciones de pertenencia para la variable lingüística Energía luego del ajuste de los umbrales.

## Capítulo 4 Resultados y Discusión

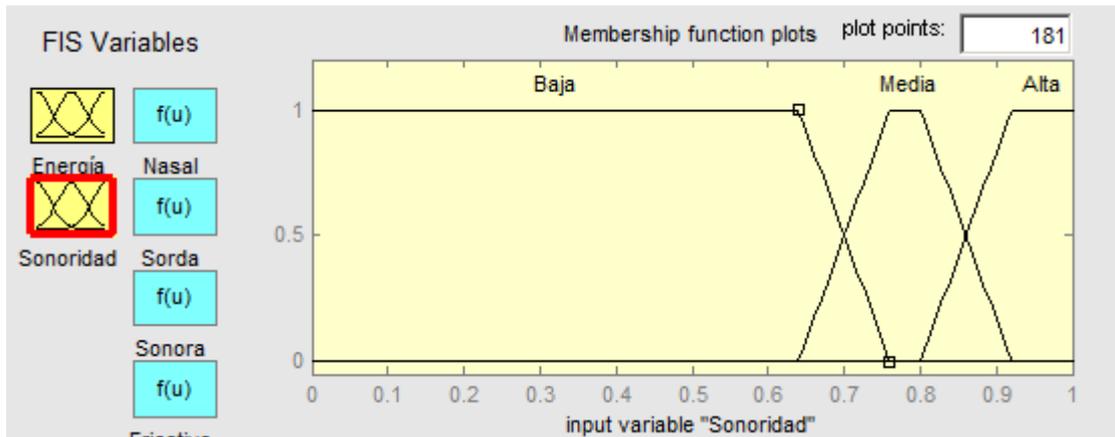


Fig.9. Funciones de pertenencia para la variable lingüística Sonoridad luego del ajuste de los umbrales.

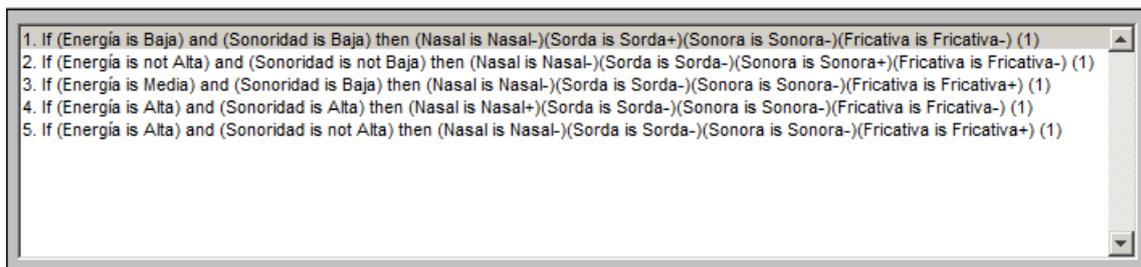


Fig. 10. Reglas de Inferencia establecidas

La salida del clasificador fue considerada indicadora de la consonante cuya salida obtuviese el mayor valor. La región de clasificación obtenida se muestra y los resultados (matriz de confusión) se muestran a continuación.

Los resultados obtenidos fueron:

- Tasa de Clasificación = 64%

## Capítulo 4 Resultados y Discusión

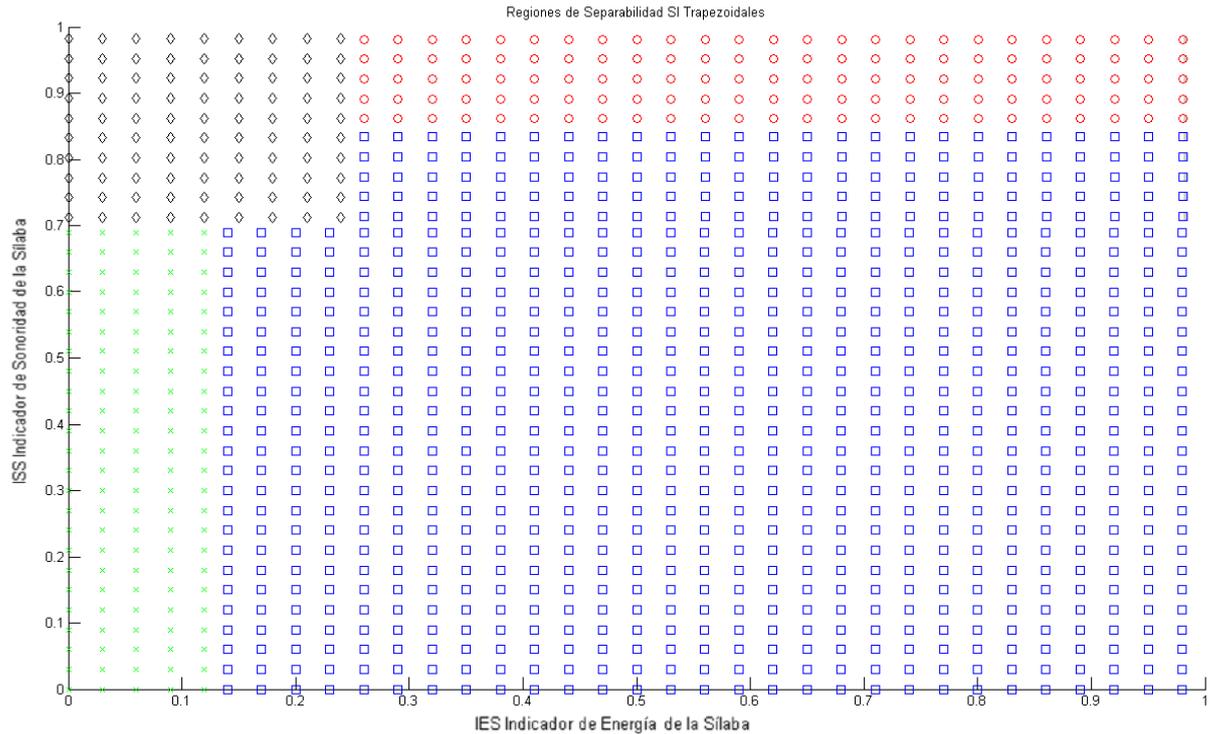


Fig. 11. Regiones de agrupación según el clasificador SIST

A continuación se observa en la Tabla 4 la matriz de confusión resultante y en la Tabla 5, el resultado de las medidas calculadas.

Tabla 4. Matriz de confusión de SIST

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	672	109	302	13
	BDG	182	403	169	69
	FSJ	137	11	472	29
	MN	7	12	71	440

Tabla 5. Valores de sensibilidad y especificidad de SIST

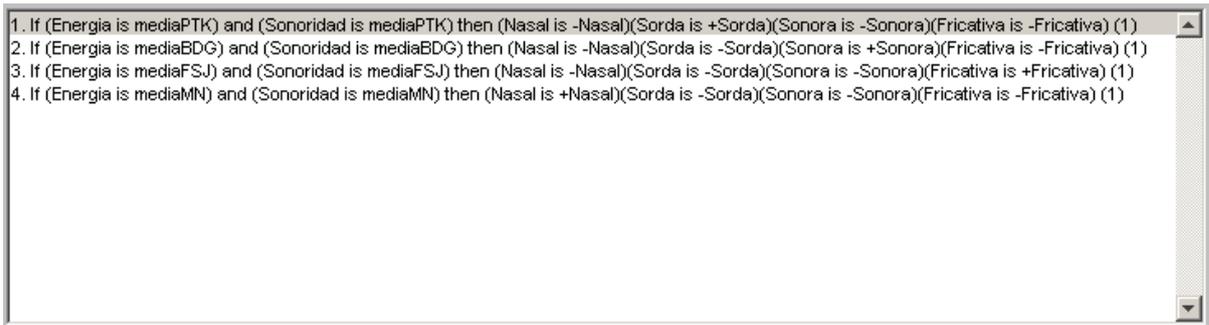
	PTK	BDG	FSJ	MN
Sensibilidad	61.2%	48.7%	72.6%	83.4%
Especificidad	83.5%	94.3%	78.1%	95.5%

Es importante comentar el valor señalado en rojo en la Tabla 4. Dichos valores evidencian un gran número de falsos positivos de las consonantes oclusivas sordas (PTK) siendo clasificadas como fricativas (FSJ). Esta gran cantidad de falsos positivos afectan seriamente la TCC. La incidencia negativa de estos resultados se puede constatar en los valores de sensibilidad de las PTK.

## Capítulo 4 Resultados y Discusión

### 4.3.2 Sistema de Inferencia Borroso de tipo Sugeno con función de transferencia gaussiana (SISG)

La configuración de las funciones de pertenencia Gaussianas se extrae de las medias y desviaciones estándar de los parámetros de energía y sonoridad mostrados en la Tabla 2 de la sección 2.3.1.1. La elaboración del sistema se realizó de la misma forma que el anterior solo se sustituyó el tipo de funciones de pertenencia cambiando a Gaussianas y se redefinieron las reglas. Estas se muestran en la Fig. 12.



```
1. If (Energia is mediaPTK) and (Sonoridad is mediaPTK) then (Nasal is -Nasal)(Sorda is +Sorda)(Sonora is -Sonora)(Fricativa is -Fricativa) (1)
2. If (Energia is mediaBDG) and (Sonoridad is mediaBDG) then (Nasal is -Nasal)(Sorda is -Sorda)(Sonora is +Sonora)(Fricativa is -Fricativa) (1)
3. If (Energia is mediaFSJ) and (Sonoridad is mediaFSJ) then (Nasal is -Nasal)(Sorda is -Sorda)(Sonora is -Sonora)(Fricativa is +Fricativa) (1)
4. If (Energia is mediaMN) and (Sonoridad is mediaMN) then (Nasal is +Nasal)(Sorda is -Sorda)(Sonora is -Sonora)(Fricativa is -Fricativa) (1)
```

Fig. 12. Reglas de Inferencia establecidas

La salida del clasificador fue considerada indicadora de la consonante cuya salida obtuviese el mayor valor. La región de clasificación obtenida se muestra y los resultados (matriz de confusión) se muestran a continuación.

Los resultados obtenidos fueron:

- Tasa de Clasificación = 62%

En la Fig. 13 se muestra las regiones de agrupación definidas por este clasificador.

## Capítulo 4 Resultados y Discusión

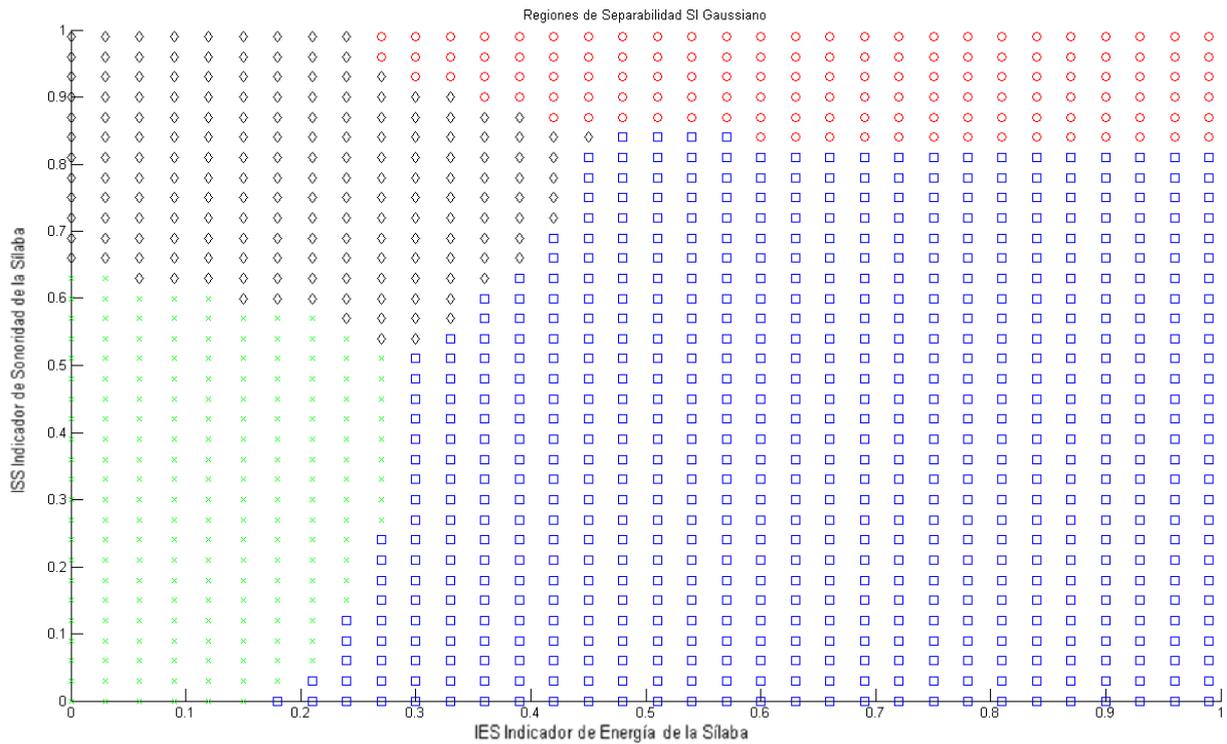


Fig. 13. Regiones de agrupación según el clasificador SISG

En la Tabla 6 y 7 se visualiza la matriz de confusión y los valores de sensibilidad y especificidad.

Tabla 6. Matriz de confusión de SISG

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	729	246	105	16
	BDG	155	558	45	65
	FSJ	275	98	238	38
	MN	7	90	37	396

Tabla 7. Valores de sensibilidad y especificidad de SISG

	PTK	BDG	FSJ	MN
Sensibilidad	66.5%	67.8%	36.6%	74.7%
Especificidad	78.1%	80.9%	92.3%	95.3%

En este caso existen gran cantidad de falsos positivos de las consonantes fricativas (FSJ) siendo clasificadas como oclusivas sordas (PTK).

### 4.3.3 Análisis cuadrático del Discriminante (QDA)

A continuación se muestran los resultados del análisis de discriminante cuadrático, como se había mencionado antes se comprobó la prueba utilizando una función lineal es decir un LDA, pero los resultados no fueron satisfactorios. El análisis de la distribución de puntos en

## Capítulo 4 Resultados y Discusión

el plano  $EvsS$  (ver Fig.4) visualiza muy bien este dilema, sería muy difícil con una función lineal obtener una frontera de separación para las aglomeraciones de estos puntos. Además es apropiado señalar que este clasificador requiere que la distribución de los rasgos de entrada sea de tipo normal, condición NO cumplida por  $IES$  e  $ISS$  tal y como se había comentado previamente. En la Fig. 14 se muestra las regiones de agrupación determinadas por QDA.

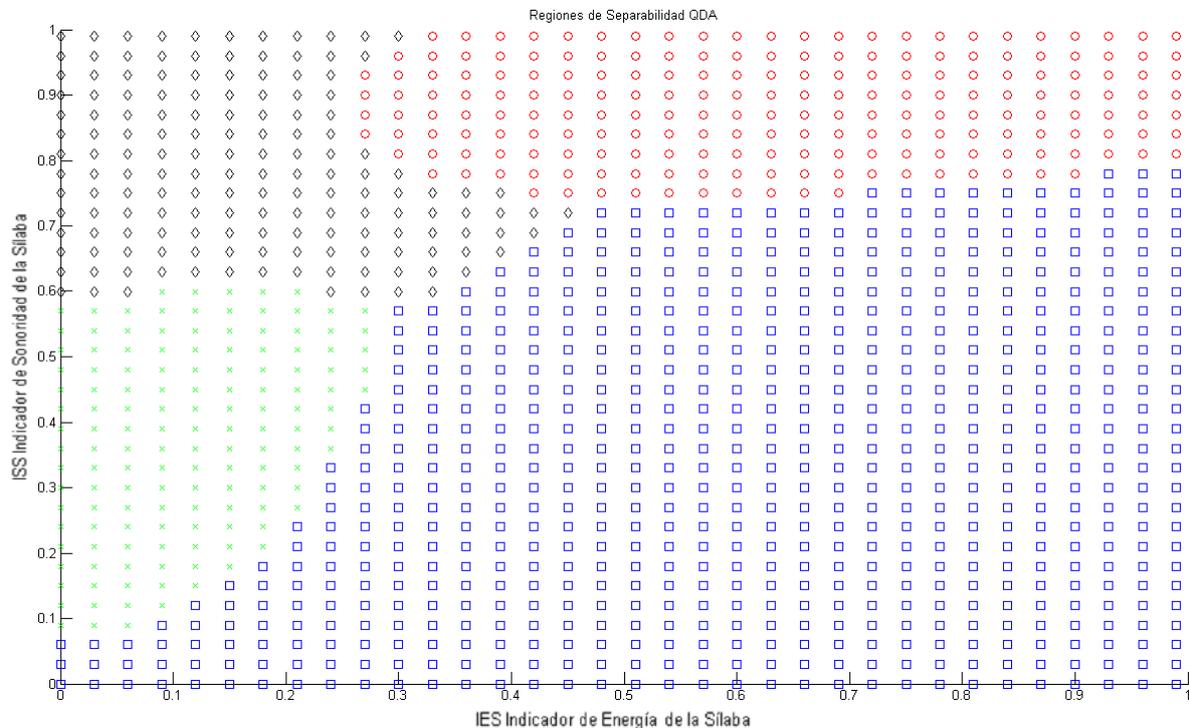


Fig. 14. Regiones de agrupación según el clasificador QDA

Por otra parte en las Tablas 8 y 9 se observan los resultados de la matriz de confusión y de los valores de sensibilidad y especificidad respectivamente.

Tabla 8 Matriz de confusión de QDA

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	608	278	172	38
	BDG	135	538	36	114
	FSJ	169	60	281	139
	MN	7	58	4	461

Tabla 9 Valores de sensibilidad y especificidad de QDA

	PTK	BDG	FSJ	MN
Sensibilidad	55.5 %	65.4 %	43.3 %	86.9 %
Especificidad	84.4 %	82.6 %	91.3 %	88.6 %

## Capítulo 4 Resultados y Discusión

La tasa de clasificación correcta (TCC) obtenida por este clasificador fue de 61%. Se muestran los falsos positivos de las consonantes oclusivas sordas (PTK) siendo clasificadas como oclusivas sonoras (BDG) y como fricativas (FSJ).

### 4.3.4 K Vecinos más cercanos (KNN)

El clasificador KNN no requiere de la condición de normalidad de sus datos de entrada. La variabilidad en el resultado de este modelo está dada por la cantidad de vecinos  $K$  que se tomen en cuenta para el análisis, mientras más vecinos se tomen en cuenta será peor la TCC. En el caso de este trabajo se requiere que la cantidad de vecinos sea suficiente grande como para que las regiones queden definidas de forma única y contigua. En la Fig. 15 se muestra la forma en que agrupa las regiones para el caso extremo de  $K=1$ .

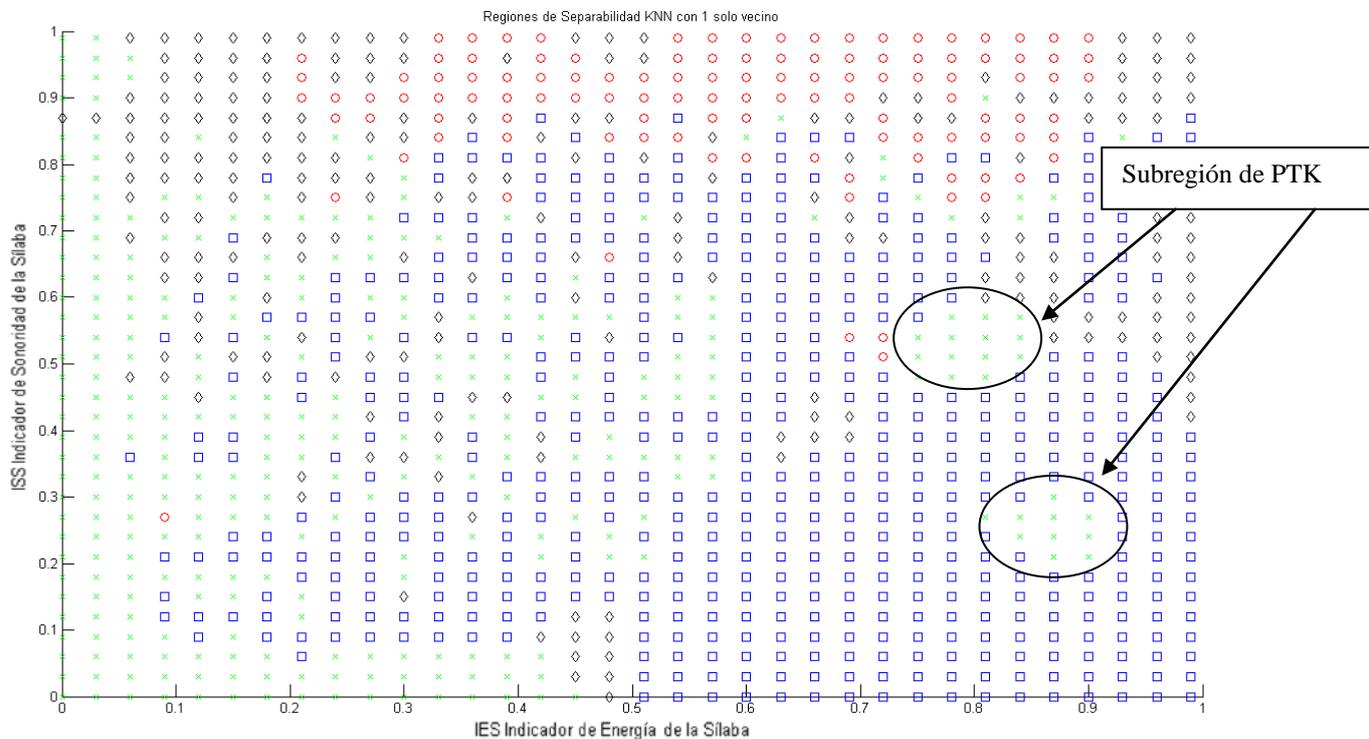


Fig.15. Regiones de agrupación según el clasificador KNN.  $K=1$ .

Como se puede apreciar los puntos se encuentran bastantes dispersos, creando subregiones y algunos de ellos se encuentran incluso aislados, tornándose en un resultado no deseable. En la Fig. 16 se muestra las regiones para  $K=1\%$  de la cantidad muestras.

## Capítulo 4 Resultados y Discusión

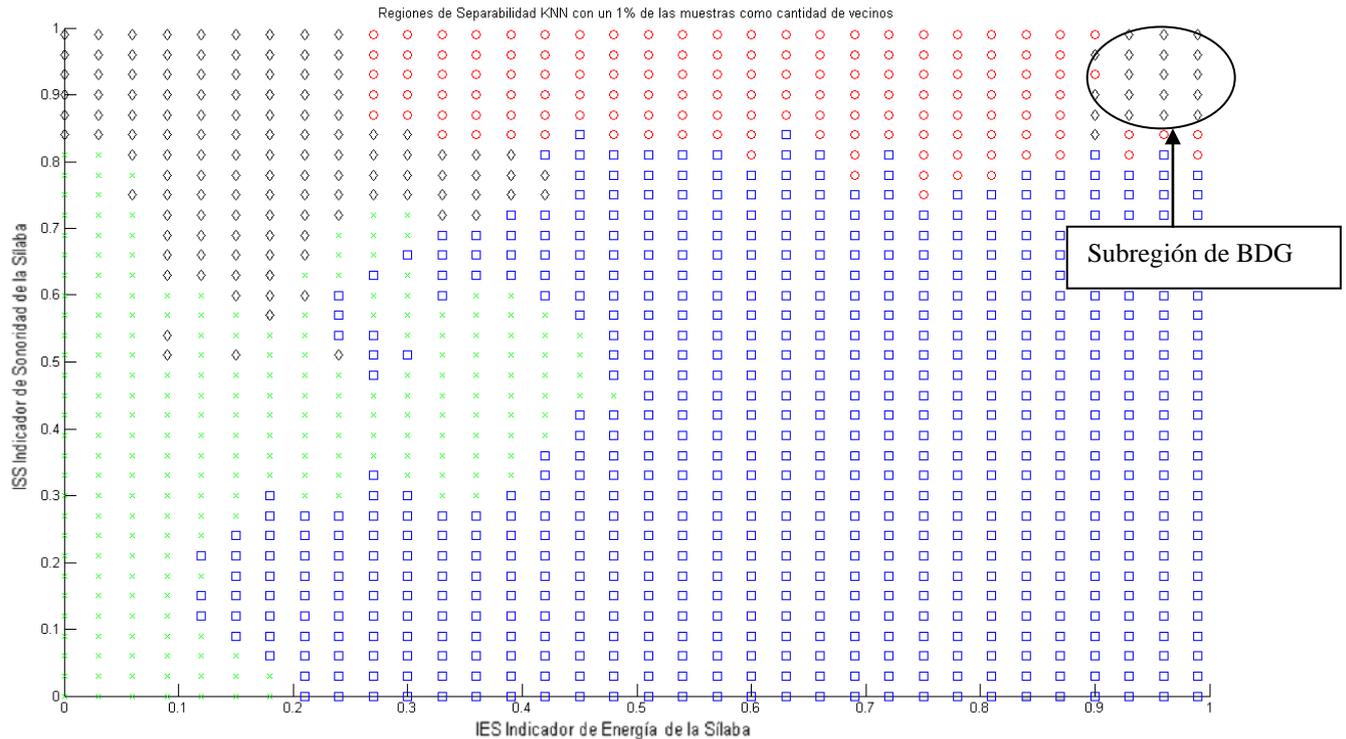


Fig.16. Regiones de agrupación según el clasificador KNN.  $K=1\%$ .

Se puede apreciar que todavía se forman “islas”, por tanto con ese valor de  $K$  no se logra el objetivo. En la Fig. 17 se elevó el valor de  $K$  al 3% de la cantidad de muestras.

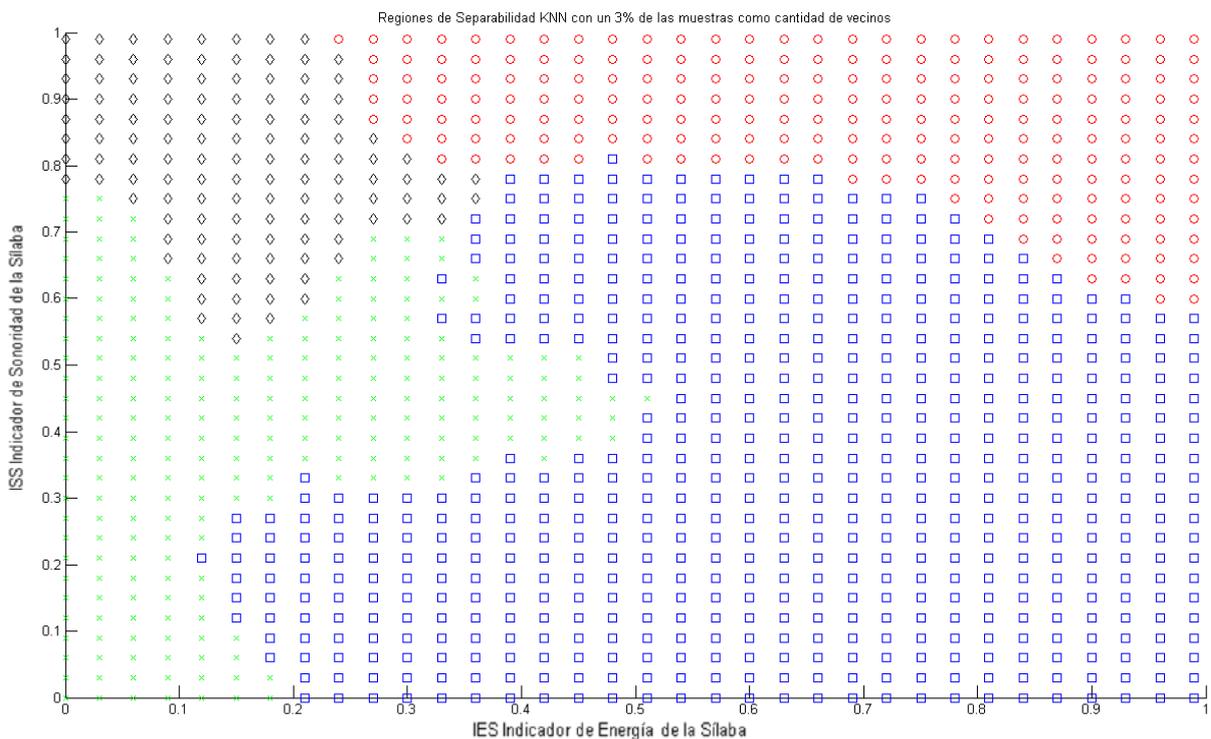


Fig.17. Regiones de agrupación según el clasificador KNN.  $K=3\%$ .

Finalmente con este valor se obtiene en gran medida el resultado deseado. No se observan subregiones en el plano y las regiones con contiguas. En la Tabla 10 y 11 se muestra la matriz de confusión y los valores de sensibilidad y especificidad para el último valor de  $K$ .

## Capítulo 4 Resultados y Discusión

Tabla 10. Matriz de confusión de KNN

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	814	138	114	30
	BDG	195	479	43	106
	FSJ	206	42	311	90
	MN	7	14	14	495

Tabla 11. Valores de sensibilidad y especificidad de KNN

	PTK	BDG	FSJ	MN
Sensibilidad	74.2 %	58.2 %	47.9 %	93.4 %
Especificidad	79.6 %	91.4 %	93.1 %	91.1 %

La TCC de este clasificador fue de 68%. Como se observa en los valores resaltados en la Tabla 10, se vuelve a presentar un gran número de falsos positivos ahora de las consonantes BDG y FSJ siendo clasificadas como PTK. Estas irregularidades disminuyen la sensibilidad de estos grupos de consonantes.

### 4.3.5 Árboles de clasificación

La aplicación de un clasificador de árboles de decisión tiene como principal características la determinación de la cantidad de niveles  $N$  a tomar en cuenta (ramas). A medida que se añaden ramas la clasificación será mejor pero se corre el riesgo de que casos aislados constituyan regiones aparte. En las Fig. 18, 19 y 20 se muestran tres casos de niveles del árbol.

## Capítulo 4 Resultados y Discusión

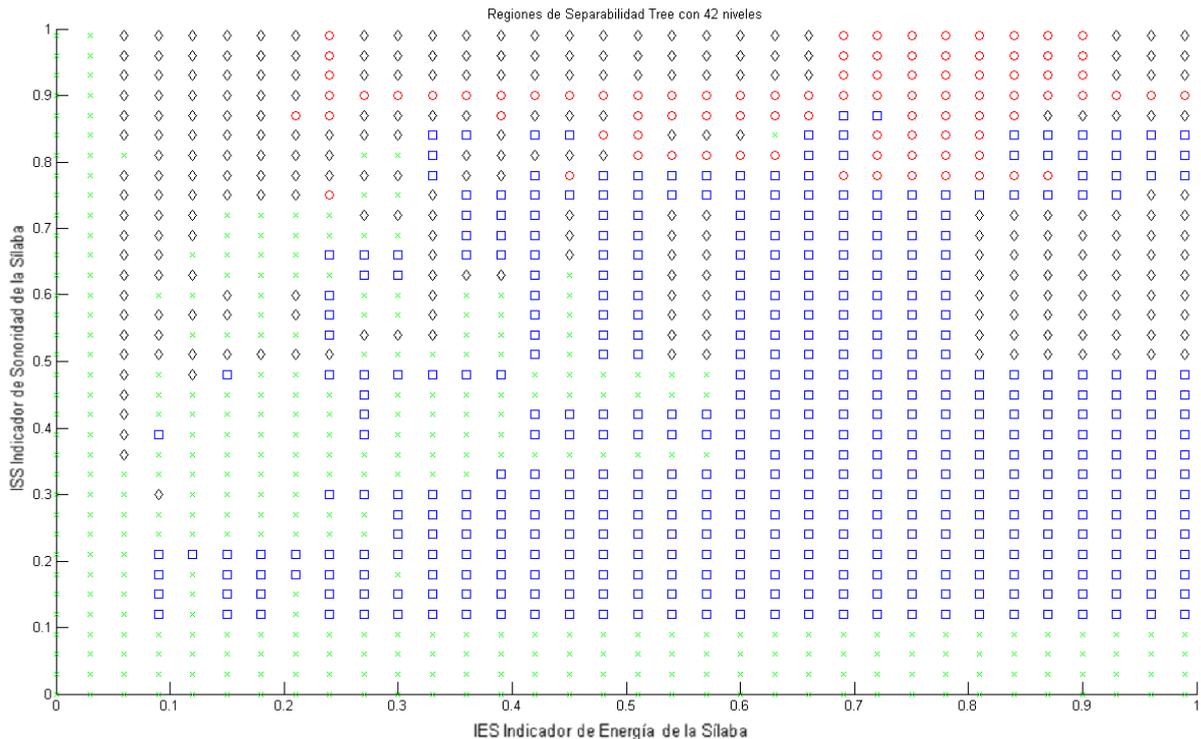


Fig.18. Regiones según el árbol de clasificación con el número de niveles  $N=42$ .

El valor de  $N=42$  es el máximo valor, con el cual el árbol resulta en una TCC de 100%. Como se observa las regiones no son únicas ni contiguas, hay presencia de subregiones. Se comenzara a disminuir el número de niveles hasta obtener la forma deseada en las regiones.

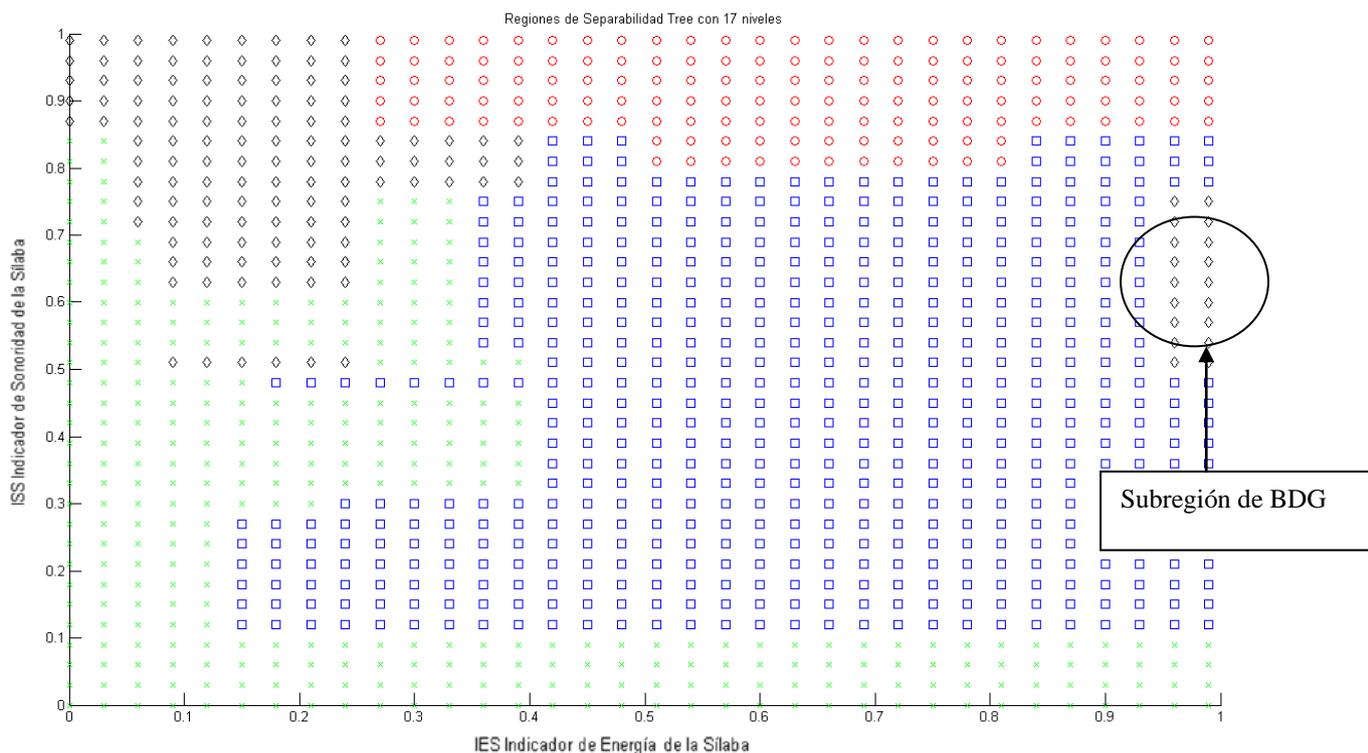


Fig.19. Regiones según el árbol de clasificación con el número de niveles  $N=17$ .

Con  $N=17$  todavía se muestran subregiones en el plano por lo que se disminuirá su valor aún más.

## Capítulo 4 Resultados y Discusión

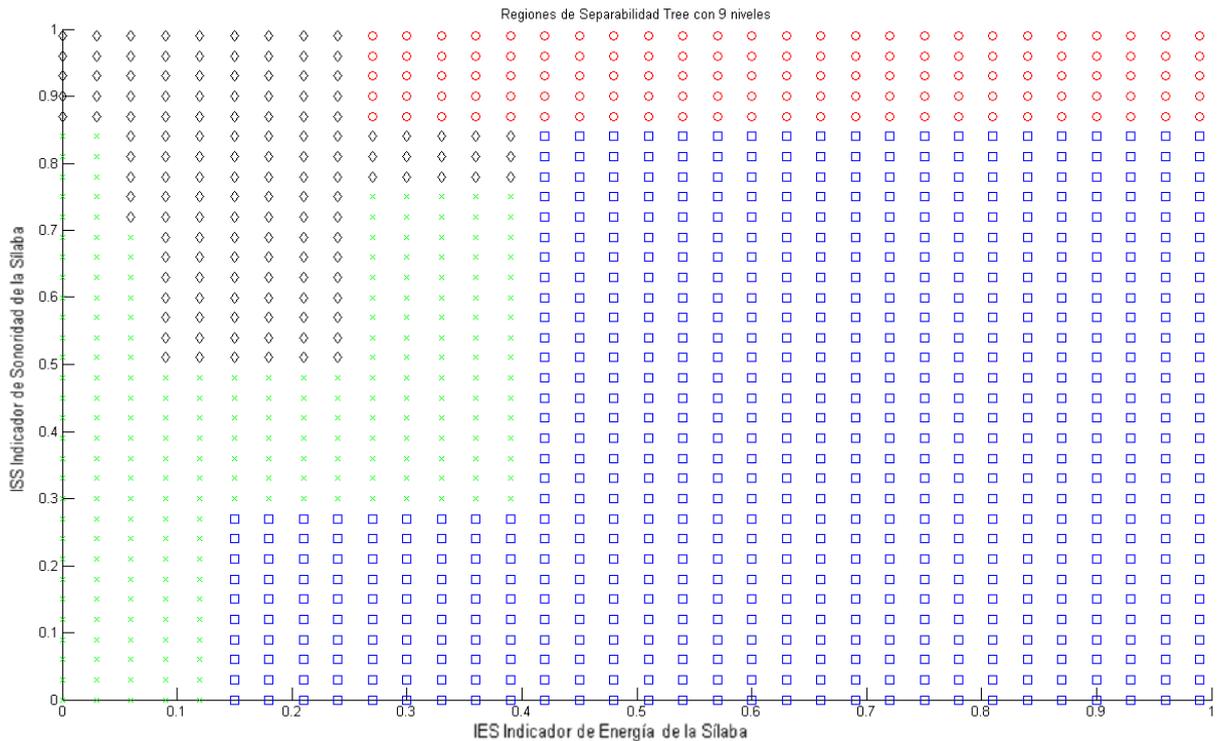


Fig. 20. Regiones según el árbol de clasificación con el número de niveles  $N=9$ .

Finalmente con este valor de  $N$  ya se obtiene el resultado deseado, las regiones son únicas y contiguas. En las Tablas 11 y 12 se muestran la matriz de confusión y los valores de sensibilidad y especificidad para este clasificador.

Tabla 12. Matriz de confusión del clasificador por Arboles de Clasificación

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	775	196	117	8
	BDG	133	567	62	61
	FSJ	216	60	353	20
	MN	10	28	61	431

Tabla 13. Valores de sensibilidad y especificidad del clasificador por Arboles de Clasificación

	PTK	BDG	FSJ	MN
Sensibilidad	70.7 %	68.9 %	54.4 %	81.3 %
Especificidad	82.1 %	87.5 %	90.2 %	96.5 %

La TCC obtenida para este clasificador es la más alta con un valor de 69%. Es importante destacar que este valor superior se debe en gran medida a que el clasificador no se basa en una distribución definida de las variables de entrada. Además la forma de clasificación a través de subdivisiones de áreas se aproxima al SIST que a su vez es una forma intuitiva de

## Capítulo 4 Resultados y Discusión

aproximación del problema. Vale destacar que su resultado se ve afectado por la gran cantidad de FSJ clasificadas como PTK.

### 4.3.6 Clasificador naive Bayes

Por último el clasificador naive Bayes. Al igual que el QDA los parámetros de entrada de este sistema deben seguir una distribución normal. En la Fig. 21 se muestra las áreas definidas.

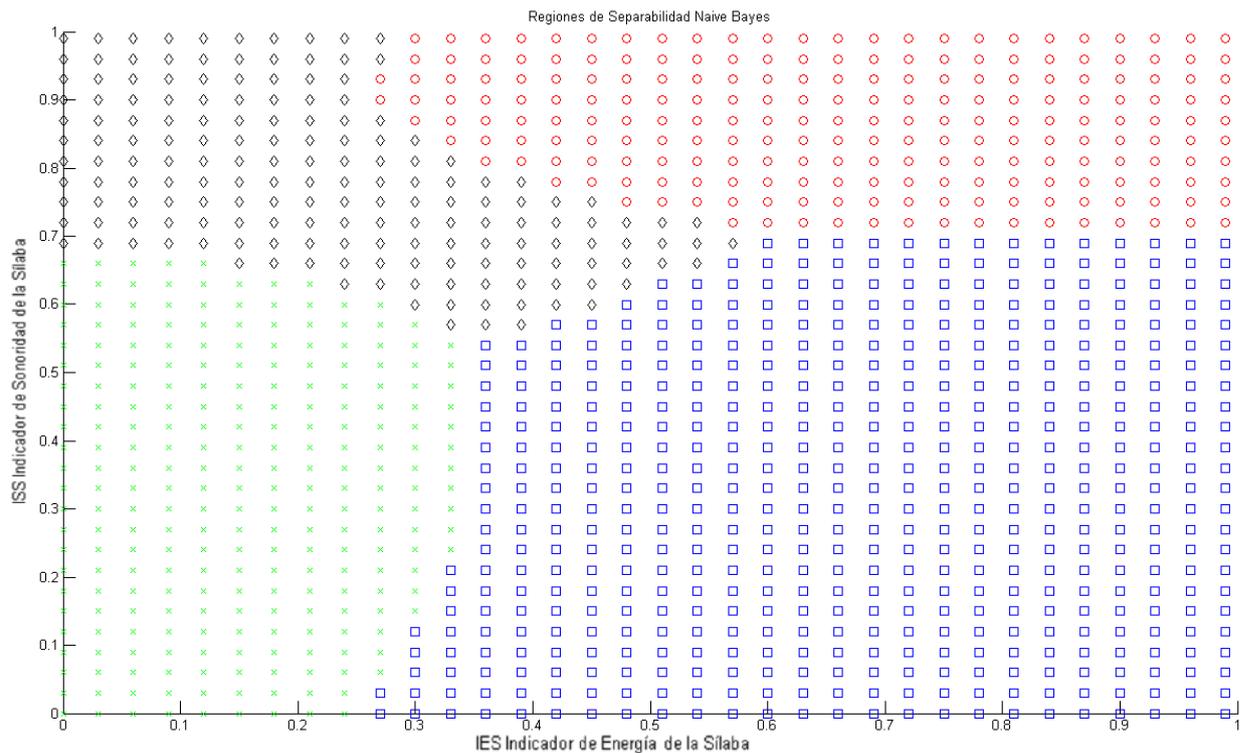


Fig.21. Regiones según el clasificador naive Bayes.

En las Tablas 14 y 15 se muestran los resultados numéricos.

Tabla 14. Matriz de confusión de naive Bayes

		Resultados de Clasificación			
		PTK	BDG	FSJ	MN
Clasificación Real	PTK	815	200	44	37
	BDG	218	488	13	104
	FSJ	345	76	112	116
	MN	8	24	1	497

Tabla 15. Valores de sensibilidad y especificidad de naives Bayes

## Capítulo 4 Resultados y Discusión

	PTK	BDG	FSJ	MN
Sensibilidad	74.4 %	59.3 %	17.3 %	93.7 %
Especificidad	71.4 %	86.8 %	97.6 %	89.9 %

El valor de TCC para este caso es de 62%. La afectación correspondiente a un valor alto de FSJ clasificadas como PTK, afecta a su vez el valor de sensibilidad de estas consonantes.

### 4.4 Conclusiones Parciales del Capítulo

Se obtuvieron poblaciones significativas y balanceadas de las distintas clases de sílabas, y se comprobó la separabilidad de las mismas en el plano (*EvsS*). Las regiones contiguas obtenidas se encuentran en posiciones correctas desde el punto de vista teórico (características acústicas que deben tener las consonantes). Se observa un elevado solapamiento de las fricativas (FSJ) con las otras clases, fundamentalmente con las sordas (PTK).

## Conclusiones

### Conclusiones

- Se obtuvo una muestra representativa de las consonantes involucradas en el trabajo (1096 PTK, 823 BDG, 649 FSJ, 530 MN).
- Se utilizaron los clasificadores implementados en la versión comercial de Matlab R2010a (QDA, NB, DT, SIB, KNN).
- Se logra una separación aceptable en cuanto a TCC (~70% en el árbol de decisión), arribando a regiones relativamente bien definidas, en las localizaciones esperadas.
- El mayor problema se encuentra en el solapamiento entre FSJ-PTK. Esto es visible gráficamente y en las matrices de confusión de los clasificadores. De los cuales se brinda un resumen en la Tabla 16.

## Conclusiones

Tabla 16. Resumen de los resultados de los clasificadores, en rojo se muestran los mayores valores por filas.

Clasificador	SI Gauss	SI Trapez.	KNN (3%)	QDA	Tree (9 LvL)	NB
Tasa de clasificación (TC)	62%	64%	68%	61%	69%	62%
Sensibilidad PTK	67%	61%	74%	56%	71%	74%
Especificidad PTK	78%	84%	80%	84%	82%	71%
Sensibilidad BDG	68%	49%	58%	65%	69%	59%
Especificidad BDG	81%	94%	91%	83%	88%	87%
Sensibilidad FSJ	37%	73%	48%	43%	54%	17%
Especificidad FSJ	92%	78%	93%	91%	90%	98%
Sensibilidad MN	75%	83%	93%	87%	81%	94%
Especificidad MN	95%	96%	91%	89%	97%	90%

## **Recomendaciones**

### **Recomendaciones**

- Se propone el desarrollo de un tercer indicador capaz de separar estas dos clases.
- Se propone utilizar las regiones obtenidas en una estimación del rasgo de CI, que brinde así una medida objetiva de la imprecisión articular, y funcione como apoyo a las valoraciones de expertos, permitiendo diagnosticar el nivel de severidad del TML en un paciente.

## Bibliografía

## Bibliografía

- [1].Ali, A. M. A., “Acoustic features of the automatic recognition of fricatives”, Technical Report, TR-CST27AUG97, Center for Sensor Technologies, University of Pennsylvania, 1997.
- [2].Berry, W.R.; Aronson, A.E.; Darley, F.L. & Goldstein, N.P. “Effects of penicillamine therapy and low-copper diet on dysarthria in Wilson’s disease.1974.
- [3].Berry, W.R.; Darley, F.L.; Aronson, A.E. & Goldstein, N.P. “Dysarthria in Wilson’s disease”. Journal of Speech & Hearing Research. 17, pp 169-183.1974.
- [4].Darley, F.L.; Aronson, A.E. & Brown, J.R. “Clusters of deviant speech dimensions in the dysarthria”. Journal of Speech & Hearing Research. 12, pp 462- 496, 1969.
- [5].Darley, F.L.; Aronson, A.E. & Brown, J.R. “Differential diagnostic patterns of dysarthria”. Journal of Speech & Hearing Research. 12, pp 246-269, 1969.
- [6].Darley, F.L.; Aronson, A.E. & Brown, J.R. “Motor Speech Disorders”. Philadelphia. Saunders. 1975.
- [7].Darley, F.L.; Aronson, A.E. & Brown, J.R. “Motor Speech Signs in Neurologic Disease”. Medical Clinics of North America, 52, 835-844. 1968.
- [8].Darley, F.L.; Brown, J.R. & Goldstein, N. “Dysarthria in multiple sclerosis”. Journal of Speech & Hearing Research. 15, pp 229-245, 1972.
- [9].Enderby, P. “Frenchay Dysarthria Assessment”. California. College Hill. 1983.
- [10].Enderby, P. “Relationship between dysarthric groups”. British Journal of Disorders of Communications. No 21. pp 189-197. 1986.
- [11].Ferrer, C. A. Cuantificación de parámetros de la voz para aplicaciones médicas. Tesis Doctoral, Universidad Central "Marta Abreu" de las Villas. 2005.
- [12].Ferrer, C. A. Cuantificación de parámetros de la voz para el diagnóstico de disartrias. Tesis de Maestría, Universidad Central "Marta Abreu" de las Villas, 2000.
- [13].Ferrer, C., González, E. “Determinación del grado de imprecisión de consonantes explosivas empleando niveles de energía y sonoridad”. Bioingeniería y Física Médica Cubana. Vol. 5, 2004
- [14].Ferrer, C.; Hernández, M.E. & González, E. Isolated Syllable Position Detector in Recordings of Patients With Motor Speech Disorders Using Speech Processing

## Bibliografía

- Techniques. Proceedings of the TELECOM'02. International Conference, Santiago de Cuba, July. 2002.
- [15]. Ferrer, C.A., del Toro, A. & González, E. Estimation of Consonant Imprecision Ratings by using Neural Networks and Acoustic Measures of Energy and Sonority. Center for Studies on Electronics and Information Technologies, Central University of Las Villas, Zip code: 54830, Santa Clara, Cuba. 2005
- [16]. Hughes, G. W. and Halle, M., "Spectral properties of fricative consonants: Journal of the Acoustical Society of America, 1956, 28, pp. 303-310.
- [17]. Jiang, J. J., Shah, A. G. & Hanson, D. G. 1999. Voice measurement: importance of voice analysis and measurement to patient management. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 7, 119-124.
- [18]. Klatt, D.H and Klatt, L. C., "Analysis, synthesis and perception of voice quality variations among female and male talkers", *Journal of the Acoustical Society of America*, 1990, 87, pp. 820-857.
- [19]. Llanes, R., Ferrer, C., Díaz, R., Torres D. "Caracterización de consonantes mediante indicadores de energía y sonoridad". *Reconocimiento de Patrones e Ingeniería Biomédica*, 2008.
- [20]. McCasland, G. P., "Noise intensity and spectrum cues for spoken fricatives", *Journal of the Acoustical Society of America*, 1979, Suppl.1, 65, S78-S79.
- [21]. Morales, L. "Sistema de diagnóstico sobre una Base de Casos y Redes Neuronales para el diagnóstico diferencial de las disartrias". Tesis para optar por el título de Master en Electrónica. UCLV. VC. Cuba. Julio. 1999.
- [22]. Ohde, R. N. & Sharf, D.J. "Phonetic Analysis of Normal and Abnormal Speech". MacMillan Publishing Co. New York. 1992.
- [23]. Perrin, E.; Berger-Brachon, C.; Kauffman. I. & Collet, L. "Acoustical recognition of laryngeal pathology using the fundamental frequency and the first three formants of vowels". *Medical & Biological Engineering & Computing*. No 35. pp 361-368. July. 1997.
- [24]. Ruiz Riol, R. "Programa para el cálculo de parámetros que cuantifican características de la voz". Tesis de Grado de Ingeniero en Telecomunicaciones y Electrónica. UCLV. VC. Cuba. Julio. 1999.

## Bibliografia

- [25].R. O.; Hart, P. E.; Stork, D. G. ‘‘Pattern Classification’’. Second Edition. 2000.
- [26].Statistic Toolbox User’s Guide. MathWorks. January. 1998.
- [27].Titze, I.R & Liang, H. ‘‘Comparison of F0 Extraction methods for high precision voice perturbation measurements’’. Journal of Speech & Hearing Research. 36. pp 1120-1139. 1993.
- [28].Wendall, R.W. ‘‘Approach to the objective evaluation of hoarseness’’. Folia Phoniatica. No 18. pp 98-108. 1966.
- [29].Wendall, R.W. ‘‘Laryngeal Analog Synthesis of Harsh voice Quality’’. Folia Phoniatica. No 15. pp 251-255. 1963.
- [30].Yanagihara, N. ‘‘Significance of harmonic changes and noise components in hoarseness’’. Journal of Speech & Hearing Research. 10, pp 531-541, 1967.
- [31].Yegnanarayana, B. & Veldhuis, R.N.J. ‘‘Extraction of Vocal-Tract System characteristics from Speech Signals’’. IEEE Transactions on Speech and Audio Processing. Vol 6. No 4. pp 313-327. July. 1998.
- [32].Yorkston, K.M.; Beukelman, D.R. & Bell, K. ‘‘Clinical Management of Dysarthric Speakers’’. PRO-ED. Austin. Texas. 1987.
- [33].Yorkston, K.M.; Beukelman, D.R. & Bell, K. ‘‘Clinical Management of Dysarthric Speakers’’. PRO-ED. Austin. Texas. 1987.
- [34].Yumoto, E. ‘‘The quantitative evaluation of hoarseness. A new harmonics to noise ratio method’’. Archives of Otolaryngology. Vol 109 pp 48-52. 1983.
- [35].Yumoto, E. Gould, W.J. & Baer, T. ‘‘The harmonic-to-noise ratio as an index of the degree of hoarseness’’. Journal of the Acoustical Society of America. Vol. 71. pp 1544-1550. 1982.
- [36].Yunik, M. & Boyanov, B. ‘‘Method for evaluation of the noise-to-harmonics-components ratios in pathological and normal voice’’. Acustica. Vol 70. pp 89-91. 1990.

## Anexos

## Anexos

### A-1 Detector de Sílabas utilizado

```
function [PosVec,StartVec,Env]=SPosDet(y,fs,bps,MinTrans,Step);
%function [PosVec,StartVec,Env]=SPosDet(y,fs,bps);
%A SPosDet se le pasa una alocaucion de PTK en Y, y en el se
%determina la posición de cada sílaba. Env está muestrada a 50 Hz
%La envolvente de la energia se devuelve en Env

%Definición de constantes usadas
S20Ms=round(0.02*fs);%20 milisegundos de ventana
Step=round(0.001*fs);%milisegundos de salto
%Obtencion de la senal de energia
%disp('haciendo conv')
MinSep=round(0.1*fs/Step);%Minima separación entre sílabas>100mS(ó 10/seg)(en
muestras)
MinMaxAmp=20;%Minima Amplitud de un maximo para ser tenido en cuenta=5*EnMin
%MinTrans=0.6;%Minimo descenso de Energy para separar dos sílabas=25%

a=abs(y-mean(y));%el valor absoluto para el promedio móvil
h=hamming(S20Ms);
Energy=SampledConv(a,h,Step,1);
%Ahora fs es de 50 Hz
EnDeriv=[0 diff(Energy)'];
i=2:length(EnDeriv);
PosMax=find(EnDeriv(i)<=0 & EnDeriv(i-1)>0);%Todos los maximos
PosMin=find(EnDeriv(i)>0 & EnDeriv(i-1)<=0);%Todos los minimos
%vector de las x
x=0:length(y)/length(Energy)/fs:(length(y)-1)/fs;
%
%figure(1)
%ClF;
%a=zeros(1,length(Energy));
%a(PosMax)=max(Energy);
%plot(x,Energy);
%hold on
%grid
%zoom
%plot(x,a,'r');

%Obtener los maximos importantes
PosAllMax=PosMax;%Dejar una copia de todos
length(PosMax);
i=2;MinEn=min(Energy);
while i<length(PosMax),%Con este while quedan dos en las esquinas
    if (((PosMax(i)-PosMax(i-1) < MinSep) & (Energy(PosMax(i)) <
Energy(PosMax(i-1))))...
        | ( PosMax(i+1)-PosMax(i) < MinSep) & (Energy(PosMax(i)) <
Energy(PosMax(i+1))))...
        | (Energy(PosMax(i))< MinMaxAmp*MinEn)),
        PosMax(i)=[];%Eliminar los que estan muy cercanos (el de menor En), o
En<K*EnMin
```

## Anexos

```
        if i>3,%corregir el count(hay menos elem.) pa'q' chequee al previo en
las new cond.
            i=i-2;
        else
            i=i-1;
        end;%MinTrans*Energy(PosMax(i))>max([min(Energy(PosMax(i-1):PosMax(i)))
min(Energy(PosMax(i):PosMax(i+1))))],
        end;
        i=i+1;
        j=i;
    end
    PosMax(1)=[];PosMax(length(PosMax))=[];%Eliminar los de la esquinas

%
%a=zeros(1,length(Energy));
%a(PosMax)=max(Energy);
%plot(x,a,'k')%En negro los grandes bien separados
%

%Ahora analizar los que quedan para determinar sílabas
PosMaxSG=PosMax;%Una copia de los separados y de gran amplitud
i=2;
while i<length(PosMax),%
    %if min(Energy(PosMax(i-1):PosMax(i))) > MinTrans*(min([Energy(PosMax(i-
1)) Energy(PosMax(i))]))...
    %    );%+ abs(Energy(PosMax(i-1)) - Energy(PosMax(i)))/10 ),%la
diferencia influye 1/10
    if (min(Energy(PosMax(i-1):PosMax(i))) > MinTrans*Energy(PosMax(i))) |...
        (min(Energy(PosMax(i):PosMax(i+1))) > MinTrans*Energy(PosMax(i))),
        PosMax(i)=[];%Eliminar donde no hubo transicion,(cruce por En<K*EnMin)
        i=i-1;
    end;
    i=i+1;
end
PosVec=PosMax;

%
%a=zeros(1,length(Energy));
%a(PosMax)=max(Energy);
%plot(x,a,'r')%En rojo los considerados como sílabas
%

%Ahora determinar el comienzo de la silaba, como el segmento más largo
%de pendiente positiva continua
%i=1:length(PosMax);
i=2;
PosOnset=[];
while i<length(PosMax),
    StartPos=find( Energy == min(Energy(PosMax(i-1):PosMax(i))) );%comienzo un
seg. en el min.
    SegLength=0;
    while StartPos < PosMax(i),
        j=StartPos+1:PosAllMax(min(find(PosAllMax>StartPos)));%un segmento de
pendiente positiva
        if SegLength < sum( sqrt( (Energy(j)-Energy(j-1)).^2) ),
            SegLength=sum( sqrt( (Energy(j)-Energy(j-1)).^2) );
        end;
    end;
    PosOnset=[PosOnset;StartPos];
    i=i+1;
end
```

## Anexos

```
        TempPosOnset=find(EnDeriv == max(EnDeriv(j)));
    end;
    StartPos=PosMin(min(find(PosMin>max(j))));
    end;
    PosOnset=[PosOnset TempPosOnset];
    i=i+1;
end;

%
%a=zeros(1,length(Energy));
%a(PosOnset)=max(Energy);
%plot(x,a,'k')%En negro los inicios de sílabas
%

StartVec=PosOnset;%Salida de las posiciones (en muestras)
Env=Energy;
```