

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnologías de la Información



TRABAJO DE DIPLOMA

**Separación de Consonantes a partir de la
combinación de los Indicadores ISS, IES, DCCM.**

Autor: Javier David Brizuela Cardoso

Tutor: Msc. Yoandy Pérez Peñate

Cotutor: Dr.T. Carlos A. Ferrer Riesgo

Santa Clara

2014

"Año 56 de la Revolución"

Universidad Central “Marta Abreu” de Las Villas

Facultad de Ingeniería Eléctrica

Centro de Estudios de Electrónica y Tecnologías de la Información



TRABAJO DE DIPLOMA

**Separación de Consonantes a partir de la
combinación de los Indicadores ISS, IES, DCCM.**

Autor: Javier David Brizuela Cardoso

e-mail: jbrizuela@uclv.edu.cu

Tutor: Msc. Yoandy Pérez Peñate

Centro de Ingeniería Clínica y Electromedicina, Villa Clara.

e-mail: yoandyp@capiro.vcl.sld.cu

Cotutor: Dr.C. Carlos A. Ferrer Riesgo

e-mail: cferrer@uclv.edu.cu

Santa Clara

2014

"Año 56 de la Revolución"



Hago constar que el presente trabajo de diploma fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de estudios de la especialidad de Ingeniería Biomédica, autorizando a que el mismo sea utilizado por la Institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos, ni publicados sin autorización de la Universidad.

Firma del Autor

Los abajo firmantes certificamos que el presente trabajo ha sido realizado según acuerdo de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

Firma del Autor

Firma del Jefe de Departamento
donde se defiende el trabajo

Firma del Responsable de
Información Científico-Técnica

DEDICATORIA

A mi Familia en especial a mis padres Gisela, Candido, Mandy y Mary por ser luz y guía en mi vida. Son los mejores

A mi hermana por el compromiso de ser cada día su ejemplo a seguir.

A mi novia por su cariño y comprensión.

A mis amigos para aquellos que en las buenas y en las malas nunca han flaqueado.

A mis profesores por la dedicación tan sagrada y humana de educarnos

AGRADECIMIENTOS

A todos aquellos que de una forma u otra me contribuyeron a la realización de este trabajo. En especial a mis tutores Yoandy y Carlos por la profesionalidad, por el tiempo dedicado, por soportarme.

TAREA TÉCNICA

- ✓ La realización de un estudio de los trabajos vinculados al tema.
- ✓ La caracterización de los diferentes tipos de consonantes.
- ✓ Propuesta de distintos índices.
- ✓ Selección de clasificadores.
- ✓ Evaluación de los índices.

Firma del Autor

Firma del Tutor

RESUMEN

Mediante este trabajo se pretende contribuir a los trabajos precedentes, en la separación de las consonantes *P*, *T*, *K* y sus sustituciones resultantes del ejercicio de diagnóstico diferencial de disartrias creado por Darley, Aronson y Brown (*DAB*) en las Clínicas Mayo, mejorar la separación de las consonantes *P*, *T*, *K* y sus sustituciones *B*, *D*, *G* y *F*, *S*, *J*. Primeramente se reprodujo los experimentos utilizando una misma Base de Datos mediante las combinaciones (Indicador de Sonoridad de la Sílabla (*ISS*)- Indicador de Energía de la Sílabla (*IES*)) y (Indicador de Energía de la Sílabla (*IES*) con Densidad de Cruces por Cero Modificado (*DCCM*)). Se evaluó el desempeño de estos tres indicadores en conjunto, para lo cual se calculan índices de clasificación, empleando varios métodos de clasificación y comparando sus resultados.

INDICE

INTRODUCCIÓN.....	1
CAPÍTULO 1. ESTADO DEL ARTE EN LA CARACTERIZACIÓN Y SEPARACIÓN DE LAS CONSONANTES.	4
1.1 Definición y clasificación de los fonemas en español.	4
1.1.1 Clasificación.	4
1.1.2 Grupo de consonantes bajo estudio.	7
1.1.2.1 Oclusivas.....	7
1.1.2.2 Fricativas.....	8
1.1.2.3 Nasaes.	8
1.2 Producción de la voz.	9
1.2.1 Características de la señal de voz en el dominio del tiempo.	9
1.2.1.1 Forma de onda de la señal de voz.....	9
1.2.1.3 Cruces por cero.	9
1.2.2 Tipo de señales de voz.	10
1.2.2.1 Señal sonora.....	10
1.2.2.2 Señal sorda.....	11
1.3 Alteraciones del habla.	11
1.3.1 Tipos de alteraciones del habla.....	11
1.3.2 Disartria.	12

1.4	Antecedentes.....	13
1.4.1	Detectores de sílabas.....	13
1.4.2	Indicadores en el dominio del tiempo.....	13
1.4.2.1	Indicador de sonoridad de la sílaba (<i>ISS</i>).....	13
1.4.2.2	Indicador de energía de la sílaba (<i>IES</i>).	14
1.4.2.3	Rasgo de Densidad de Cruces por Cero.....	14
1.4.3	Estudios <i>ISS</i> vs. <i>IES</i>	15
1.4.4	Estudios <i>IES</i> vs. <i>DCCM</i>	17
CAPÍTULO 2. MATERIALES Y MÉTODOS.....		20
2.1	Bases de Datos.	20
2.1.1	Grabaciones utilizadas como material de estudio.....	20
2.2	Detector de sílabas.	21
2.2.1	Detector de centros de sílabas.....	22
2.3	Indicadores en el dominio del tiempo.	22
2.3.1	Indicador de sonoridad de la sílaba (<i>ISS</i>).....	22
2.3.2	Indicador de energía de la sílaba (<i>IES</i>).	23
2.3.3	Indicador de Densidad de Cruces por Cero Modificado (<i>DCCM</i>).	24
2.4	Métodos de clasificación empleados e índices de cuantificación para comprobar los resultados.....	25
2.4.1	Clasificación mediante el método de Árbol de decisión.....	26
2.4.2	Análisis Discriminante.....	27
2.4.3	Método de los k vecinos más cercanos (<i>KNN</i>).....	29
2.4.4	Índices de cuantificación de los resultados.....	31
CAPÍTULO 3. RESULTADOS DE LA CLASIFICACIÓN.....		33
3.1	Distribución de consonantes usando indicadores de la señal de voz.	33

3.1.1	Distribución mediante la sonoridad de las sílabas (<i>ISS</i>).	33
3.1.2	Distribución mediante la energía de las sílabas (<i>IES</i>).	34
3.1.3	Distribución mediante la densidad de Cruces por Cero (<i>DCCM</i>).	35
3.2	<i>ISS</i> vs. <i>IES</i>	35
3.2.1	<i>ISS</i> vs. <i>IES</i> en la separación de los cuatro grupos de consonantes.....	37
3.2.1.1	Clasificación de las consonantes mediante análisis discriminante con distancia de Mahalanobis (<i>AD-DM</i>).	37
3.2.1.2	Clasificación mediante análisis discriminante cuadrático con matriz de covarianza diagonal (<i>AD-MCD</i>).	40
3.2.1.3	Clasificación mediante el método del árbol de decisión.....	42
3.2.1.4	Clasificación mediante el método de los k vecinos más cercanos (<i>KNN</i>). ..	44
3.2.2	Discusión de los resultados de los clasificadores en <i>ISS</i> vs. <i>IES</i>	46
3.3	<i>IES</i> vs. <i>DCCM</i>	46
3.3.1	<i>IES</i> vs. <i>DCCM</i> en la separación de los cuatro grupos de consonantes.....	48
3.3.1.1	Clasificación de las consonantes mediante análisis discriminante cuadrático con matriz de covarianza diagonal (<i>AD-MCD</i>).	48
3.3.1.2	Clasificación de las consonantes mediante análisis discriminante con distancia de Mahalanobis (<i>AD-DM</i>).	50
3.3.1.3	Clasificación mediante el método de los k vecinos más cercanos (<i>KNN</i>). ..	52
3.3.1.4	Clasificación mediante el método de árbol de decisión.....	54
3.3.2	Discusión de los resultados de los clasificadores en <i>IES</i> vs. <i>DCCM</i>	56
3.4	Comparación entre <i>IES</i> vs. <i>DCCM</i> y <i>ISS</i> vs. <i>IES</i> en la clasificación de consonantes.	56
3.5	<i>ISS</i> , <i>IES</i> y <i>DCCM</i> , de conjunto, en la separación de los cuatro grupos consonantes.	58
3.5.1	Clasificación a partir de los distintos métodos empleados.	59

3.5.1.1	Clasificación mediante análisis discriminante cuadrático con matriz de covarianza diagonal (<i>AD-MCD</i>).	59
3.5.1.2	Clasificación mediante el análisis discriminante con distancia de Mahalanobis (<i>AD-DM</i>).	60
3.5.1.3	Clasificación mediante el método de k vecinos más cercanos.	61
3.5.1.4	Clasificación mediante el método de árbol de decisión.	63
3.5.2	Discusión de los resultados de los clasificadores en ISS, IES y DCCM en conjunto.	64
3.5.3	Clasificación mediante la separación empírica manual de las regiones.	65
3.5.3.1	Ajuste fino de los valores y cálculo de la TCT.	67
3.5.4	Conclusiones parciales de los resultados.	69
CONCLUSIONES Y RECOMENDACIONES		70
REFERENCIAS BIBLIOGRÁFICAS		72

INTRODUCCIÓN

Durante el transcurso de la vida del hombre, le ha sido necesario comunicarse mediante el lenguaje hablado, condición tan importante para su desarrollo como ser social y racional. Si generalmente para algunas personas es muy simple establecer una conversación, para otras que padecen trastornos en la comunicación oral les resulta algo bien complicado. Debido al mal funcionamiento de los mecanismos fisiológicos que llevan consigo la emisión del habla, en los últimos tiempos se le ha dedicado una especial atención a los estudios realizados con el fin de mejorar las técnicas para el diagnóstico y tratamiento de las patologías neurológicas como son los trastornos motores del lenguaje (*TML*). Dentro de estos trastornos se encuentran dos grupos: la Apraxia y la Disartria, las cuales se diferencian en que la Disartria es provocada por el descontrol en aquellos músculos que intervienen directamente en la producción de la voz, mientras que la Apraxia no es causada por dicho descontrol.

Darley, Aronson y Brown (*DAB*) llevaron a cabo una serie de estudios [1, 2, 3, 4, 5, 6, 7] en las Clínicas Mayo a fines de la década de los 60' y principios de los 70' donde a partir de estos crearon una metodología para el análisis y diagnóstico diferencial de disartrias. Dicha metodología se basa en las valoraciones subjetivas de 38 rasgos perceptuales que emite un panel de expertos y determinan la presencia de agrupaciones distintivas de rasgos por disartrias, tras escuchar de un grupo de pacientes los ejercicios siguientes: la fonación de la vocal "a" sostenida durante un tiempo determinado, repetición de los monosílabos explosivos "Pa", "Ta" y "Ka" y la lectura de un segmento de un párrafo modelo. En la actualidad aún estos estudios son considerados los cimientos del diagnóstico diferencial clínico de disartrias, pero resulta inconforme la obtención de resultados totalmente

dependientes de las experiencias que pueda tener un grupo de especialistas, dado por subjetividad de las evaluaciones.

Debido a lo anterior planteado en el Centro de Estudios de Electrónica y Tecnología de la Información, Ferrer y sus colaboradores desarrollaron una serie de trabajos [12] [13] [14] [15] con el fin de determinar, a través de técnicas de procesamiento digital de señales, una medida objetiva de imprecisión articulatoria correspondiente con los resultados obtenidos tras el rasgo de Consonantes Imprecisas (CI) en la pronunciación repetida de los monosílabos. Con estos trabajos se llegó a la conclusión de que la imprecisión está dada a través de las sustituciones de las consonantes explosivas sordas como son las fricativas, nasales y explosivas sonoras.

Luego para la caracterización y clasificación de estos grupos de consonantes se emplearon en [12] [9] [10] algunos índices como son el Indicador de Sonoridad de la Sílabla (*ISS*) y el Indicador de Energía de la Sílabla (*IES*). Aunque estos rasgos fueron capaces de separar los grupos de consonante: explosivas sonoras de las nasales así como las explosivas sordas de las explosivas sonoras; no fue así entre las fricativas y las explosivas sordas que se encuentran altamente solapadas. Por tal motivo se creó y se evaluó en [11] un nuevo índice el cual denominó Densidad de Cruces por Cero Modificado (*DCCM*) combinándose con el (*IES*). En comparación con los resultados obtenidos en la combinación de *ISS* y *IES*; fue posible mejorar la separación de las explosivas sordas del resto; pero aún se puede lograr mejores resultados.

Por lo que este trabajo está orientado a contribuir a los experimentos mencionados, mediante el uso en conjunto de estos indicadores en el dominio del tiempo para la separación de los grupos de consonante.

Interrogante científica

¿Es posible mediante la combinación de los indicadores *ISS*, *IES* y *DCCM* mejorar los índices de clasificación logrados a partir de : *ISS* vs. *IES* y *IES* vs. *DCCM*?

Objetivos

Objetivo general

Evaluar la separación de consonantes combinando los indicadores *ISS*, *IES* y *DCCM*.

Objetivos específicos

- Reproducir los experimentos anteriores: ISS vs. IES y IES vs. DCCM para verificar la equivalencia de la base de datos.
- Evaluar la efectividad de los indicadores ISS, IES y DCCM, de conjunto, en la separación de las consonantes, a través del empleo de diferentes clasificadores multiclase.

Estructura del trabajo

En el Capítulo 1 se exponen los principios teóricos que fundamentan este trabajo, así como aspectos generales en torno al tema en cuestión, abordando de manera sintetizada los resultados de los trabajos que anteceden y motivan el presente.

En el Capítulo 2 se describe brevemente el detector de sílabas utilizado para localizar las consonantes dentro de las grabaciones. Se comentan los indicadores ISS, IES y DCCM usado en estudios precedentes los cuales se emplean en la separación de consonantes. Luego se prosigue a la descripción de los métodos de clasificación implementados, y se detallan los índices a partir de los cuales se cuantifica la efectividad de los clasificadores.

En el Capítulo 3 se muestran los resultados de reproducir las combinaciones precedentes de los indicadores y de conjunto los tres indicadores en la separación de los distintos grupos de consonantes.

Por último se exponen las conclusiones finales de la tesis y las recomendaciones para trabajos futuros.

El cuerpo de la tesis consta de 67 páginas, con 32 tablas, 52 figuras y 18 ecuaciones contenidas en el texto.

CAPÍTULO 1. ESTADO DEL ARTE EN LA CARACTERIZACIÓN Y SEPARACIÓN DE LAS CONSONANTES.

En este capítulo se exponen los aspectos generales de los fonemas sobre todo de las consonantes. Se hace una introducción al tema de las alteraciones del habla entre ellas las disartrias, su definición, causas y algunos ejemplos. Se hace un resumen de los trabajos relacionados con la separación de las consonantes así como resultados obtenidos de los distintos indicadores empleados para obtener dicha separación.

1.1 Definición y clasificación de los fonemas en español.

A las unidades lingüísticas más pequeñas e indivisibles se les conoce como fonemas, su número es reducido. No tienen significado por sí mismo, pero el significado de una palabra cambia si se intercambia dos fonemas, estos cambios o diferencias se establecen por el fenómeno de la oposición, o sea no pueden distinguirse una cosa de otra si no se oponen entre sí. Los fonemas en español se clasifican en Vocales y Consonantes [8].

En la emisión de los sonidos de la voz, que tienen más o menos obstáculos en su recorrido a lo largo de todo el aparato fonador se denominan consonantes. Una consonante presenta en el tracto vocal mayor grado de constricción que la vocal, esto se debe al acercamiento o contacto de los órganos de articulación de tal manera que ocasiona turbulencia audible. Precisamente en el español se hace necesario que toda sílaba presente al menos una vocal, debido a que las consonantes no tendrían pronunciación.

1.1.1 Clasificación.

Para la clasificación de los fonemas (Tabla 1) se tienen en cuenta los siguientes criterios:

- Según el punto de articulación: indica el lugar de las cavidades supraglóticas donde se produce un determinado fonema. De acuerdo con los diferentes puntos de articulación se pueden distinguir los siguientes fonemas:

Consonantes:

- Bilabiales: Contacta el labio superior e inferior, ejemplo: /b, m, p/.
- Labiodentales: Contacta el labio inferior con incisivos superiores, ejemplo: /f/.
- Linguointerdentales: Se sitúa la lengua entre los incisivos superiores e inferiores, ejemplo: /θ/.
- Linguodentales: Contacta el ápice de la lengua y la parte trasera de los incisivos superiores, ejemplo: /t, d/.
- Linguoalveolares: Contacta el ápice o predorso de la lengua con los alvéolos superiores, ejemplo: /l, s, n, r/.
- Linguopalatales: Contacta el predorso de la lengua el paladar duro, ejemplo: /tʃ, ʎ/
- Linguovelares: Se aproxima o toca el postdorso de la lengua con el velo del paladar, ejemplo: /x, k, g/.

Vocales:

- Anteriores: La lengua se aproxima a la región delantera o zona del paladar duro, ejemplo: /i, e/.
 - Centrales: La lengua se encuentra en la parte central del paladar, ejemplo: /a/.
 - Posteriores: La lengua se aproxima a la zona velar, ejemplo: /o, u/.
- Según el modo de articulación: depende de la posición que adopte los órganos durante el habla, es decir, se refiere los elementos o procesos que se ponen en marcha para pronunciar un fonema. Se distinguen entre explosivas y no explosivas. Las primeras son aquellas que se producen cuando hay mayor o menor obstaculización en el aparato fonador, mientras que las segundas se pronuncian de una forma más directa.

Consonantes:

- Oclusivas: Se establece un cierre completo de los órganos articulatorios y el aire sale de forma explosiva tras la interrupción, ejemplo: /b, d, g, p, t, k/, ejemplo de no explosivas: /m, n/.
- Fricativas: Existe un estrechamiento de dos órganos articulatorios donde pasa el aire espirado de manera explosiva, ejemplo: /f, θ, s, x, j/.
- Africadas: Se forma por combinación de una oclusiva seguida de una fricativa, ejemplo: /tʃ/.
- Laterales: Durante su emisión el aire se escapa por un lado o por los dos, por lo que es no explosiva, ejemplo: /l, λ/.
- Vibrantes: Se produce una o varias vibraciones del ápice de la lengua lo que la hace no explosiva, ejemplo: /r/.

Vocales:

- Cerradas: La lengua se encuentra muy cerca del paladar, ejemplo: /i, u/.
- Medias: La lengua está una distancia interna del paladar, ejemplo: /e, o/.
- Abiertas: La lengua se separa totalmente del paladar, /a/.
- Según el modo de fonación: se relaciona con la vibración o no vibración de las cuerdas vocales:
 - Sordas: No existe vibración de las cuerdas vocales, ejemplo: /p, t, k, f, θ, tʃ, s, x/.
 - Sonoras: Existe vibración de las cuerdas vocales, ejemplo: /b, d, g, j, m, n, l, λ, r/ y /a, e, i, o, u/.
- Según la acción del velo del paladar:
 - Nasales: El velo del paladar está separado de la pared faríngea produciéndose un escape de aire por la cavidad nasal, ejemplo: /m, n/.
 - Orales: El velo paladar no está separado de la pared faríngea y no permite el paso de aire a la cavidad nasal.

Tabla 1. Clasificación de las consonantes.

		Posterioridad →						
		bilabial	Labio-dental	Linguo-interdental	Linguo-dental	Linguo-alveolar	Linguo-palatal	Linguo-velar
Oclusiva	sonora	/b/			/d/			/g/
	sorda	/p/			/t/			/k/
Fricativa	sonora							
	sorda		/f/	/θ/		/s/	/j/	/x/
Africada	sonora							
	sorda						/tʃ/	
Nasal	sonora	/m/				/n/	/ɲ/	
Lateral	sonora					/l/	/ʎ/	
Vibrante	simple					/r/		
	múltiple					/r̄/		

1.1.2 Grupo de consonantes bajo estudio.

1.1.2.1 Oclusivas.

También conocidas como **explosivas**, u **oclusivas orales** son aquellas consonantes en las que hay una oclusión (bloqueo) completa de las cavidades oral y nasal del tracto vocal, por lo que no fluye el aire. En el español tenemos, las ya mencionadas, /p t k/ (sordas) y, en algunas posiciones, /b d g/ (sonoras). Si la consonante es sonora, esta resonancia es el único sonido realizado durante la oclusión; si es sorda, la consonante es totalmente silenciosa. Lo que oímos como /p/ o /k/ es el efecto que el inicio de la oclusión tiene en la vocal anterior y la liberación del aire, con el efecto que ésta tiene en la vocal siguiente. La forma y posición de la lengua (el *punto* de articulación) determinan cavidad resonante que da a las diferentes oclusivas sus sonidos característicos. Todas las lenguas cuentan con este tipo de consonantes.

Las oclusivas pueden dividirse además en aspiradas y no aspiradas, que se caracterizan por la producción de un soplo (un ruido sordo, que se consigue manteniendo la glotis abierta

durante la explosión) entre la explosión y la vocal siguiente. Si la aspiración es muy fuerte, las aspiradas tienden a pasar al grupo de las africadas.

1.1.2.2 Fricativas.

Una **consonante fricativa** es aquella que se produce por un estrechamiento o constricción de dos órganos articulatorios, que modifican la corriente de aire, originando una fricción turbulenta; esta es su característica más significativa (en las aproximantes el flujo no llega a ser tan turbulento). Durante su emisión el aire pasa por un espacio muy estrecho que se forma al juntarse ciertas partes de la boca. Las fluctuaciones aleatorias de velocidad en el flujo pueden actuar como una fuente de sonido, y el sonido generado de esta manera es llamado ruido de turbulencia. Si el ruido es generado en o cerca de la constricción glotal es nombrado ruido de aspiración, mientras que si es producido más allá de la glotis es llamado ruido de fricación. Las consonantes fricativas al presentar un comportamiento aleatorio similar al ruido blanco tienen una alta Densidad de Cruces por Cero. La aleatoriedad de este tipo de consonante dificulta la determinación de patrones que las identifiquen.

Algunos estudios más recientes [16], [17] y [18] permiten una mayor profundidad en la investigación de las características consonantes fricativas. Además se han desarrollado estudios encaminados a la detección de estas en [19].

1.1.2.3 Nasales.

Una **consonante nasal** se produce cuando el aire fluye libremente a través de la nariz. La cavidad bucal todavía actúa como compartimiento de la resonancia del sonido, pero el aire no sale por la boca mientras que es bloqueado por la lengua. Acústicamente, las consonantes nasales son sonorantes, es decir, no restringen la salida de aire, y desde el punto de vista lingüístico casi siempre son sonoras. (Compárense las oclusivas orales, que obstruyen el aire completamente, y las fricativas, que obstruyen el aire con un canal estrecho.

Sin embargo, las consonantes nasales también se consideran oclusivas por su articulación, que impide completamente el paso del aire por la cavidad bucal. Tal dualidad, un flujo de aire sonorante por la nariz junto con la oclusión de la boca, implica que las nasales se

comportan como sonorantes al tiempo que como obstruyentes (en general clases opuestas). Para la descripción acústica en general se consideran sonorantes, pero en muchas lenguas se pueden desarrollar a partir de, o devenir en, oclusivas.

1.2 Producción de la voz.

Para la producción del habla se dan [8] los siguientes elementos:

- Una fuente de energía, proporcionada por el aire a presión que se expulsa en la expiración.
- Un órgano vibratorio, las cuerdas vocales.
- Una caja de resonancia, las fosas nasales, la cavidad bucal y la faringe.
- Un sistema de articulación del sonido, úvula, lengua, dientes y labios.

1.2.1 Características de la señal de voz en el dominio del tiempo.

1.2.1.1 Forma de onda de la señal de voz

La señal de voz está constituida por un conjunto de sonidos generados por el aparato fonador. La representación de la señal de voz en función del tiempo es importante puesto que brinda información sobre características importantes como la energía y los cruces por cero, las cuales facilitan su estudio y análisis.

1.2.1.2 Energía.

En una señal continua, la Energía total E en el intervalo de tiempo t_1 a t_2 está definida como (1):

$$E = \int_{t_1}^{t_2} |x(t)|^2 dt \quad (1)$$

Para el caso de las señales discretas donde N es el número de muestras de la señal, la energía se define por (2):

$$E = \sum_{m=0}^{N-1} x(m)^2 \quad (2)$$

1.2.1.3 Cruces por cero.

Los cruces por cero, indican el número de veces que una señal continua toma el valor de cero. Para las señales discretas, un cruce por cero ocurre cuando dos muestras consecutivas difieren de signo, o bien una muestra toma el valor de cero Figura 1. Consecuentemente, las señales con mayor frecuencia presentan un mayor valor de esta característica, el ruido también genera un gran número de cruces por cero.

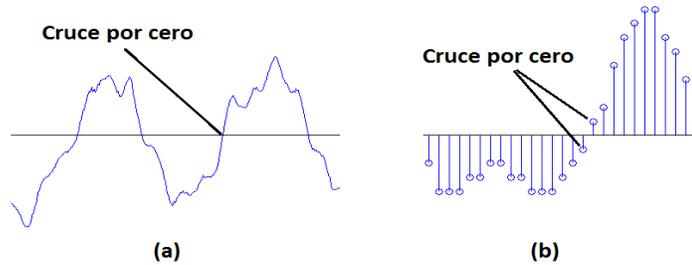


Fig. 1. Cruces por Cero. a) Señal Continua. b) Señal Discreta.

La formulación matemática de la Cantidad de Cruces por Cero (*CCC*) para señales discretas está representada en la siguiente fórmula (3), en la cual, *sign* es la función signo y *N* es el número de muestras de la señal.

$$CCC = \sum_{m=0}^{N-1} |sign[x(m)] - sign[x(m-1)]| \quad (3)$$

1.2.2 Tipo de señales de voz.

Existen múltiples tipos de clasificación de los sonidos (Tabla 1), pero en este epígrafe se describen aquellos directamente vinculados a este trabajo.

1.2.2.1 Señal sonora.

La señal sonora, en la Figura 2 se genera por la vibración de las cuerdas vocales al fluir el aire a través de ella. Estas señales se caracterizan por tener alta energía y un contenido frecuencial en el rango de los 300 Hz a 4000 Hz presentando cierta periodicidad, es decir son de naturaleza cuasiperiódica. El tracto vocal, actúa como una cavidad resonante reforzando la energía en torno a determinadas frecuencias (formantes). En la siguiente figura se muestra el comportamiento de este tipo de señales en el tiempo.

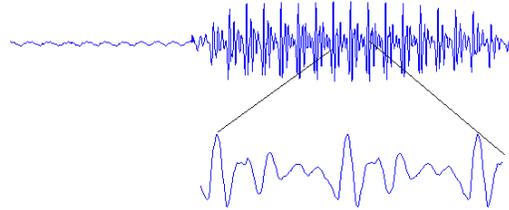


Fig. 2. Señal sonora. Fragmento de la vocal 'a'.

1.2.2.2 Señal sorda.

A esta señal también se le conoce como señal no sonora Figura 3, y se caracteriza por tener un comportamiento aleatorio en forma de ruido blanco. Tienen una alta Densidad de Cruces por Cero y baja Energía comparadas con las señales de tipo sonora. Durante su producción no se genera vibración de las cuerdas vocales, ya que, el aire atraviesa un estrechamiento, y genera una turbulencia. Las consonantes que producen este tipo sonidos son las fricativas entre otras. La siguiente figura muestra la forma de onda de una señal no sonora.

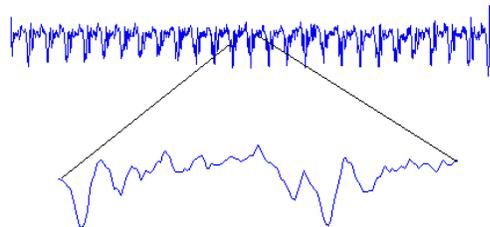


Fig. 3. Señal no sonora o sorda.

1.3 Alteraciones del habla.

1.3.1 Tipos de alteraciones del habla.

Las alteraciones del habla pueden clasificarse en tres tipos, de acuerdo con la escala de tiempo en que se perciben [20], Alteraciones de la Calidad Vocal, Alteraciones de la Prosodia y Alteraciones de la Articulación:

- Alteraciones de la Calidad Vocal: Son perturbaciones que se perciben en sonidos aislados, generalmente vocales. Son característicos de este grupo las perturbaciones de la periodicidad de la señal de excitación generada en la glotis y la hiper/hipo-nasalidad, correspondiente a un inadecuado acople de la cavidad nasal al tracto vocal, que modifica su estructura de formantes [17] [21]. Se reportan múltiples parámetros de la

señal acústica que resultan anómalos a nivel de sonido individual (*jitter*, *shimmer*, ruido aditivo, medidas espectrales) [22].

- **Alteraciones de la Prosodia:** La prosodia es una rama de la lingüística que analiza y representa formalmente aquellos elementos de la expresión oral, tales como el acento, los tonos y la entonación. Desde el punto de vista fonético-acústico la variación de la frecuencia fundamental, la duración y la intensidad constituyen los parámetros prosódicos físicos [23]. Se perciben en unidades del habla de mayor duración, como frases u oraciones y también tienen causa en el control muscular. Como ejemplos pueden mencionarse la acentuación igual y excesiva, las ráfagas cortas del habla, los silencios inapropiados y las monotonías de intensidad y tono.
- **Alteraciones de la Articulación:** Estas perturbaciones se aprecian en las transiciones entre sonidos, por lo que requieren que la persona pronuncie como mínimo sílabas. La causa está en un inadecuado control de los músculos que regulan la conformación del tracto vocal [24], y algunos rasgos representativos son: los fonemas prolongados, o repetidos, las vocales distorsionadas y las consonantes imprecisas, con el presente trabajo se pretende contribuir a la estimación de esta alteración.

1.3.2 Disartria.

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 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.

1.4 Antecedentes.

1.4.1 Detectores de sílabas.

Ubicar las sílabas dentro de la grabación constituye el primer paso con vistas a efectuar el cálculo de los indicadores. En [25] se hace una propuesta de mejora al detector de sílabas basado en una máquina de estado de [26], logrando superar su desempeño. En [12] donde se describe un detector de sílabas, basado en la ubicación de máximos locales que cumplen con determinadas condiciones y desechando los que no las cumplen. Los máximos que pasan la prueba son considerados centros de sílabas y a partir de estos se ubican las posiciones de los segmentos de interés dentro las grabaciones, a este detector de sílabas en [11] se le realizan algunas modificaciones.

1.4.2 Indicadores en el dominio del tiempo.

1.4.2.1 Indicador de sonoridad de la sílaba (*ISS*).

El Indicador de Sonoridad de la Sílabas (*ISS*) 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 (2-14.3ms, 70-500 Hz). La función de autocorrelación Figura 4 se define como la correlación cruzada de la señal consigo misma. Frecuentemente las autocorrelaciones (R) con un desfase (j), se calcula para una señal x_n centrada alrededor del cero, es decir con un valor principal de cero. En ese caso la definición de la autocorrelación viene dada por:

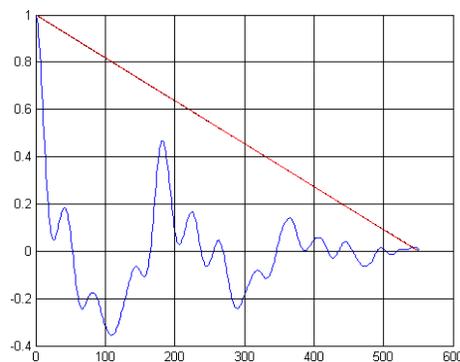


Fig. 4. Representación gráfica de la función de autocorrelación utilizando una señal de voz.

$$R(j) = \sum_n x_n x_{n-j} \quad (4)$$

1.4.2.2 Indicador de energía de la sílaba (*IES*).

La variación de energía en la señal de voz se debe a la variación de la presión subglotal y de la forma del tracto vocal. La energía es útil para distinguir segmentos sordos y sonoros en la señal de voz, debido a que los valores de esta característica aumentan en los sonidos sonoros respecto a los sordos. Este indicador en un inicio tratado como Imprecisión de la Consonante. [12].

Este indicador tuvo función patológica como consonantes imprecisas (*IC*) en el ejercicio de la repetición de las sílabas “*Pa*”, “*Ta*” y “*Ka*”, donde se suponía que *IC* tienda a cero para los pacientes normales y a uno para los patológicos.

En trabajos posteriores [9] [10] [11] [27-30], se usaron distintos valores para tiempo de guarda (T_r) y tiempo de análisis (T_a) se denominó Indicador de Energía de las Sílabas (*IES*), al igual que en este trabajo.

1.4.2.3 Rasgo de Densidad de Cruces por Cero.

Para determinar el indicador Densidad de Cruces por Cero (*DCC*) [28-30] de cada sílaba, Figura 5, se utilizan los mismos segmentos de T_r y T_a descritos en el epígrafe anterior. El *DCC* (5) se define como el cociente entre la Cantidad de Cruces por Cero ($CCC(n)$) de la señal en el segmento de T_a , correspondiente a la consonante y la *CCC* del segmento de T_a de la vocal.

$$DCC = \frac{CCC(Consonante)}{CCC(Vocal)} = \frac{\sum_{T_{on}-T_r}^{T_{on}-T_r} CCC(n)}{\sum_{T_{on}+T_r}^{T_{on}+T_r} CCC(n)} \quad (5)$$

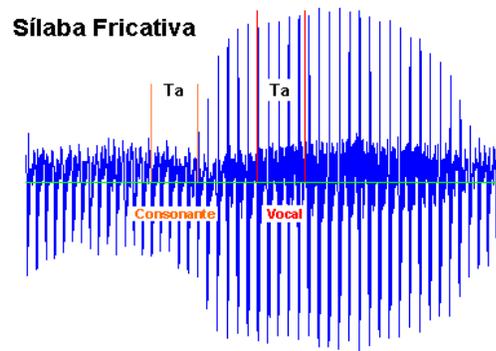


Fig. 5. Representación de una sílaba fricativa, con sus respectivos segmentos de consonante y vocal.

1.4.3 Estudios ISS vs. IES.

Los indicadores *ISS* y *IES* se han utilizado en varios trabajos [9] [10] [27] con el objetivo de separar las diferentes imprecisiones presentes en el ejercicio de diagnóstico de disartrias establecido por las Clínicas Mayo. En [27] se plantea la hipótesis de que las consonantes oclusivas sordas (“normales” en el caso del ejercicio del *Pa-Ta-Ka*) y sus tres sustituciones posibles (sonorización, fricación y nasalización), en la Figura 6, están distribuidas en el plano *IES* vs. *ISS* de la manera siguiente:

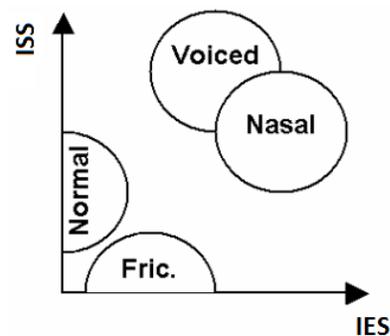


Fig. 6. Distribución hipotética de las consonantes explosivas y sus sustituciones más frecuentes en el plano *IES* vs. *ISS*.

Esta hipótesis se basa en las características fonéticas asumidas para cada uno de los grupos de consonantes, aunque los valores reales de los “centroides” no se evaluaron en dicho trabajo. En [9] se trazó el objetivo de determinar las regiones típicas de los diferentes tipos de consonantes, determinadas por sus valores medios y desviaciones estándar. Los resultados obtenidos se muestran en la Figura 7.

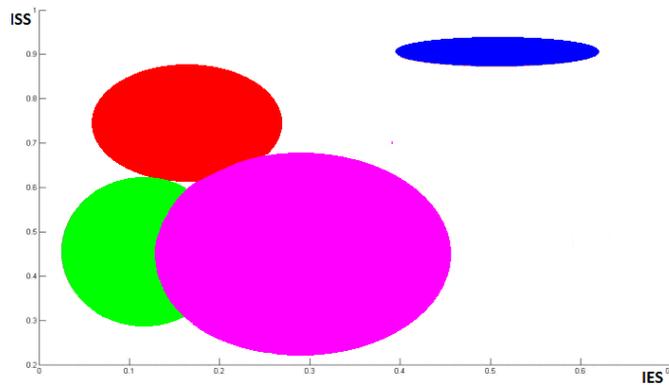


Fig. 7. Regiones de agrupación de las consonantes en el plano IES vs. ISS , donde: fricativas - magenta, explosivas sordas - verde, explosivas sonoras - rojo, nasales - azul.

En el artículo [9] al igual que en este, se llevó a cabo un análisis del desempeño de los indicadores IES vs ISS en la caracterización de las consonantes explosivas sordas y sus sustituciones usuales a partir del plano Sonoridad vs. Energía. Se ha planteado que el mayor problema se encuentra en el solapamiento entre las fricativas (F , S , J) y las explosivas sordas (P , T , K).

Con el objetivo de separar los cuatro grupos de consonantes bajo estudio, explosivas sonoras (B , D , G), explosivas sordas (P , T , K), fricativas (F , S , J) y nasales (M , N), en el plano IES vs. ISS en [10] se utilizaron varios clasificadores estos fueron: Sistemas de Interferencia Borroso a partir de funciones de pertenencia gaussianas (SI Gauss) y a partir de funciones de pertenencia trapezoidales (SI Trapez); Análisis Discriminante Cuadrático (QDA); Clasificador de Naive Bayes (NB); k vecinos más cercanos (KNN); Árboles de Clasificación ($TREE$). En la Tabla 2 aparecen los resultados para cada tipo de clasificador, según [10].

Tabla 2. Resultados de los diferentes clasificadores usados en [10].

	SI Gauss	SI Trapez	KNN	QDA	TREE	NB
TCT	62.0%	64.1%	67.7%	60.9%	68.6%	61.7%
Sens FSJ	36.6%	72.6%	47.9%	43.3%	54.4%	17.3%
Espec FSJ	92.3%	78.1%	93.1%	91.3%	90.2%	97.6%
Sens PTK	66.5%	61.2%	74.2%	55.5%	70.9%	74.4%
Espec PTK	78.1%	83.5%	79.6%	84.4%	82.1%	71.4%
Sens BDG	67.8%	48.7%	58.2%	65.4%	68.9%	59.3%
Espec BDG	80.9%	94.3%	91.4%	82.6%	87.5%	86.9%
Sens MN	74.7%	83.4%	93.4%	86.9%	81.3%	93.7%
Espec MN	95.3%	95.5%	91.1%	88.6%	96.5%	89.9%

1.4.4 Estudios IES vs. DCCM.

Los Indicadores *DCC* y *IES* fueron utilizados en varios trabajos [28-30] con el fin clasificar las consonantes en monosílabos y determinar las consonantes fricativas. En [11] se le introdujo modificaciones en el Indicador *DCC*, con el objetivo de reducir la inconveniente de la deriva de la línea base, que se produce debido a artefactos en el proceso de grabación lo cual causa que en el segmento de la consonante no sea representativo o incluso equivalente a cero.

En la Figura 8, se muestran elipses descriptivas de las regiones de cada grupo de consonantes, de manera similar a como se representaron en el plano *IES vs. ISS*. El centro de estas se ubican en las medias del *DCCM* y el *IES* por clase y el radio en cada dirección se corresponde con la desviación estándar del *DCCM* y del *IES*, según corresponda. En esta figura se observa que con respecto a las explosivas sordas el mayor solapamiento lo presentan las explosivas sonoras, las nasales quedan bien separadas y las fricativas aunque se solapan en una pequeña zona del plano *DCCM vs. IES*, este solapamiento no es representativo si se compara con el área que estas últimas ocupan.

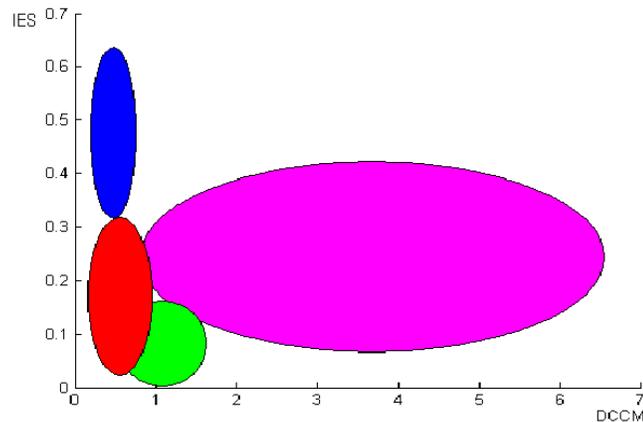


Fig. 8. Regiones de agrupación de las consonantes en DCCM vs. IES, donde: fricativas - magenta, explosivas sordas - verde, explosivas sonoras - rojo, nasales – azul.

En [11] su objetivo consistió originalmente en la separación de las consonantes explosivas sordas del resto, haciendo énfasis en la separación de las fricativas, que había resultado el mayor problema en trabajos anteriores. A este fin se le dedicaron varios análisis de separación de estos dos grupos de interés: fricativas-resto, sordas-resto y sordas-fricativas, unos de los análisis fue separación mediante fronteras de clasificación, Ver en [11].

Para evaluar los resultados de la separación de los cuatro grupos se utilizaron los clasificadores como: Búsqueda Heurística en Rejilla (*BHR*), Análisis Discriminante a partir de la Distancia de Mahalanobis (*AD-DM*), Análisis Discriminante con matriz de covarianza diagonal (*AD-MCD*) y mediante método de los k vecinos más cercanos (*KNN*). En la Tabla 3 aparecen los resultados para cada tipo de clasificador, según [11].

Tabla 3. Resultados de los diferentes clasificadores usados en [11].

	BHR	AD-DM	AD-MCD	KNN
TCT	76.3%	75.2%	73.8%	78.0%
Sens FSJ	85.8%	76.3%	60.5%	69.8%
Espec FSJ	83.7%	90.6%	97.7%	95.3%
Sens PTK	77.8%	71.9%	82.7%	81.5%
Espec PTK	91.2%	93.8%	88.2%	88.9%
Sens BDG	62.5%	72.3%	73.4%	76.7%
Espec BDG	95.9%	91.2%	88.6%	90.2%
Sens MN	82.3%	85.5%	78.1%	87.0%
Espec MN	96.9%	91.4%	90.3%	95.5%
Tasa de Clasificación Total (TCT)=	71%			

CAPÍTULO 2. MATERIALES Y MÉTODOS.

En este capítulo se describe el detector de sílabas mediante el cual se localizan las consonantes dentro de las grabaciones procesadas. Se definen los indicadores ISS, IES y DCCM, que de conjunto se usarán en la separación de las consonantes. Se detallan los índices a partir de los cuales se cuantifica la robustez de los indicadores en la separación de los diferentes conjuntos de consonantes y se exponen los métodos de clasificación implementados.

2.1 Bases de Datos.

Para la determinación de los rasgos o indicadores, que definirán las supuestas regiones donde se agrupan los 4 grupos de consonantes; al igual que en [10]-[11] se contó con la locución de sujetos sanos entre 20 y 35 años de edad de ambos sexos. Debido a que existen algunas grabaciones con 8 bit éstas se desecharon usando solo las que tienen 16 bit de resolución que constipen la mayoría. Como se describe en los epígrafes anteriores del capítulo 1 *_ver* (Tabla 2 y Tabla 3), la TCT en [11] es superior a la obtenida en [10]. Con el fin de verificar la equivalencia de la base de datos se reproduce en este trabajo los experimentos anteriores: *ISS vs. IES* y *DCCM vs. IES*.

2.1.1 Grabaciones utilizadas como material de estudio.

Cada grabación consiste en la repetición sucesiva de una misma consonante durante 5 segundos. Se efectuaron grabaciones de 11 señales diferentes que contienen las sílabas del ejercicio original: ``Pa``, ``Ta`` y ``Ka`` que pertenecen a las consonantes sordas, así como sus desviaciones más comunes, es decir, ``Ba``, ``Da`` y ``Ga``, sonoras; ``Fa``, ``Sa`` y ``Ja``, fricativas; y las nasales: ``Ma`` y ``Na`` (Tabla. 4).

En cada grabación se detectaron las sílabas existentes y a cada una se le calculó el *ISS*, *IES* y la *DCCM* en conjunto, se analizaron en total 3182 sílabas, de estas 815 fricativas, 869 explosivas sonoras, 1023 explosivas sordas y 475 nasales. Los resultados se agruparon como vectores que contienen los valores *ISS*, *IES* y *DCCM* según corresponda a cada tipo de consonantes, para luego buscar la mejor forma de separarlas. Se analizaron estos tipos de consonantes, ya que son las sustituciones más comunes en pacientes con trastornos motores del habla.

Tabla 4. Grupos de consonantes analizadas.

Grupos de Consonantes	Labial	Dental	Palatal
Explosivas Sordas	P	T	K
Explosivas Sonoras	B	D	G
Fricativas	F	S	J
Nasales	M	N	-

2.2 Detector de sílabas.

Para el cumplimiento del algoritmo se empleó el software matemático *Matlab* (versión 7.8.0 R2009a), que ofrece un entorno de desarrollo mejorado, es un programa que se utiliza para el cálculo de vectores y matrices, disponible para *Windows*. El primer paso, antes del cálculo de los rasgos es el localizar las sílabas dentro de cada grabación; así como también la transición de consonante a vocal. Con este fin se empleó el detector de sílabas reportado en un principio en [12] y modificado en [11] descrito sintéticamente a continuación en el epígrafe 2.2.1 las cuales se emplearon en este trabajo.

El detector se basa en la detección de máximos en la envolvente de la señal de energía, primero en [11], se suprimió el nivel *DC* de la señal de voz, después se le halló el valor absoluto, para posteriormente aplicarle un filtro pasobajo de *Butterworth IIR*, con frecuencia de corte de 10 Hz, de orden 3, destinado eliminar los componentes de alta frecuencia y así suavizar la envolvente.

2.2.1 Detector de centros de sílabas.

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 son las siguientes:

- Los máximos tienen que tener una amplitud mayor que determinado umbral. Se tomó como umbral el 70% de la media de la envolvente de energía (6).

$$Umbral = 0.7 * mean(Envolvente_de_Energía) \quad (6)$$

- Los máximos deben estar separados un tiempo moderado. La separación mínima aceptada entre máximos se fijó a 100 milisegundos (cuando dos máximos están más cercanos que este valor el que posee menor amplitud no se tomara en cuenta).
- Entre dos máximos de diferentes sílabas, debe haber un mínimo considerable. Se usó como condición, que cada máximo para ser considerado centro de sílaba tendrá la presencia de un mínimo de menos del 75 % de su máxima amplitud entre él y los máximos posteriores y anteriores.
- La sílaba debe tener una duración determinada. La separación entre el mínimo más pronunciado anterior y el posterior a la sílaba, debe ser suficiente para que esta sea percibida. Esta separación mínima se fijó a 50 milisegundos.

Una vez que las sílabas son detectadas, el instante de liberación o de transición de consonante hacia la vocal 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.

2.3 Indicadores en el dominio del tiempo.

2.3.1 Indicador de sonoridad de la sílaba (ISS).

La definición se describe en el epígrafe I.4.2.1. A este indicador se le hace una corrección en [9] y se empleó en [10] al igual que en este trabajo. La modificación desagravia la reducción del pico de autocorrelación causada por el tamaño finito del segmento de análisis. En la Figura. 9 se muestran los elementos involucrados en el cálculo, tanto del ISS

original, como la corrección y la fórmula se brinda en (7) y (8), donde WL es el tamaño de la ventana de análisis, 30 milisegundos. La línea roja representa el valor máximo que puede alcanzar la función de autocorrelación. Se tuvo en cuenta el mismo que [10] tiempo de guarda final (Tr), para el cálculo del ISS de 16ms y el valor de tiempo de análisis final (Ta), de 25ms.

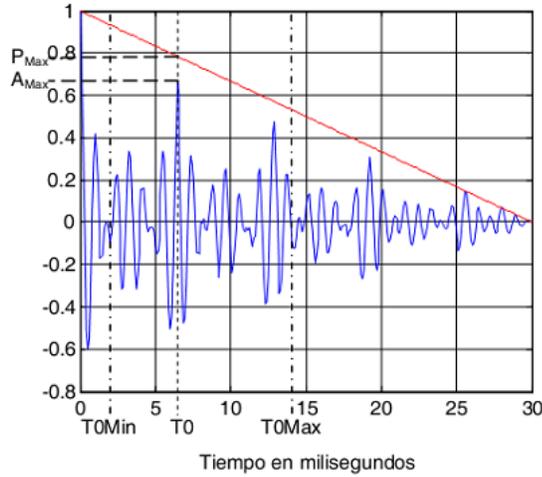


Fig. 9. Representación gráfica de los elementos usados para el cálculo de ISS.

$$ISS_{Orig} = A_{Máx} \quad (7)$$

$$ISS = \frac{A_{Máx}}{P_{Máx}} = \frac{A_{Máx} W_L}{W_L - T_0} \quad (8)$$

2.3.2 Indicador de energía de la sílaba (IES).

Indicador que se define [10] [11] como la relación entre el área de la envolvente de energía $E(t)$ antes y después de la explosión. Los elementos que se tienen en cuenta para el cálculo (9) de este rasgo se muestran en la Figura 10, en la misma se ilustra A_b y A_a , las áreas bajo la curva envolvente de la amplitud de la señal de voz o energía de la sílaba. El instante de liberación de la consonantes, o lo que es lo mismo, el inicio de la vocal se denota como T_{on} . Los intervalos de tiempo de guarda T_r durante el cual se asume que está ocurriendo la transición y se supone que no aporta información útil. El tiempo de análisis T_a se precisa

como el segmento temporal donde se encuentran los fragmentos de menor y mayor área bajo la curva.

Para los valores de estos elementos se consideró el mismo criterio que en [11] de (20 ms) cada uno, debido a que con ello se obtuvieron muy buenos resultados.

$$IES = \frac{A_b}{A_a} = \frac{\sum_{T_{on}-T_r}^{T_{on}-T_r-T_a} E(t)}{\sum_{T_{on}+T_r}^{T_{on}+T_r+T_a} E(t)} \quad (9)$$

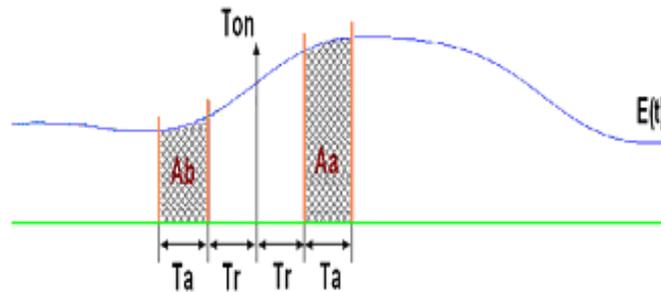


Fig. 10. Representación gráfica de los elementos y los segmentos temporales empleados en el cálculo del IES.

2.3.3 Indicador de Densidad de Cruces por Cero Modificado (DCCM).

La modificación se realiza, debido a que en las señales de audio en ocasiones no deseadas se introducen artefactos de manera momentánea con el proceso de grabación. Por lo que ocasiona que la señal de voz no se cruce con cero en el segmento temporal el cual contiene la porción referido a la consonante, este efecto se conoce como la deriva de la línea base como se ilustra en la Figura 11. Al calcular DCC como bien se describe en el epígrafe I.4.2.3, en estas sílabas el resultado no sería representativo, ya que es igual a cero.

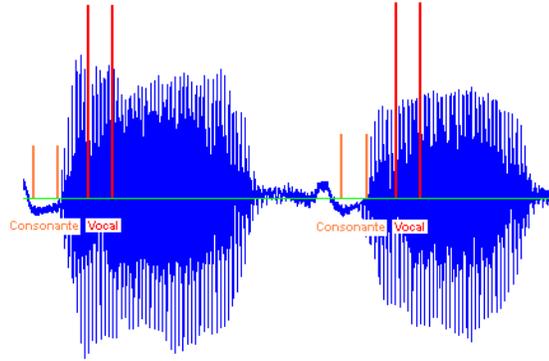


Fig. 11. En las sílabas que ocurra este problema, la determinación de los cruces por cero no aportará resultados efectivos y la clasificación de estas sílabas pudiera ser incorrecta.

Con el fin de suprimir este inconveniente se ideó ingeniosamente en [11] y donde se expone que: el indicador Densidad de Cruces por Cero Modificado *DCCM* no se calcula la Cantidad de Cruces por Cero (*CCC*) en el segmento temporal de la consonante, sino que se fragmenta el mismo, en cinco tramas como se muestra en Figura 12. Se calcula en cada trama los cruces por su media y finalmente se suman los cruces de las cinco tramas, siendo este valor Cantidad de Cruces por la Media de las Tramas (*CCMT*), el utilizado para la implementación del indicador *DCCM* (10).

$$DCCM = \frac{CCMT(Consonante)}{CCC(Vocal)} \quad (10)$$

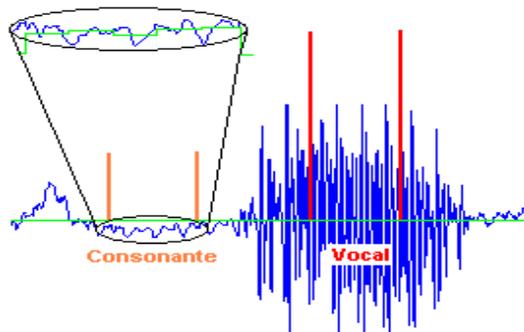


Fig. 12. Representación de la implementación del indicador *DCCM*.

2.4 Métodos de clasificación empleados e índices de cuantificación para comprobar los resultados.

A continuación se exponen el funcionamiento de los cuatros clasificadores que se utilizaron y los términos para valorar su efectividad.

2.4.1 Clasificación mediante el método de Árbol de decisión.

Este modelo de clasificación que se utilizó en [10] es un algoritmo de aprendizaje inductivo, es decir, busca establecer leyes o principios generales sobre la base de la observación de varios o todos los componentes de un conjunto o clase. Aplica la estrategia “*divide y vencerás*” para hacer la clasificación de los objetos y asociarlos a una clase, tomando en cuenta los valores de los atributos de los objetos. Un árbol de decisión toma como entradas objetos o situaciones que se caracterizan mediante un conjunto de propiedades. El árbol proporciona como salida una “*decisión*” que puede ser si o no, para las funciones booleanas, aunque también es posible representar funciones con un mayor rango de salidas como es el caso que nos concierne.

La idea básica es de determinar, para un conjunto de ejemplos dados, el atributo más importante, o sea, aquel que posea el mayor poder discriminatorio para dicho conjunto; éste atributo es usado para la clasificación de la lista de objetos, basados en los valores asociados con él mismo. Después de haber hecho la primera prueba de atributo, esta arrojará un resultado, el cual es en sí mismo un nuevo problema de aprendizaje de árbol de decisión, con la diferencia que contará con menos ejemplos y un atributo menos, por lo que, cada atributo que se selecciona se descarta para la siguiente prueba.

Este modelo cuando se aplica mediante el software matemático *Matlab*, el mismo te provee de dos métodos que se suelen denominar árbol de clasificación o árbol de regresión. La diferencia se encuentra en que el árbol de clasificación se activa por defecto si las clases se etiquetan con texto o una variable categórica, mientras que el árbol de regresión se activa por defecto si se etiqueta con números como se muestra en la Figura 13. En cualquiera de estas estructuras las hojas del árbol representan clases mientras que las ramas son conjunciones de rasgos que resultan en esas clases.

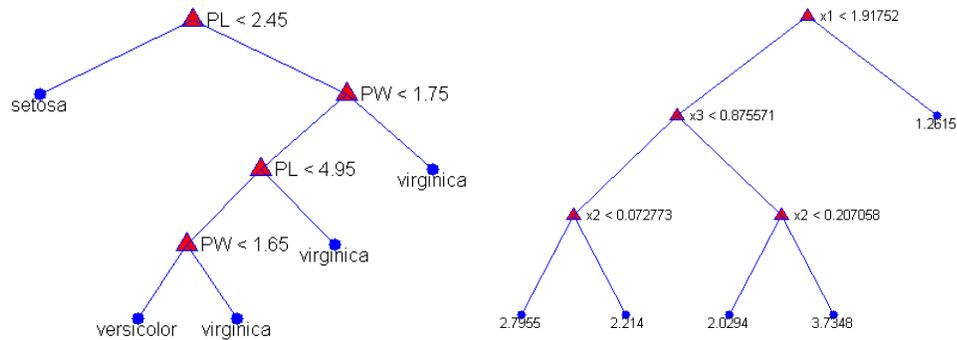


Fig. 13. Representación de izquierda a derecha ejemplos de Árbol de clasificación y Árbol de regresión ambos con 4 niveles (ramas).

Para que se obtengan resultados eficientes con el uso de Árbol de decisión en un problema en cuestión, dichos problemas deben satisfacer las siguientes características:

- La descripción de los objetos debe efectuarse en términos de parejas, el atributo y su valor.
- Cada salida debe ser distinta, es decir, los nodos terminales deben contener una respuesta única.
- Debe haber alternativas entre 2 o más cosas, para hacer posible la clasificación.
- El entrenamiento debe estar libre de errores, o sea, el problema a clasificar debe contener la información correcta, para que el resultado sea eficiente.
- El entrenamiento debe contener información suficiente para poder hacer la clasificación.

2.4.2 Análisis Discriminante.

El Análisis Discriminante, [31] es una técnica estadística, enfocada a identificar las características que diferencian (discriminan) a dos o más grupos, y crear una función capaz de distinguir los miembros de uno u otro grupo. La pertenencia a los grupos, conocida de antemano, se utiliza como variable dependiente (una variable categórica con tantos valores como grupos). Las variables que se suponen diferencian a los grupos se utilizan como variables dependiente o variables de clasificación (también llamadas variables discriminantes). El objetivo del análisis discriminante, es encontrar la combinación lineal de variables independientes que mejor permiten diferenciar (discriminar) a los grupos. Una

vez encontrada esta combinación (la función discriminante), podrá ser utilizada para clasificar nuevos casos; se trata de una técnica de análisis multivariante, que es capaz de aprovechar las relaciones existentes entre una gran cantidad de variables independientes, para maximizar la capacidad de discriminación. El análisis discriminante es conceptualmente muy similar al análisis de varianza multivariante de un factor.

El análisis discriminante, permite diferenciar entre cualquier número de grupos. A continuación se explica el caso de dos grupos, en la Figura 14, están representadas, en el espacio bivariable X_1 y X_2 , las nubes de puntos de puntos correspondiente a dos grupos hipotéticos. Los dos grupos representados se diferencian entre sí en ambas variables, pero no por completo, pues, se solapan en una pequeña región situada entre ambos. También está representada la función D , que es una combinación lineal de ambas variables. Sobre la función D se representa la proyección de las dos nubes de puntos, en forma de histograma, como si la función D cortara a las dos nubes de puntos en dirección de su eje. Las dos líneas punteadas de cada uno de los histogramas representan la ubicación proyectada de los puntos medios de cada grupo (centroides).

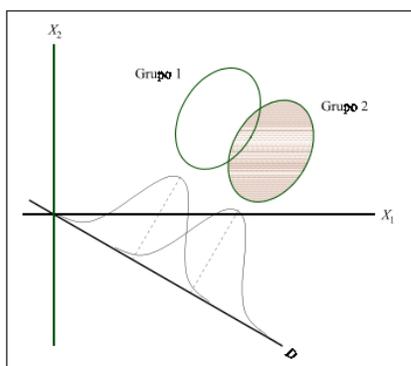


Fig. 14. Diagramas de dispersión de dos grupos en dos variables de clasificación.

El propósito del análisis discriminante, consiste en aprovechar la información contenida en las variables independientes, para crear una función D combinación lineal (puede no ser lineal) de X_1 y X_2 capaz de diferenciar lo más posibles a ambos grupos. La función discriminante es de la forma:

$$D = b_1 X_1 + b_2 X_2 \quad (11)$$

Donde b_1 y b_2 son las ponderaciones de las variables independientes que consiguen hacer que los sujetos de uno de los grupos obtengan puntuaciones máximas en D y los sujetos del otro grupo puntuaciones mínimas.

Los grupos deben diferenciarse con anterioridad en las variables independientes, combinando esta, el análisis busca diferenciar los grupos al máximo; pero si los grupos no se diferencian en las variables independientes, el análisis será infructuoso: no podrá encontrar una dimensión en la que los grupos difieran. Dicho de otro modo, si el solapamiento entre los grupos es excesivo, los centroides se encontrarán en la misma o parecida posición en el espacio p -dimensional y en esas condiciones no será posible encontrar una función discriminante útil para la clasificación.

El software matemático *Matlab versión 7.8.0 R2009a*. permite implementar varios tipos de análisis discriminantes; el lineal, el lineal con matriz de covarianza diagonal, el cuadrático, el cuadrático con matriz de covarianza diagonal y el análisis discriminante a partir de distancias de Mahalanobis.

2.4.3 Método de los k vecinos más cercanos (KNN).

El método de los k vecinos más cercanos descritos en [31] [32], es un método de clasificación supervisada, (Aprendizaje, estimación basada en un conjunto de entrenamiento y prototipos) que sirve para estimar la función de densidad $F(x/C_j)$ de las predictoras x por cada clase C_j .

Es un método de clasificación no paramétrico, que estima el valor de la función de densidad de probabilidad de pertenencia de un elemento x a la clase C_j a partir de la información proporcionada por el conjunto de prototipos. En el proceso de aprendizaje no se hace ninguna suposición acerca de las variables predictoras.

Los ejemplos de entrenamiento son vectores en un espacio característico multidimensional, cada ejemplo está descrito en términos de p atributos considerando q clases para la clasificación. Los valores de los atributos del i -ésimo ejemplo (donde $1 \leq i \leq n$) se representan por el vector p -dimensional (12).

$$x_i = (x_{1i}, x_{2i}, \dots, x_{pi}) \in X \quad (12)$$

El espacio es particionado en regiones por localizaciones y etiquetas de los ejemplos de entrenamiento. Un punto en el espacio es asignado a la clase C si esta es la clase más frecuente entre los k ejemplos de entrenamiento más cercanos. Generalmente se usa la distancia euclidiana (13).

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^p (x_{ir} - x_{jr})^2} \quad (13)$$

La fase de entrenamiento del algoritmo consiste en almacenar los vectores característicos y las etiquetas de las clases de los ejemplos de entrenamiento. En la fase de clasificación, la evaluación del ejemplo (del que no se conoce su clase) es representada por un vector en el espacio característico. Se calcula la distancia entre los vectores almacenados y el nuevo vector y se seleccionan los k ejemplos más cercanos. El nuevo ejemplo es clasificado con la clase que más se repite en los vectores seleccionados.

Este método supone que los vecinos más cercanos nos dan la mejor clasificación y esto se hace utilizando todos los atributos; el problema de dicha suposición es que es posible que se tengan muchos atributos irrelevantes que dominen sobre la clasificación: dos atributos relevantes perderían peso entre otros veinte irrelevantes.

Para corregir el posible sesgo se puede asignar un peso a las distancias de cada atributo, dándole así mayor importancia a los atributos más relevantes. Otra posibilidad consiste en tratar de determinar o ajustar los pesos con ejemplos conocidos de entrenamiento. Finalmente, antes de asignar pesos es recomendable identificar y eliminar los atributos que se consideran irrelevantes. El método k -nn consta de dos algoritmos: el de entrenamiento y el de clasificación.

La mejor elección de k depende fundamentalmente de los datos; generalmente, valores grandes de k reducen el efecto de ruido en la clasificación, pero crean límites entre clases parecidas. Un buen k puede ser seleccionado mediante una optimización de uso. El caso especial en que la clase es predicha para ser la clase más cercana al ejemplo de entrenamiento (cuando $k=1$) es llamada *Nearest Neighbor Algorithm*, (algoritmo del vecino más cercano).

La exactitud de este algoritmo puede ser severamente degradada por la presencia de ruido o características irrelevantes, o si las escalas de características no son consistentes con lo que uno considera importante. Muchas investigaciones y esfuerzos fueron puestos en la selección y crecimiento de características para mejorar las clasificaciones. Particularmente una aproximación en el uso de algoritmos que evolucionan para optimizar características de escalabilidad. Otra aproximación consiste en escalar características por la información mutua de los datos de entrenamiento con las clases de entrenamiento.

2.4.4 Índices de cuantificación de los resultados.

Los resultados obtenidos, pueden ser valorados en términos de tasa de clasificación, tasa de error, sensibilidad y especificidad; estos son definidos como las proporciones de sujetos en cada subclase, correcta o incorrectamente clasificados por la prueba, donde:

Verdadero Positivo (*VP*): Sílabo correctamente clasificada dentro de un tipo de consonante.

Falso Negativo (*FN*): Sílabo que por error fue excluida del tipo de consonante bajo análisis, y sin embargo pertenecía a dicho grupo.

Verdadero Negativo (*VN*): Sílabo correctamente descartada del subconjunto en investigación.

Falso Positivo (*FP*): Sílabo no perteneciente a una clase pero que fue enumerada dentro de esta.

La tasa de clasificación (*TC*) es la fracción de eventos correctamente detectados con respecto al total de eventos.

$$Tasa_de_Clasificación(TC) = \frac{VP + VN}{VP + VN + FP + FN} \quad (14)$$

La tasa de error (*TE*) es la fracción de eventos incorrectamente detectados con respecto al total de eventos.

$$Tasa_de_Error(TE) = \frac{FP + FN}{VP + VN + FP + FN} \quad (15)$$

La sensibilidad (*Sens*) se define como la probabilidad del algoritmo de detectar correctamente el evento positivo.

$$Sensibilidad = \frac{VP}{VP + FN} \quad (16)$$

La especificidad (*Espec*) se define como la probabilidad del proceder para detectar los elementos de la clase complementaria (eventos negativos) como tales.

$$Especificidad = \frac{VN}{VN + FP} \quad (17)$$

Los índices anteriores serán usados para cuantificar los resultados en la clasificación de cada grupo de consonante de forma independiente.

Para evaluar la efectividad de cada método de clasificación de forma global, se deben tener en cuenta los resultados en la separación de cada tipo de evento (grupo de consonantes). Teniendo en cuenta que estamos en presencia de un problema multiclase (4 grupos de consonantes), se utilizó la tasa de clasificación total (*TCT*).

$$TCT = \frac{VP_Eventos_1 + VP_Eventos_2 + VP_Eventos_n}{Cantidad_Total_de_Eventos_Analizados} \quad (18)$$

La matriz de confusión muestra la cantidad de cada tipo de consonante que fue clasificada correctamente, así como el número de estas que fue apuntada erróneamente en cada uno de los demás grupos. Las filas representan el total de consonante de cada tipo que fue procesada, mientras las columnas revelan la cantidad que fue clasificada dentro de cada grupo de consonantes.

CAPÍTULO 3. RESULTADOS DE LA CLASIFICACIÓN.

En el capítulo se muestran los histogramas de cada grupo de consonante para los tres indicadores usados. Se exhiben los resultados del desempeño al implementar los indicadores mediante las combinaciones y en conjunto, en la separación de los grupos de consonantes. Así como los resultados de emplear los 4 tipos de clasificadores seleccionados para la clasificación.

3.1 Distribución de consonantes usando indicadores de la señal de voz.

3.1.1 Distribución mediante la sonoridad de las sílabas (ISS).

Teniendo en cuenta la distribución de cada tipo de consonante se estima, algunos rasgos de semejanzas entre las explosivas sonoras y las nasales (acústicamente, las consonantes nasales son sonorantes es decir, no restringen la salida de aire, y desde el punto de vista lingüístico casi siempre son sonoras). Por lo tanto como se puede interpretar en la Figura 15, se cumple que las explosivas sonoras presentan *ISS* alto y las nasales se pueden considerar con un *ISS* muy alto. (Tabla 5) (se generan por la vibración de las cuerdas vocales manteniendo la glotis abierta, lo que permite que el aire fluya a través de ellas). De manera diferente se comportan las explosivas sordas y las fricativas, por lo que se corrobora que las explosivas sordas tengan un *ISS* medio y las fricativas en su mayoría se consideren con un *ISS* bajo (durante su ejecución no se generan vibraciones en las cuerdas vocales). A pesar de las diferencias que presentan los grupos analizados, el *ISS* no es suficiente separarlos, el nivel de solapamiento es elevado entre (explosivas sonoras y nasales) y (explosivas sordas y fricativas).

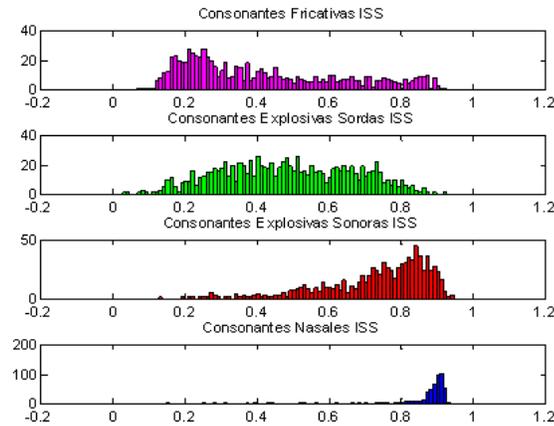


Fig. 15. Histograma ISS para cada tipo de consonante.

3.1.2 Distribución mediante la energía de las sílabas (*IES*).

Como bien se muestra en la Figura 16, al igual que en [9] [10] [11] la distribución entre cada grupo de consonantes en *IES* es diferente, aunque no es capaz de separarlos ya que existe un alto nivel de solapamiento (como en *ISS*), con prácticamente la misma distribución en las explosivas sonoras y sordas (estas se originan por el bloqueo completo de la cavidad oral y nasal causando la detención del flujo de aire y posterior liberación en forma de explosión), por lo tanto se consideran bajo el *IES* y en las otras dos clase (fricativas y nasales), una muy elevada dispersión (durante su emisión el aire se escapa pasando por un espacio muy estrecho, las fricativas al juntarse ciertas partes de la boca y las nasales al separarse el velo-la parte carnosa del paladar), como se puede comparar en los histogramas, las fricativas se catalogaron con un *IES* medio y las nasales alto (Tabla 5).

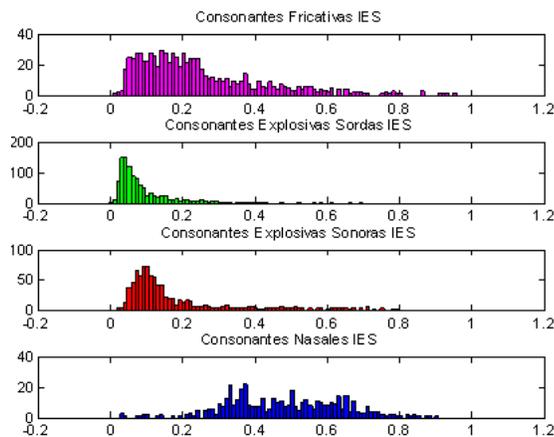


Fig. 16. Histogramas IES para cada tipo de consonante.

3.1.3 Distribución mediante la densidad de Cruces por Cero (DCCM).

En la Figura 17 se muestran los histogramas de DCCM para los cuatro grupos de consonantes. Se aprecia que existe nuevamente un elevado grado de solapamiento entre ellas, por lo que no es posible la separación mediante la determinación de umbrales, aunque sí se aprecian diferencias notables; por ejemplo: las fricativas se pueden considerar con DCCM alto, son prácticamente las únicas que presentan DCCM mayores que 4, aunque tan disperso, que hay una parte considerable por debajo de este valor. En el resto de las consonantes se observa una tendencia ascendente, en cuanto a los valores de DCCM de las nasales a las explosivas sonoras y sordas, en ese orden. (Tabla 5).

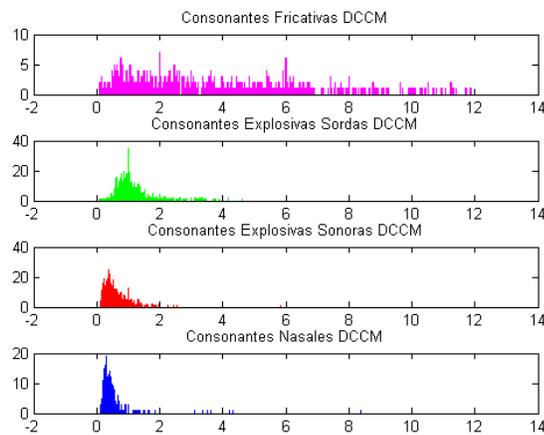


Fig. 17. Histograma DCCM para cada tipo de consonante.

Tabla 5. Caracterización de las consonantes analizadas según los histogramas de ISS, IES y DCCM.

Grupos de Consonantes	ISS	IES	DCCM
Fricativas	Baja (en su mayoría)	Media	Alta (en su mayoría)
Explosivas Sordas	Media	Muy Baja	Relativamente Baja
Explosiva Sonoras	Alta	Baja	Baja
Nasales	Muy Alta	Alta	Muy Baja

3.2 ISS vs. IES.

Al igual que en [9] y [10], se analizaron las sílabas detectadas según sus valores de ISS y IES como se muestra en el plano de coordenadas de la Figura 18, se puede identificar cada grupo de consonante por colores, igual que en la tabla anterior. Se aprecian como se

encuentran solapados y dispersos los puntos que definen a cada grupo de consonante, aunque se observan relativas aglomeraciones que brindan una posible ubicación en el plano. Para tener una mejor idea como en [9] [10] [11], en la Figura 19, se representan las posibles regiones en forma de elipse. Para determinar las elipses se calcula las coordenadas del centro con el promedio de *ISS* y *IES* del grupo de consonante correspondiente, así como los radios vertical y horizontal mediante la desviación estándar de *ISS* y *IES* respectivamente, de cada grupo en cuestión. En la Figura 20, son rectángulos que definen regiones correspondientes a cada clase, cuyos centros se establecen por la mediana de *ISS* y de *IES*. La posición de sus lados está dada por los valores correspondientes al 25 y 75 percentiles propio de cada grupo analizado.

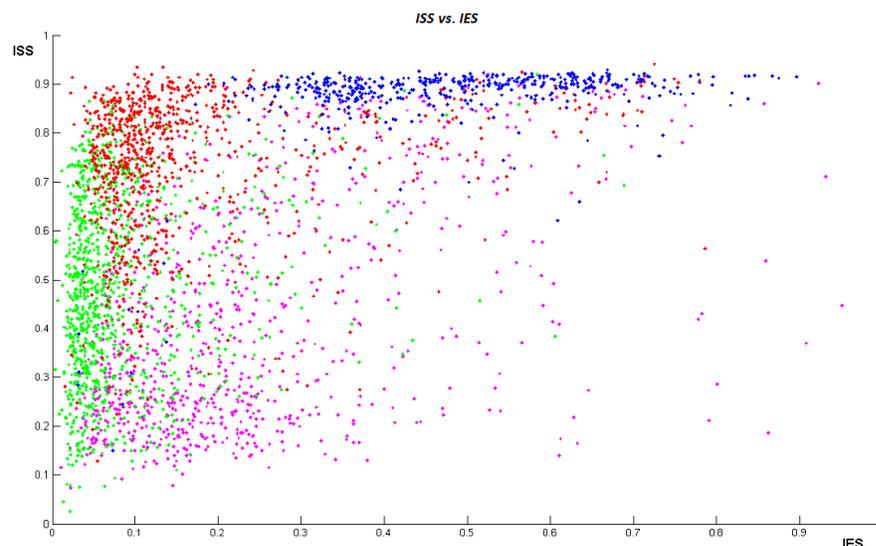


Fig. 18. Ubicación de cada sílaba en el plano coordenado *ISS* vs. *IES*.

En la Figura 19, de las elipses descriptivas, se aprecia cómo se solapan las regiones definidas por las explosivas sonoras, sordas y fricativas, sobre todo entre estas dos últimas. El solapamiento entre las explosivas sordas y las fricativas ha sido el problema principal abordado en trabajos precedentes, en los cuales se realizaron varios análisis de separación con estos dos grupos de sumo interés [11]. Se visualiza como están separadas el grupo de las nasales.

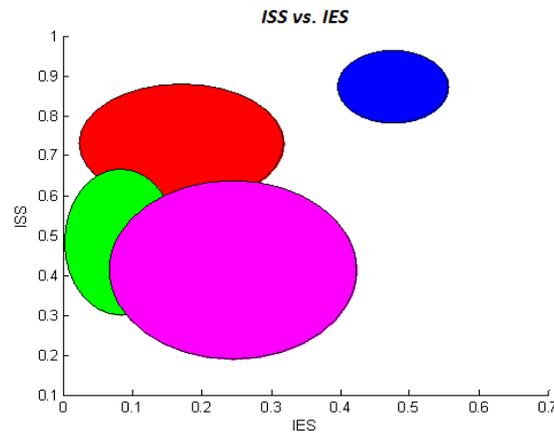


Fig. 19. Regiones de agrupación de las consonantes con los rasgos ISS y IES.

Como se muestra en la siguiente Figura 20, los rectángulos, nos proporcionan una mejor descripción de las posibles regiones en cuestión. Si bien en [10] se aprecia que existe solapamiento entre las fricativas y explosivas de sordas, en esta figura aparecen los 4 grupos de consonantes separadas totalmente. Ocurre debido a lo descrito en el epígrafe 2.1.

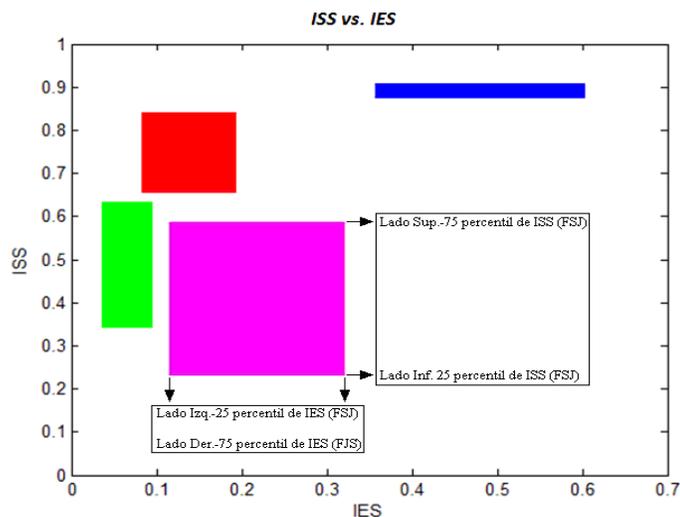


Fig. 20. Distribución de las regiones en el plano ISS vs. IES.

3.2.1 ISS vs. IES en la separación de los cuatro grupos de consonantes.

A continuación se evalúa la capacidad de *ISS* y *IES* en la separación de los cuatro grupos de consonantes mediante el uso de clasificadores.

3.2.1.1 Clasificación de las consonantes mediante análisis discriminante con distancia de Mahalanobis (AD-DM).

Las sílabas separadas por grupos se le pasaron como entrenamiento a procesos de clasificación aplicando análisis discriminante [31] [10], para lo cual se utilizó la función *classify* de *Matlab* versión 7.8.0 R2009a. Las dos variantes a utilizar son: análisis discriminante cuadrático con matriz de covarianza diagonal y análisis discriminante con distancia Mahalanobis.

La distancia de Mahalanobis es una medida que se utiliza a menudo en la separación de conjunto de datos multivariados. En la Figura 21, aparecen las regiones de clasificación al utilizar análisis discriminante a partir de la distancia de Mahalanobis. Como es de esperar la región referente al grupo de las fricativas (magenta) ocupa la mayor parte. El rojo pertenece a las explosivas sonoras, el azul a las nasales y el verde a las explosivas sordas.

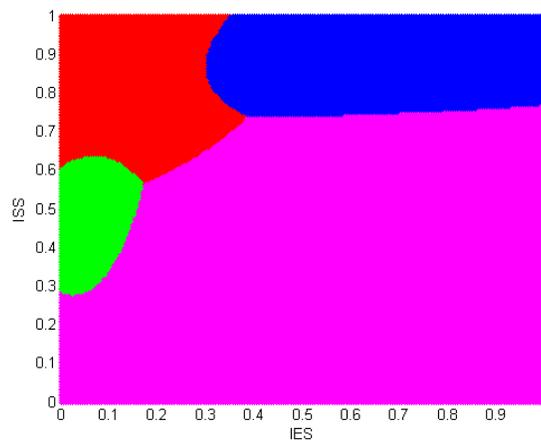


Fig. 21. Regiones de clasificación de los grupos de consonantes a partir del análisis discriminante con distancia de Mahalanobis (AD-DM).

En la Figura 22, se aprecia la ubicación real de las sílabas detectadas separadas por grupos, en conjunto con las regiones predeterminadas a partir del método de clasificación en cuestión. Por otra parte en las Tablas 6 y 7 se muestran la matriz de confusión y los resultados teniendo en cuenta la Tasa de Clasificación, Tasa de Error, así como Sensibilidad y Especificidad.

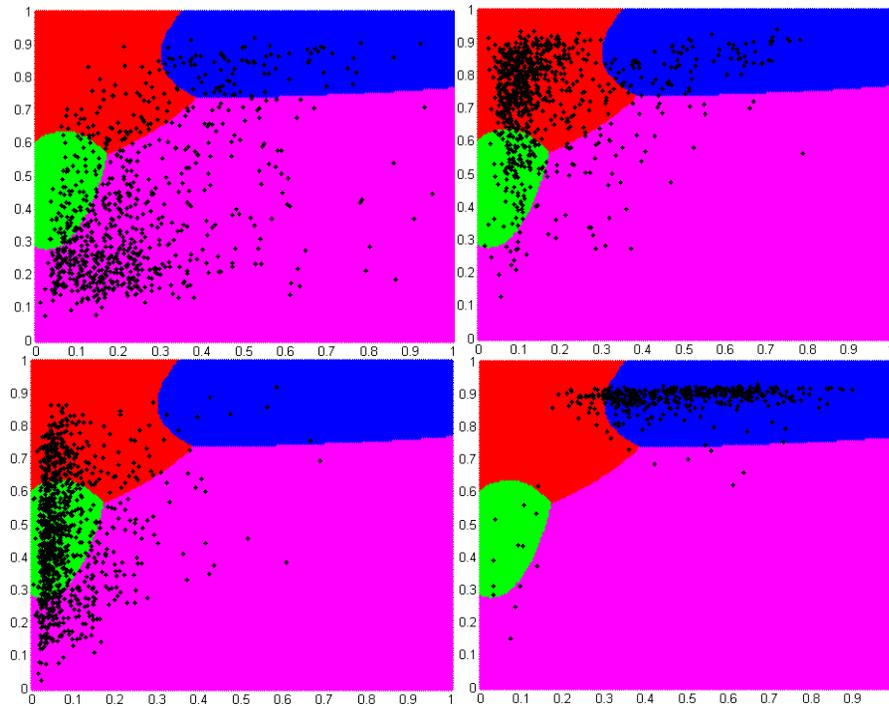


Fig. 22. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el (AD-DM). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 6. Matriz de confusión deriva de la clasificación mediante el análisis discriminante con distancia de Mahalanobis (AD-DM).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	577	81	77	80
Explosivas sordas	262	495	258	8
Explosivas sonoras	74	126	583	86
Nasales	9	8	44	414

Tabla 7. Resultados de los índices de la clasificación mediante el análisis discriminante con distancia de Mahalanobis (*AD-DM*).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	81.68%	18.32%	70.80%	85.42%
Explosivas sordas	76.65%	23.35%	48.38%	90.04%
Explosivas sonoras	79.10%	20.89%	67.09%	83.61%
Nasales	92.61%	7.38%	87.16%	93.57%

Tasa de Clasificación Total (*TCT*)= 65.02 %

3.2.1.2 Clasificación mediante análisis discriminante cuadrático con matriz de covarianza diagonal (*AD-MCD*).

En la siguiente Figura 23 se exponen las regiones de clasificación después de emplear el análisis discriminante cuadrático con matriz de covarianza diagonal. Como se había visto anteriormente, el magenta pertenece a las fricativas, el verdes a las explosivas sordas, el rojo a las explosivas sonoras y el azul a las nasales.

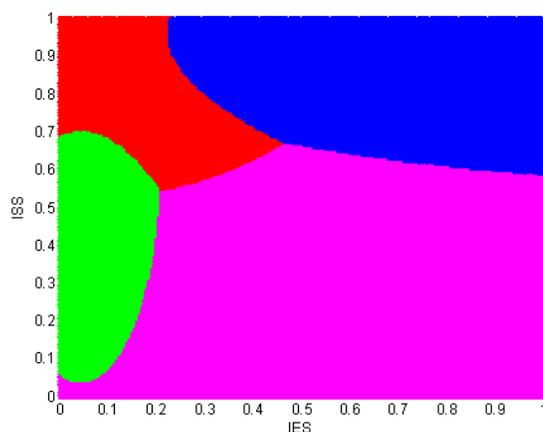


Fig. 23. Representación de las regiones de clasificación de los grupos de consonantes utilizando el análisis discriminante cuadrático con matriz de covarianza diagonal (*AD-MCD*).

Seguido en la Figura 24 se presentan las gráficas con las ubicaciones reales de cada sílaba procesada, separada por grupos, conjuntamente con las regiones que les corresponden luego de la clasificación usando el (*AD-MCD*). Las tablas 8 y 9 muestran la matriz de confusión y los resultados de los índices de clasificación mediante el análisis discriminante con distancia de Mahalanobis.

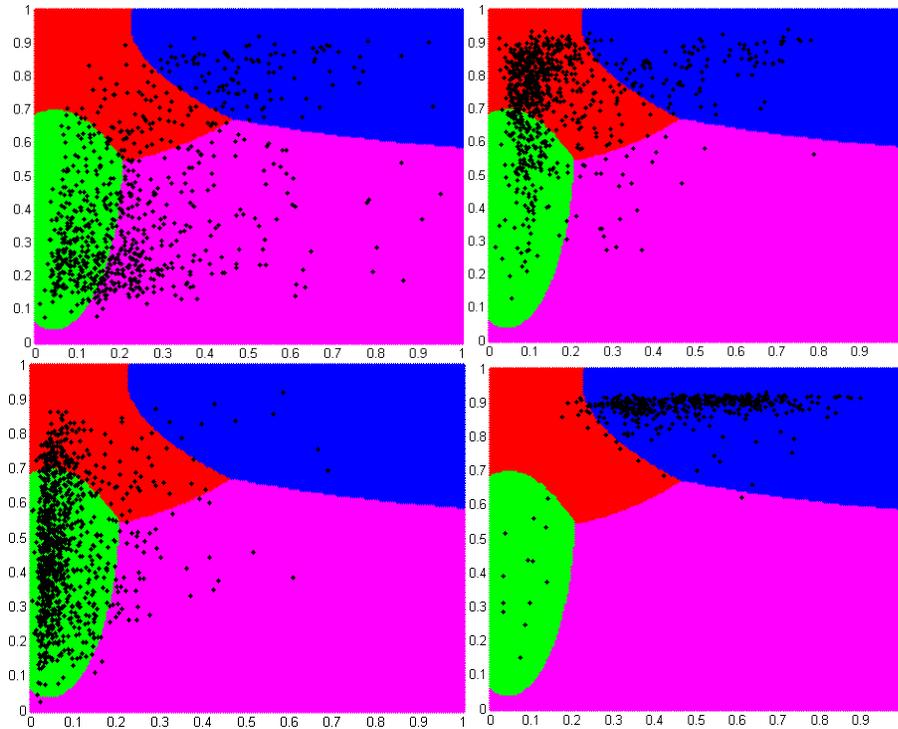


Fig. 24. *Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el (AD-MCD). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).*

Tabla 8. *Matriz de confusión deriva de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (AD-MCD).*

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	315	316	84	100
Explosivas sordas	50	779	184	10
Explosivas sonoras	29	189	544	107
Nasales	1	17	13	444

Tabla 9. Resultados de los índices de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (AD-MCD).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	81.77%	18.23%	38.65%	96.62%
Explosivas sordas	76.05%	23.95%	76.15%	76%
Explosivas sonoras	80.83%	19.17%	62.60%	87.68%
Nasales	92.20%	7.79%	93.47%	91.98%

Tasa de Clasificación Total (TCT)= 65.43 %

3.2.1.3 Clasificación mediante el método del árbol de decisión.

Para aplicar la clasificación a partir del método de árbol de decisión se empleó la función *classregtree* del *Matlab 7.8.0*; [10]. La función *classregtree* brinda dos métodos para la clasificación, se seleccionó *Clasification* que se activa por defecto si se etiqueta las clases con texto o una variable categórica.

La variación en los resultados está dada por el número de niveles (N) o ramas del árbol; se necesita conocer primero el número máximo de niveles que alcanza el árbol, en este caso es $N=38$ con el cual se aprecia la presencia de subregiones en la figura, para resolver este problema se empleó la función *prune* (podador) con la cual se eliminan ramas del árbol de decisión. Para alcanzar el número de ramas óptimo, se tuvo en cuenta el compromiso entre el número de ramas y las regiones conexas clasificadas. El número de ramas que cumple con la condición es $N=5$, En la Figura 25 se ilustra como quedaron agrupadas las regiones conexas con $N=5$.

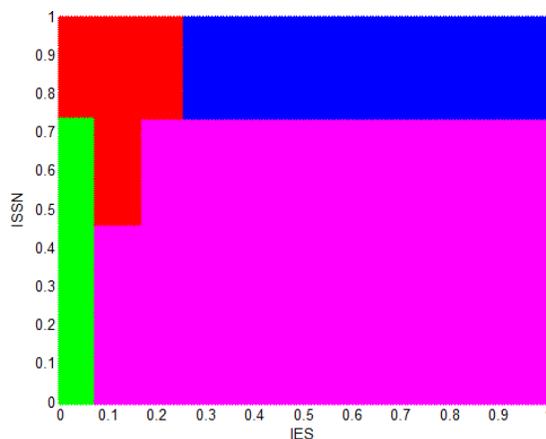


Fig. 25. Regiones de clasificación de los grupos de consonantes a partir del método de árbol de decisión (TREE).

En la Figura 26 se muestra como quedan ubicadas las regiones de cada grupo de consonante con sus respectivas aglomeraciones de puntos que identifican las posiciones reales de estos grupos. En las tablas 10 y 11 se brinda los resultados estadísticos obtenido al utilizar el (TREE).

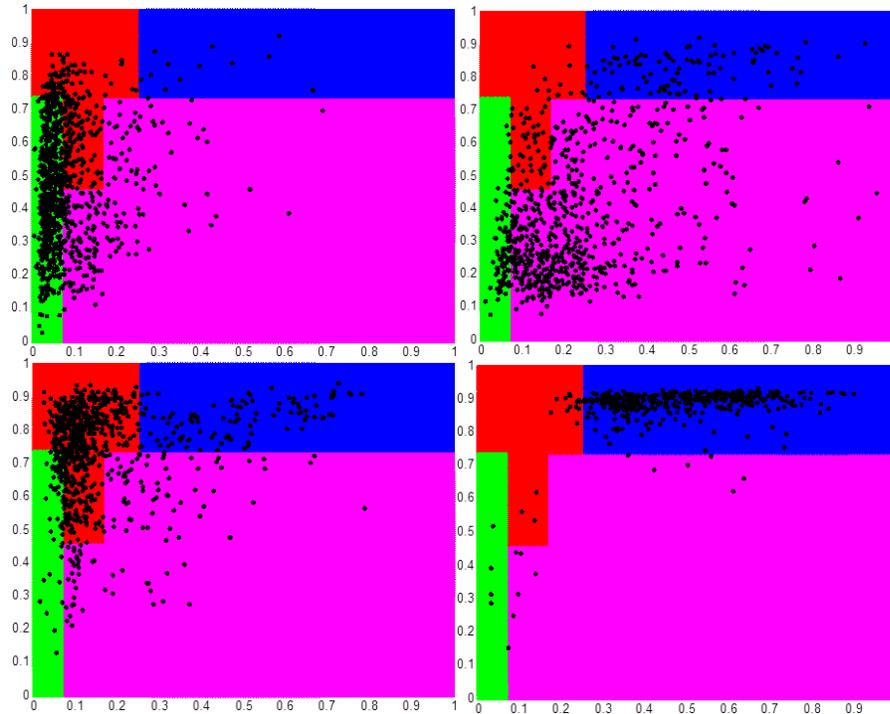


Fig. 26. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el método (TREE). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 10. Matriz de confusión deriva de la clasificación mediante el método de árbol de decisión (TREE).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	579	87	58	91
Explosivas sordas	205	591	215	12
Explosivas sonoras	110	60	596	103
Nasales	12	4	16	443

Tabla 11. Resultados de los índices de la clasificación mediante el método de árbol de decisión (*TREE*).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	82.30%	17.69%	71.04%	86.18%
Explosivas sordas	81.80%	18.19%	57.77%	93.19%
Explosivas sonoras	82.34%	17.66%	68.58%	87.50%
Nasales	90.16%	9.83%	93.26 %	89.62%
Tasa de Clasificación Total (TCT)= 69.42 %				

3.2.1.4 Clasificación mediante el método de los k vecinos más cercanos (*KNN*).

Para la implementación de este modelo se utilizó una de las funciones del *Matlab* versión 7.8.0R2009a llamada *knnclassify*. En este método la variabilidad de los resultados está dada por la cantidad de k vecinos que se tengan en cuenta, con cantidad menores k vecinos se obtienen índices de clasificación mayores (TCT hasta el 100%); pero en las regiones clasificadas puede que se observen lagunas o regiones no conexas, por lo contrario si se le introducen números mayores k vecinos los resultados estadísticos serán menores.

Como no existe un algoritmo para implementar el número de k vecinos óptimo a utilizar en la clasificación, la vía más factible para ello es la experimentación de la manera siguiente: se fue aumentando la cantidad de k vecinos más cercanos hasta que no se observen subregiones en la gráfica o que es lo mismo hasta que se agrupen los cuatro grupos en proceso. Se crea un estado de compromiso entre el número de k vecinos más cercanos y las regiones conexas clasificadas de la gráfica. El número de k vecinos que cumple con la condición es, $k=95$ que es aproximadamente el 3% del total de sílabas procesadas. En la Figura 27 se aprecia la forma en que se agrupa las regiones clasificadas con $k=95$, es de destacar que se empleó con la distancia Euclidiana [31] [11]. El magenta representa las fricativas, el rojo las explosivas sonoras, el verde a las explosivas sordas y el azul a las nasales.

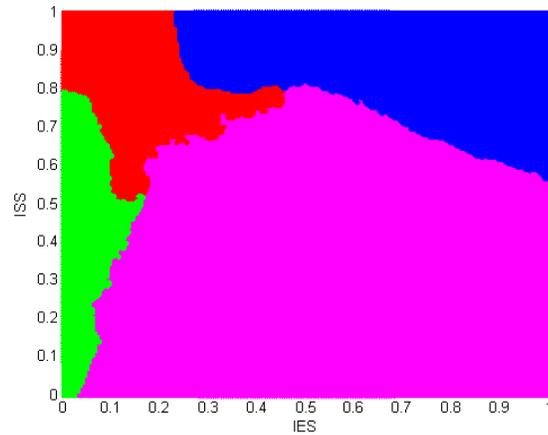


Fig. 27. Regiones de clasificación de los grupos de consonantes a partir del método de k vecinos más cercanos (KNN).

En la Figura 28 se ilustra la ubicación real de las sílabas procesadas separadas en sus grupos correspondientes, conjuntamente con cada región asignada por la clasificación. Se a destacar la clasificación de las nasales que produjo una mejorías en los resultados .A diferencia de en [10] con este método se obtuvieron los mejores resultados como se puede apreciar en las tablas 12 y 13.

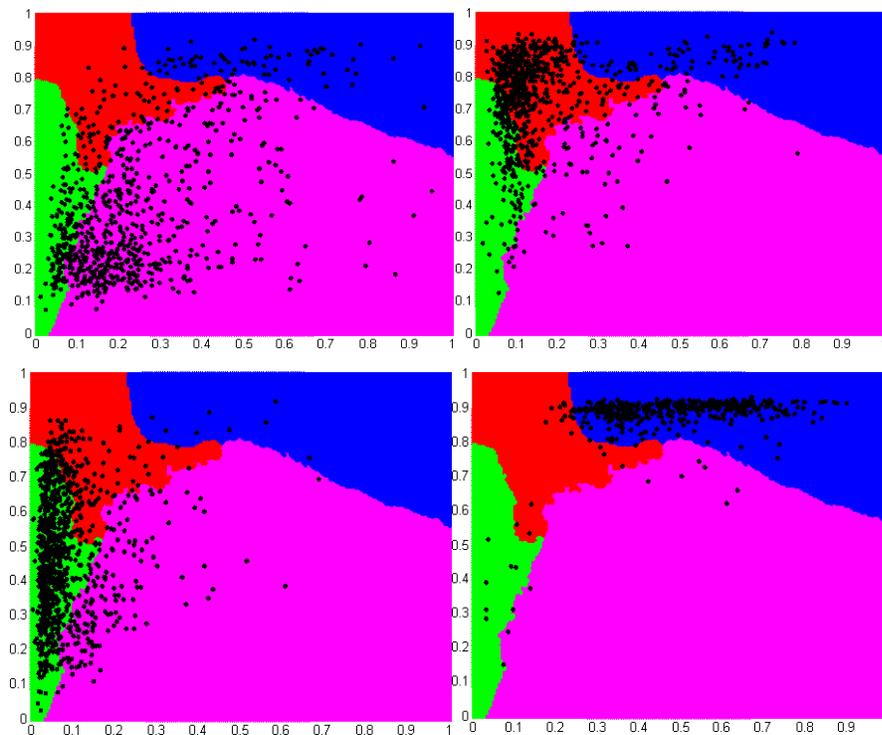


Fig. 28. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el método (KNN). La ubicación real de los grupos de sílabas

(*superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales*).

Tabla 12. Matriz de confusión deriva de la clasificación mediante el método de k vecinos más cercanos (KNN).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	551	127	69	68
Explosivas sordas	168	703	144	8
Explosivas sonoras	84	141	567	77
Nasales	10	7	19	439

Tabla 13. Resultados de los índices de la clasificación mediante el método de k vecinos más cercanos (KNN).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	83.47%	16.53%	67.60%	88.93%
Explosivas sordas	81.30%	18.70%	68.72%	87.26%
Explosivas sonoras	83.22%	16.78%	65.24%	89.97%
Nasales	94.06%	5.94%	92.42%	94.35%

Tasa de Clasificación Total (TCT)= 71%

3.2.2 Discusión de los resultados de los clasificadores en ISS vs. IES.

Con la aplicación de los modelos de análisis discriminante los resultados alcanzados son infructuosos, tal vez debido al solapamiento excesivo de algunos grupos *_ver* epígrafe 2.4.2, donde el análisis discriminante es deficiente. Los dos restantes el árbol de decisiones y k vecinos más cercanos superan de manera elocuente al análisis discriminante, y entre esto resultó ser superior k vecinos más cercanos con mejor TCT (71%).

3.3 IES vs. DCCM.

También se analizaron las sílabas procesadas detectadas según sus valores de IES y DCCM [11]. La Figura 29, ilustra los pares coordinados en las ordenadas (IES) y en las abscisas (DCCM), con la ubicación de cada una de las sílabas agrupadas como puntos que se identifican con los mismos colores que (ISS vs. IES), en la figura se aprecia que tanto las explosivas sonoras, como explosivas sordas y las nasales se encuentran aglomeradas en

áreas relativamente definidas. Las fricativas a parte que se encuentra dispersas en el plano, presentan bajos niveles de solapamiento para valores mayores de 4 *DCCM*. Al igual que con *ISS* vs. *IES*, se determinaron posibles regiones en forma de elipse y de rectángulo a cada grupo de sílaba analizada como se observa en las Figuras 30 y 31.

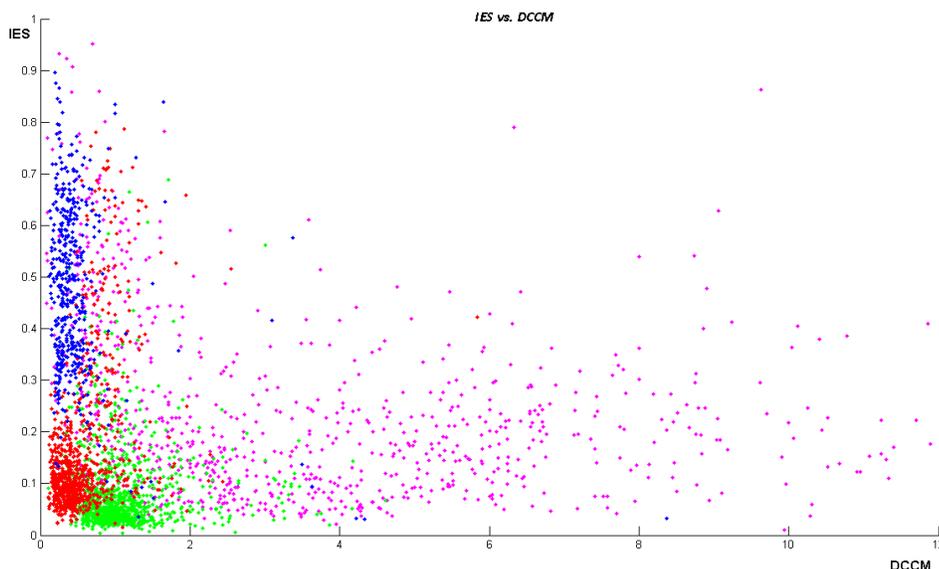


Fig. 29. Ubicación de cada grupo de sílaba en el plano coordenado *IES* vs. *DCCM*.

En la Figura 30, aparecen las elipses descriptivas, en la cual se destaca el insignificante solapamiento que existe entre una mínima parte de las fricativas y una pequeña porción de las explosivas sordas, todo lo contrario a como se observa en el plano *ISS* y *IES*.

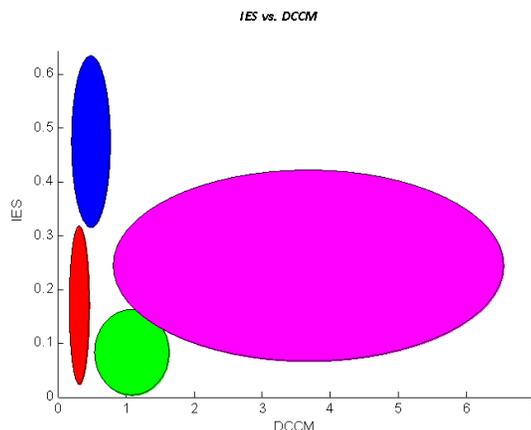


Fig. 30. Regiones de agrupación de las consonantes con los rasgos *IES* y *DCCM*.

En la siguiente Figura 31, se distingue como se ubican las posibles regiones en forma de rectángulos bien separadas unas de otras.

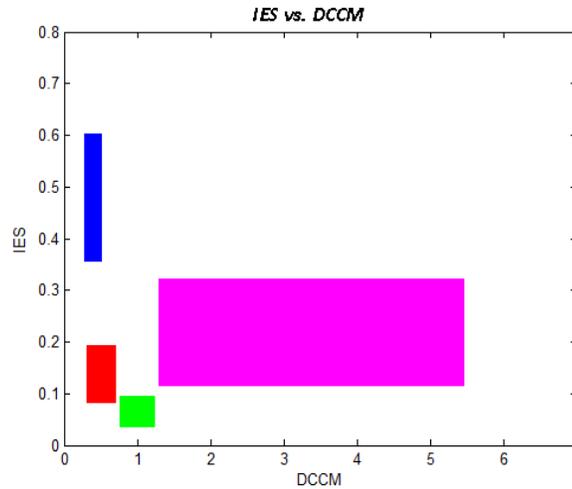


Fig. 31. Distribución de las regiones en el plano IES vs. DCCM.

3.3.1 IES vs. DCCM en la separación de los cuatro grupos de consonantes.

A continuación se evalúa la capacidad de IES y DCCM en la separación de los cuatro grupos de consonantes mediante el empleo de los mismos clasificadores destinados para esta tarea.

3.3.1.1 Clasificación de las consonantes mediante análisis discriminante cuadrático con matriz de covarianza diagonal (*AD-MCD*).

Al igual que con el plano *ISS vs. IES* en *IES vs. DCCM* las sílabas procesadas fueron expuestas a un proceso de clasificación aplicando análisis discriminante como en [11] obteniéndose los mismos resultados, tal como se muestra en la siguiente Figura 32.

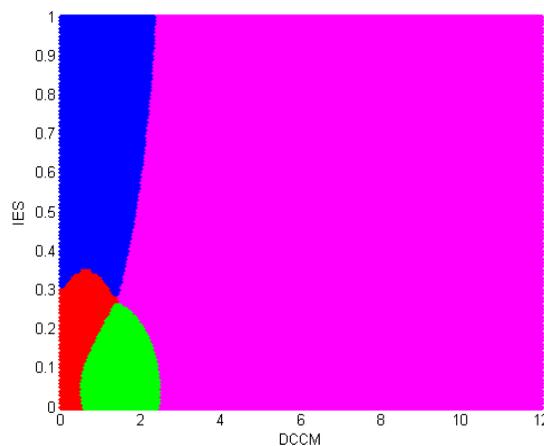


Fig. 32. Representación de las regiones de clasificación de los grupos de consonantes a partir del análisis discriminante cuadrático con matriz de covarianza diagonal (AD-MCD).

La Figura 33, ilustra las ubicaciones de cada grupo de sílabas procesadas con sus correspondidas áreas clasificadas en el plano *IES vs. DCCM*. En las tablas 14 y 15, se muestra la matriz de confusión y los resultados estadísticos que evalúa la clasificación.

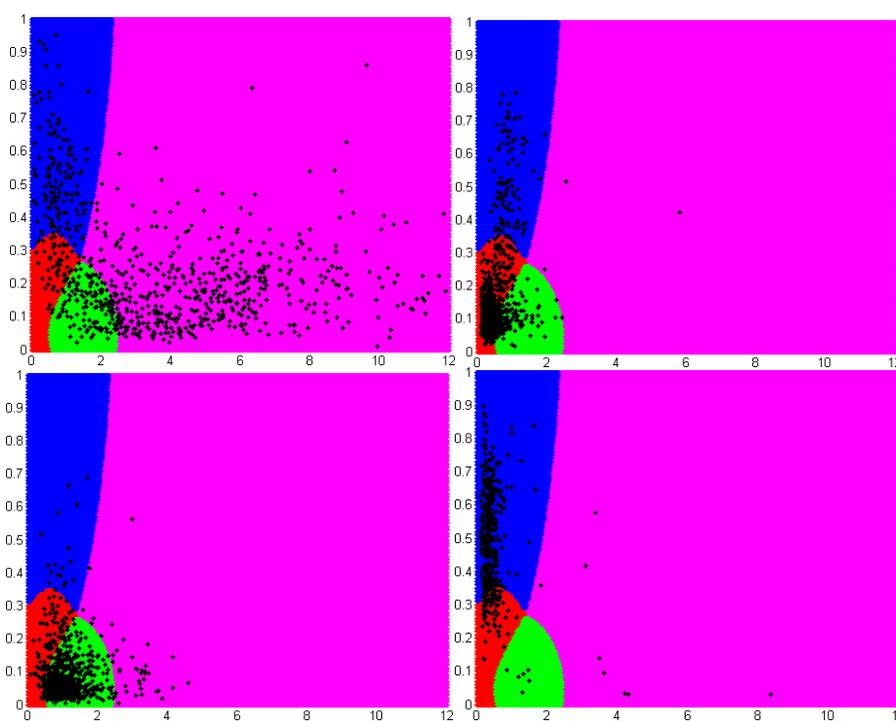


Fig. 33. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el (AD-MCD). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 14. Matriz de confusión derivada de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (AD-MCD).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	493	124	56	142
Explosivas sordas	42	846	118	17
Explosivas sonoras	5	125	638	101
Nasales	8	6	90	371

Tabla 15. Resultados de los índices de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (*AD-MCD*).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	88.15%	11.85%	60.49%	97.68%
Explosivas sordas	86.42%	13.58%	82.70%	88.19%
Explosivas sonoras	84.44%	15.56%	73.42%	88.59%
Nasales	88.56%	11.44%	78.10%	90.39%

Tasa de Clasificación Total (*TCT*)= 73.79%

3.3.1.2 Clasificación de las consonantes mediante análisis discriminante con distancia de Mahalanobis (*AD-DM*).

En la Figura 34 se muestra las regiones de clasificación a partir del uso del análisis discriminante con distancia de Mahalanobis. Destacar que a diferencia que con el (*AD-MCD*) aquí se puede apreciar discontinuidades en la región que representa a las fricativas. Se compararon los resultados derivados de usar estos tipos de análisis discriminantes, a diferencia del plano de *ISS* vs. *IES* se lograron los mejores resultados a partir de utilizar el (*AD-DM*).

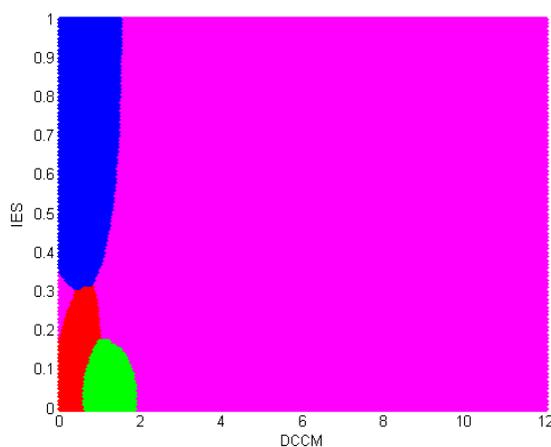


Fig. 34. Regiones de clasificación de los grupos de consonantes a partir del análisis discriminante con distancia de Mahalanobis (*AD-DM*).

A continuación se muestran las ubicaciones de cada sílaba detectadas separadas; en conjunto con las regiones clasificadas por el (*AD-DM*) Figura 35. Es significativa la mejora

que se obtiene en la Tasa de Clasificación de las explosivas sonoras, lo que influye sustancialmente en el resultado final (*TCT*) como se presenta en la tabla 16.

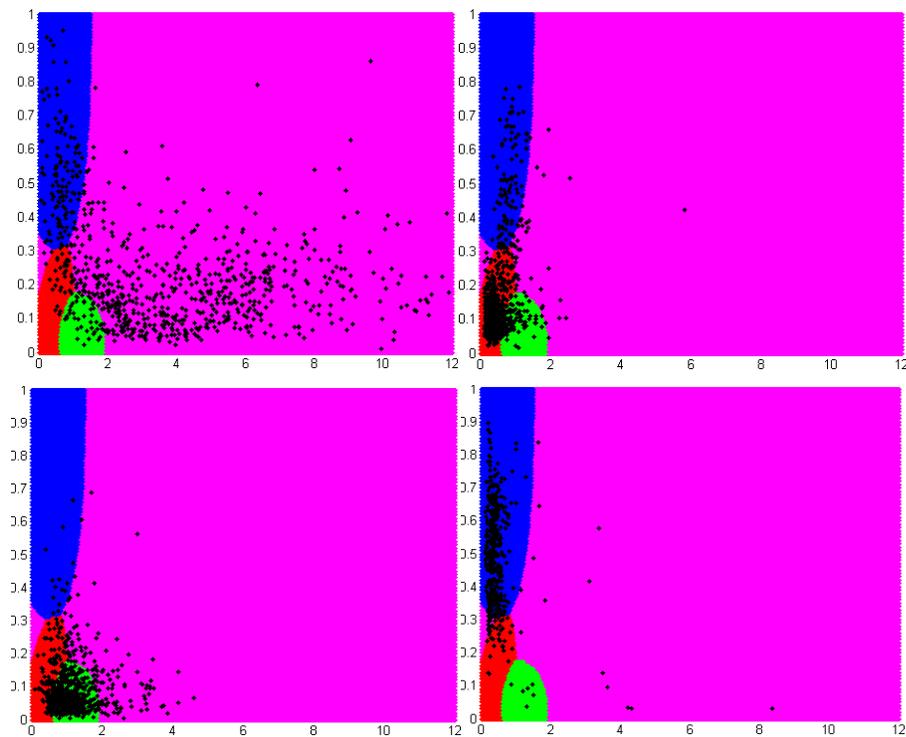


Fig. 35. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el (AD-DM). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 15. Matriz d confusión deriva de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (AD-DM).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	622	39	34	120
Explosivas sordas	130	736	143	14
Explosivas sonoras	57	88	628	96
Nasales	36	6	27	406

Tabla 16. Resultados de los índices de la clasificación mediante el análisis discriminante cuadrático con matriz de covarianza diagonal (AD-DM).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	86.93%	13.07%	76.32%	90.58%
Explosivas sordas	86.80%	13.20%	71.95%	93.84%
Explosivas sonoras	86.02%	13.98%	72.26%	91.18%
Nasales	90.60%	9.40%	85.47%	91.50%
Tasa de Clasificación Total (TCT)= 75.17 %				

3.3.1.3 Clasificación mediante el método de los k vecinos más cercanos (KNN).

El número de k vecinos más cercanos es, $k=95$ al igual que en *ISS vs. IES*. Se obtuvieron los mismos resultados de [11]. En de la Figura 36 se aprecia que la región que corresponde a las explosivas sonoras (rojo) debido al proceso de clasificación a pesar de ser contigua, irrumpe en áreas donde los valores de *IES* son altos, no justificada del punto de vista de las características y la distribución propia de este grupo con respecto a sus valores calculados de energía. Ver epígrafe 3.1.2.

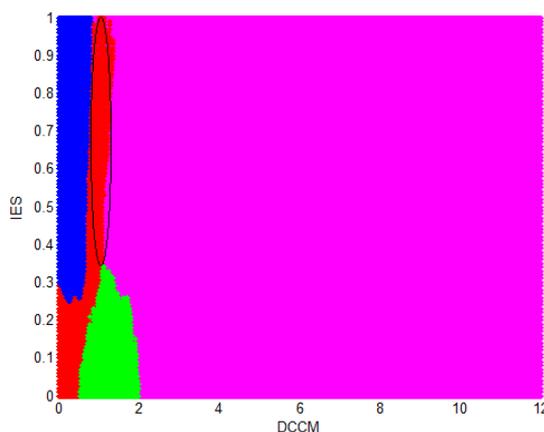


Fig. 36. Regiones de clasificación de los grupos de consonantes a partir del método de k vecinos más cercanos (KNN).

En la Figura 37 se destaca la buena agrupación lograda en las nasales, por su caso mejora sustancialmente los resultados; y por otra parte la no deseada incursión de las explosivas sonoras en zona de altas *IES*, influye en el aumento de su Tasa de Clasificación como se muestran en las tablas 17 y 18.

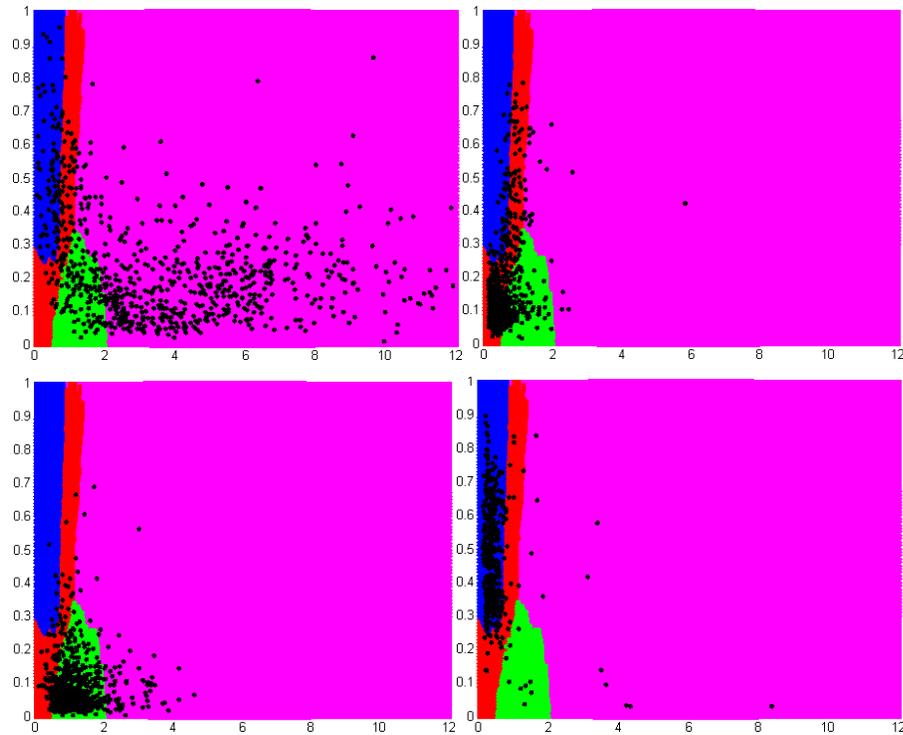


Fig. 37. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el método (KNN). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 17. Matriz de confusión deriva de la clasificación mediante el método de k vecinos más cercanos (KNN).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	569	95	77	74
Explosivas sordas	73	834	110	6
Explosivas sonora	26	136	667	40
Nasales	13	9	40	413

Tabla 18. Resultados de los índices de la clasificación mediante el método de k vecinos más cercanos (KNN).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	88.75%	11.25%	69.82%	95.27%
Explosivas sordas	86.52%	13.48%	81.52%	88.88%
Explosivas sonoras	86.52%	13.48%	76.75%	90.19%
Nasales	94.15%	5.85%	86.95%	95.42%

Tasa de Clasificación Total (TCT) = 78.03 %

3.3.1.4 Clasificación mediante el método de árbol de decisión.

Para el plano *IES* vs. *DCCM*, se obtuvo $N=36$ como máximo número de niveles del árbol, con el que se observan una vez más regiones no contiguas. Para resolver este problema se eliminaron 30 niveles, quedando los 6 restantes para la clasificación y cumpliéndose así la condición de compromiso. Los resultados alcanzados se muestran en la Figura 38, en ella se aprecia nuevamente la incursión no justificada de la región clasificada como explosivas sonoras (rojo), en zonas que marcan altos valores de energía.

Con el método del árbol de decisión se adquieren pequeñas mejorías en el resultado final superando el alcanzado en (*KNN*), según como indica la Tasa de Clasificación Total (*TCT*), debido a la buena agrupación lograda de las explosivas sordas y de las nasales tal como se muestra en la Figura 39, reduciendo en estos casos los falsos negativos presentes en el método anterior (entre los dos grupos, de 360 falsos negativos en (*KNN*) a 210 en el (*TREE*)). Los resultados estadísticos así como la matriz de confusión aparecen en las tablas siguientes 19 y 20.

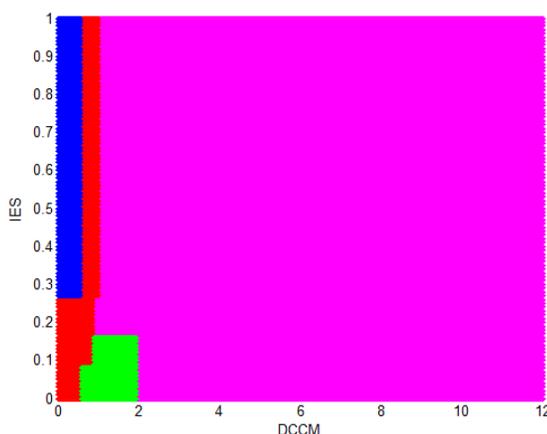


Fig. 38. Regiones de clasificación de los grupos de consonantes a partir del método de árbol de decisión (TREE).

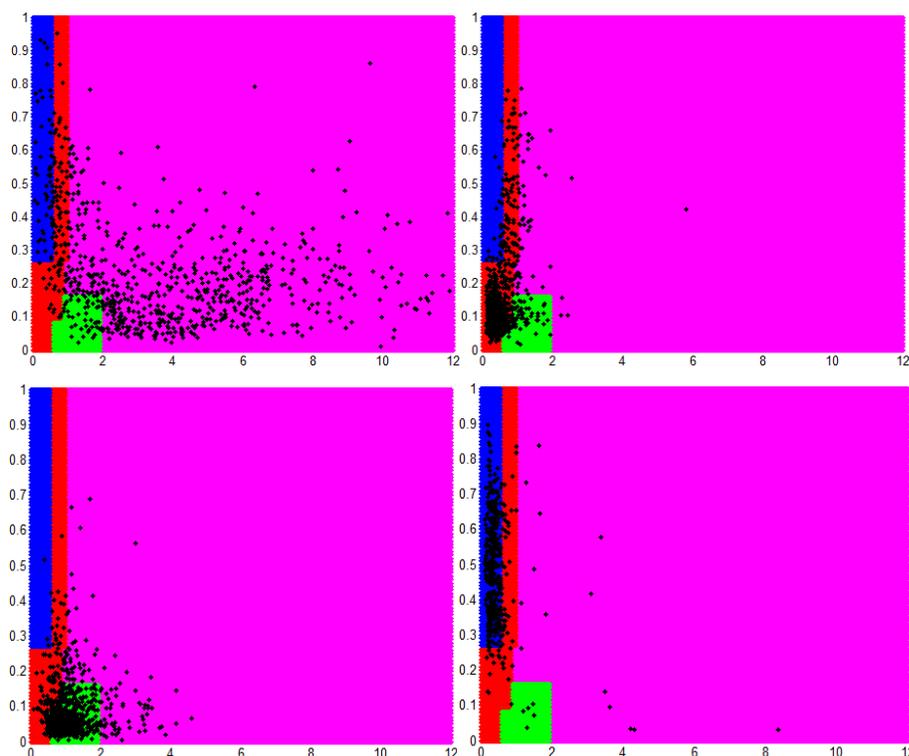


Fig. 39. Ubicación de los cuatro grupos de consonantes según corresponde con la clasificación mediante el método (TREE). La ubicación real de los grupos de sílabas (superior izquierdo-fricativas, superior derecho-explosivas sonoras, inferior izquierdo-explosivas sordas e inferior derecho-nasales).

Tabla 19. Matriz de confusión deriva de la clasificación mediante el método de árbol de decisión (TREE).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	631	41	88	55
Explosivas sordas	117	755	148	3
Explosivas sonoras	53	80	710	26
Nasales	15	5	59	396

Tabla 20. Resultados de los índices de la clasificación mediante el método de árbol de decisión (*TREE*).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	88.40%	11.59%	77.42%	94.16%
Explosivas sordas	87.61%	12.38%	73.80%	88.88%
Explosivas sonoras	85.73%	14.27%	81.70%	87.24%
Nasales	94.87%	5.12%	83.37%	96.89%
Tasa de Clasificación Total (TCT) = 78.32%				

3.3.2 Discusión de los resultados de los clasificadores en IES vs. DCCM.

Nuevamente se observa la inferioridad de los modelos del análisis discriminante, seguramente producto a lo ya mencionado en el epígrafe 3.2.2. En cuanto a los restantes métodos esta vez resulta superior el modelo denominado árbol de decisiones, superando por décimas con una $TCT = 78.32\%$ al k vecinos más cercanos que presenta una $TCT = 78\%$, pero ambos con una clusterización lingüísticamente cuestionable. Por tal motivo se considera el modelo clasificador de análisis discriminante con distancia de Mahalanobis, el más cerca de la realidad con una $TCT = 75.17\%$.

3.4 Comparación entre IES vs. DCCM y ISS vs. IES en la clasificación de consonantes.

Para comparar los resultados obtenidos en los planos *IES vs. DCCM* y *ISS vs. IES*, se analiza el desempeño de estos en la separación de los 4 grupos de consonantes en cuestión, con el uso de distintos clasificadores. Dicha comparación se concentra en el indicador TCT .

Para la comparación en cuanto al rendimiento en la clasificación de los 4 grupos de consonantes las tablas 21 y 22 contienen la tasa de clasificación (TC) lograda por cada uno de los clasificadores empleados.

Tabla 21. Resultados para los cuatro grupos de los clasificadores en cuanto a la Tasa de Clasificación en el plano ISS vs. IES.

	AD-MCD	AD-DM	KNN	TREE
TC FSJ	81.77%	81.68%	83.47%	82.30%
TC PTK	76.05%	76.65%	81.30%	81.80%
TC BDG	80.83%	79.10%	83.22%	82.34%
TC MN	92.20%	92.61%	94.06%	90.16%

Tabla 22. Resultados para los cuatro grupos de los clasificadores en cuanto a la Tasa de Clasificación en el plano IES vs. DCCM.

	AD-MCD	AD-DM	KNN	TREE
TC FSJ	88.15%	86.93%	88.75%	88.40%
TC PTK	86.42%	86.80%	86.52%	87.61%
TC BDG	84.44%	86.02%	86.52%	85.73%
TC MN	88.56%	90.60%	94.15%	94.87%

En las fricativas todas las *TC* alcanzadas en el plano *ISS vs. IES* son inferiores a las logradas a partir de *DCCM vs. IES*, en el primer caso la más alta fue 83.47% (*KNN*) mientras en el segundo la más baja fue de 86.93% (*AD-DM*).

Con *IES vs ISS* las explosivas sordas nunca lograron una *TC* superior al 82 %, obteniendo como mejor resultado $TC = 81.80\%$ con (*TREE*) sin embargo con *DCCM vs IES* todas fueron superiores al 86 %, con 87.61% como mejor resultado con el mismo clasificador.

Al igual que en los anteriores grupos, las explosivas sonoras se mejoran los resultados logrados; pero sin embargo en las nasales se asemejan bastante los resultados en ambos planos, se obtuvieron *TC* superiores al 90%. Obteniéndose el mejor como es habitual en el plano *IES vs. DCCM* empleando el (*TREE*) como clasificador con un 94.87%.

A continuación se presenta la comparación según el rendimiento global empleado como indicador *TCT* a la hora de separar los cuatro grupos de consonantes. Tabla 23.

Tabla 23. Tasa de Clasificación Total (TCT) para cada clasificador.

	AD-MCD	AD-DM	KNN	TREE
TCT ISS vs. IES	65.43%	65.02%	71%	69.42%
TCT IES vs. DCCM	73.79%	75.17%	78.03%	78.32%

Los resultados son superiores en todos los casos al usar los indicadores *DCCM* y *IES*, a partir de los cuales la *TCT* más alta fue 78.32 % con *TREE*. La mayor *TCT* de la combinación *IES vs ISS* fue de 71%, inferior a la menor de la combinación *DCCM vs IES*, de un 73.8% aproximadamente.

3.5 ISS, IES y DCCM, de conjunto, en la separación de los cuatro grupos consonantes.

En el siguiente epígrafe se describe el eje fundamental de este trabajo, donde se evaluará la eficiencia de los rasgos en conjunto, en la tarea de separar las consonantes bajo estudio. La idea surge con el motivo de mejorar los resultados alcanzados por las combinaciones analizadas en los trabajos precedentes [10] [11], reproducidos en los epígrafes previos. Para el cumplimiento de esta faena se recurre a los mismos clasificadores aplicados, *AD-MCD*, *AD-DM*, *KNN* y *TREE*. Para visualizar los diferentes grupos de consonantes en el espacio se consideró que el eje de las abscisas (x) estuviera representado por el indicador *DCCM*, por las ordenadas (y) los valores de *IES* y en la altura (z) el tercer indicador *ISS*, de la forma en que se observa en la Figura 40. En ella se identifican los diferentes grupos o clases de igual manera, el magenta representa las fricativas, el rojo a las explosivas sonoras, el verde a las explosivas sordas y por último el azul a las nasales.

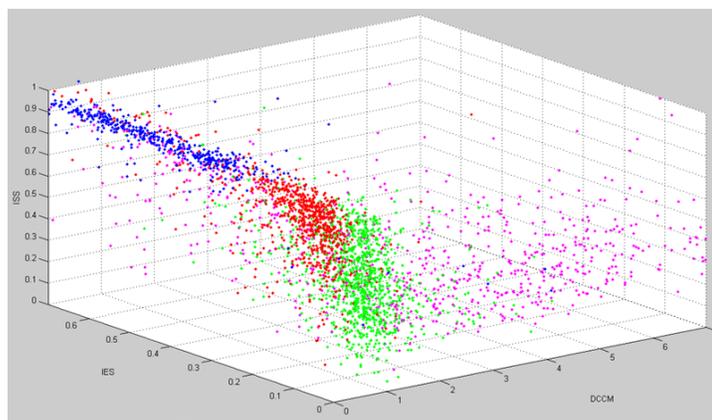


Fig. 40. Ubicación de cada grupo de sílabas en el espacio coordinado por ISS, IES y DCCM.

En la figura se puede apreciar la complejidad del problema que afrontaran los clasificadores, aún se percibe solapamiento y dispersión entre los grupos en análisis, destacándose el grupo perteneciente a las fricativas. Las nasales y sonoras aparecen relativamente bien ubicadas, con las sordas un poco menos concentradas en cuanto a "sonoridad" de lo esperado.

Este hecho pone en duda la capacidad de los clasificadores de obtener regiones útiles de separación de las clases a partir de esta Base de Datos en específico.

3.5.1 Clasificación a partir de los distintos métodos empleados.

3.5.1.1 Clasificación mediante análisis discriminante cuadrático con matriz de covarianza diagonal (AD-MCD).

La clasificación a partir de (AD-DCM) en ISS, IES y DCCM en conjunto, se llevó a cabo de igual forma con que se manejó las sílabas procesadas en las combinaciones anteriores. El hecho de introducir un tercer rasgo en el proceso, sí requirió de un costo computacional mayor (con todos los clasificadores). En la Figura 41 se percibe como quedaron agrupadas las cuatro clases de sílabas en tres dimensiones (3-D).

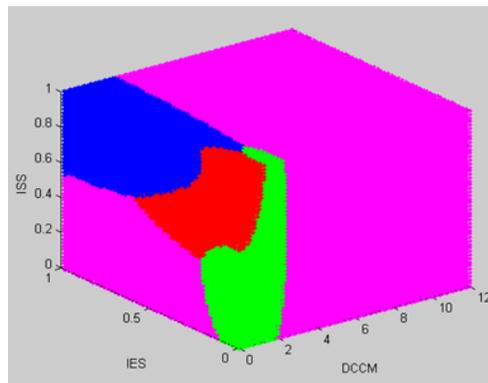


Fig. 41. Representación en 3-D de las regiones de clasificación de los grupos de consonantes a partir del (AD-MCD)

En la figura se aprecia la correlación con la realidad lingüística sobre todo las regiones correspondientes a las nasales y a las explosivas sonoras. En las Tablas 24 y 25 se presentan la matriz de confusión y los resultados estadísticos propios a la clasificación.

Tabla 24. Matriz de confusión deriva de la clasificación mediante (AD-MCD).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	524	133	62	96
Explosivas sordas	49	768	196	10
Explosivas sonora	11	125	641	92
Nasales	7	6	28	434

Tabla 25. Resultados de los índices de la clasificación mediante (AD-MCD).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	88.75%	11.25%	64.29%	97.16%
Explosivas sordas	86.20%	13.79%	75.07%	87.95%
Explosivas sonoras	83.81%	16.18%	73.76%	87.63%
Nasales	83.40%	16.59%	91.36%	82.01%

Tasa de Clasificación Total (TCT) = 74.4 %

3.5.1.2 Clasificación mediante el análisis discriminante con distancia de Mahalanobis (AD-DM).

En la Figura 42 se observa el resultado obtenido tras aplicar el modelo de (AD-DM). En ella se percibe como corresponde las regiones agrupadas clasificadas con la ubicación real de los grupos de sílabas aunque las explosivas sordas incursionan no justificadamente del punto de vista lingüístico en espacios que se reflejan altos valores de sonoridad con estas dos variantes como se puede ver se obtienen muy buenos resultados, pero aún los resultados estadísticos no superan los alcanzados con la combinación *IES* y *DCCM* en el uso del (*TREE*) como se ilustra en las tablas 26 y 27.

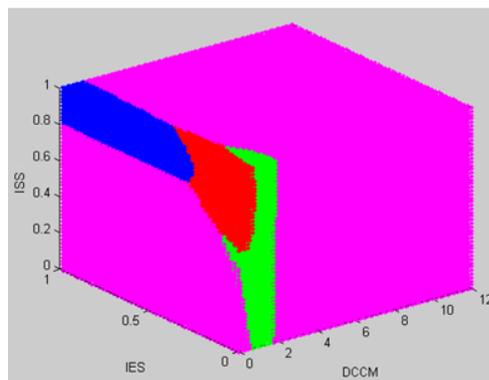


Fig. 42. Representación en 3-D de las regiones de clasificación de los grupos de consonantes a partir del método (AD-DM).

Tabla 26. Matriz de confusión deriva de la clasificación mediante el método de (AD-DM).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	683	37	25	70
Explosivas sordas	160	736	116	11
Explosivas sonoras	79	115	589	86
Nasales	21	6	36	412

Tabla 27. Resultados de los índices de la clasificación mediante el método de (AD-DM).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	87.68%	12.31%	83.80%	89.01%
Explosivas sordas	86.20%	13.79%	71.94%	92.95%
Explosivas sonoras	85.64%	14.36%	67.77%	92.34%
Nasales	91.07%	8.92%	86.73%	91.84%

Tasa de Clasificación Total (TCT) = 76.07 %

3.5.1.3 Clasificación mediante el método de k vecinos más cercanos.

El número de k vecinos más cercanos que cumple con la condición es, $k=105$ que equivale al 3.3% del total de sílabas procesadas (3182). No obstante en la Figura 43, puesto que aparecen incursiones no deseadas e injustificadas no son muy confiables los resultados. Se aprecia que las regiones correspondientes a las nasales (en zonas donde tiende a baja sonoridad), explosivas sonoras (se produce el mismo efecto de la combinación *IES* y *DCCM*) y las explosivas sordas (en zonas de alta sonoridad). Por tal de resolver estas inconvenientes se fue aumentando sucesivamente el número de k vecinos más cercanos hasta alcanzar un 21.6% del total (equivale a $k=688$) lo cual no fue posible, ya que aunque disminuye el efecto negativo en la región que concierne a las explosivas sonoras, la región de las explosivas sordas empeora (se mantiene en zonas de alta sonoridad, irrumpiendo en altos valores energía) como se ilustra en la Figura 44, por tal caso es lógico que el

parámetro que denota el rendimiento de este clasificador este muy por debajo de los resultados hasta ahora alcanzados.

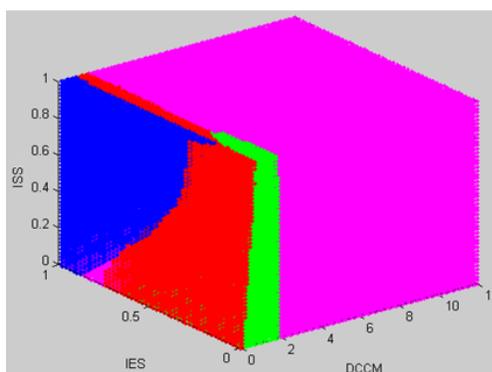


Fig. 43. Representación en 3-D de las regiones de clasificación de los grupos de consonantes a partir del método (KNN) con $k = 105$.

A continuación se presenta las tablas 28 y 29 con la matriz de confusión y los índices de los resultados tras la clasificación mediante (KNN) para $k=105$, en este trabajo no se describe para $k= 688$ porque lo obtenido no es relevante.

Tabla 28. Matriz de confusión deriva de la clasificación mediante el método de (KNN).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	574	116	65	60
Explosivas sordas	75	851	97	0
Explosivas sonoras	18	155	663	33
Nasales	11	8	31	425

Tabla 29. Resultados de los índices de la clasificación mediante el método de (KNN).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	89.16%	10.84%	70.43%	95.61%
Explosivas sordas	86.01%	13.98%	83.19%	87.35%
Explosivas sonoras	87.46%	12.54%	76.29%	91.65%
Nasales	93.27%	6.72%	89.47%	93.94%

Tasa de Clasificación Total (TCT)= 79%

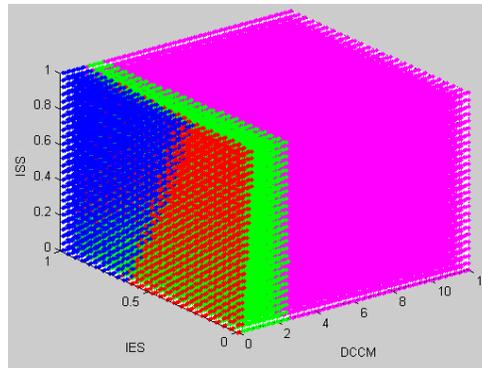


Fig. 44. Representación en 3-D de las regiones de clasificación de los grupos de consonantes a partir del método (KNN) para $k=688$.

3.5.1.4 Clasificación mediante el método de árbol de decisión.

El nivel máximo posible es de $N = 30$, pero se emplean solamente $N = 6$, con el objetivo de eliminar las subregiones y cumplir con una supuesta ubicación de las 4 regiones en cuestión que se observan en la Figura 45. Como se viene apreciando las regiones clasificadas que pertenecen a los grupos de las explosivas (sonoras y sordas) es incoherente con respecto a las características de estos sonidos según los rasgos de sonoridad y energía. Las explosivas sonoras deberían ubicarse entre (0.5-1) y las explosivas sordas entre (0-0.6) de ISS, en la Figura 46 se muestra la irrupción de las explosivas sonoras con altos índices de energía. La tasa de clasificación obtenida supera a la alcanzada con KNN con un 79.98% Tabla 30 y 31.

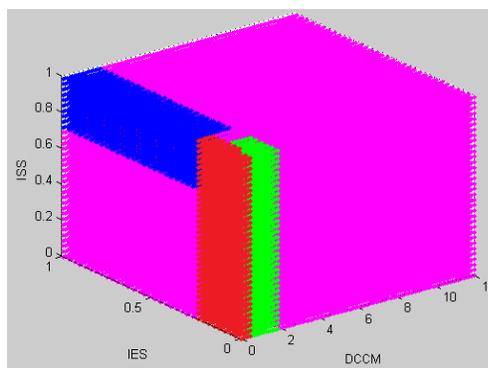


Fig. 45. Representación en 3-D de las regiones de clasificación de los grupos de consonantes a partir del método (TREE) con $N = 6$.

Tabla 30. Matriz de confusión deriva de la clasificación mediante el método de (*TREE*).

	Fricativas	Explosivas sordas	Explosivas sonoras	Nasales
Fricativas	656	45	70	44
Explosivas sordas	128	774	117	4
Explosivas sonoras	60	85	690	34
Nasales	12	5	32	426

Tabla 31. Resultados de los índices de la clasificación mediante el método de (*TREE*).

	Tasa de Clasificación	Tasa de error	Sensibilidad	Especificidad
Fricativas	88.72%	11.28%	80.49%	91.55%
Explosivas sordas	88.06%	11.94%	75.66%	93.93%
Explosivas sonoras	87.49%	12.51%	79.40%	90.53%
Nasales	95.57%	4.43%	89.68%	96.60%

Tasa de Clasificación Total (*TCT*)= 79.98%

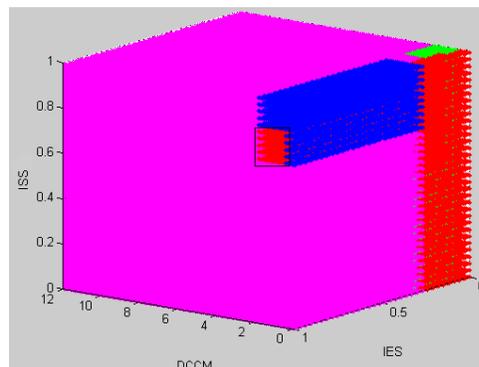


Fig. 46. Representación en 3-D de la incursión no justificada de la región clasificada correspondiente a las explosivas sonoras en altos índices de energía.

3.5.2 Discusión de los resultados de los clasificadores en ISS, IES y DCCM en conjunto.

La *TCT* es superior a las obtenidas en las combinaciones anteriores, para los modelos clasificadores de análisis discriminante existe una relativa correlación con la ubicación real de las sílabas a clasificar, pero aún el índice de rendimiento global está por debajo de lo alcanzado anteriormente por *TREE*. La clasificación mediante los métodos *KNN* y *TREE*

solamente supera en 1% a los alcanzados a partir de combinar *IES* y *DCCM*; pero la ubicación de las regiones clasificadas no son del todo lógicas.

3.5.3 Clasificación mediante la separación empírica manual de las regiones.

Debido a lo planteado anteriormente, surge la idea de definir manualmente regiones de pertenencia de cada clase, y evaluar si la aparente separabilidad es lograble de una manera sencilla. En la Figura 47b muestra la región conformada para aislar las nasales de manera empírica a partir de lo observado en la Figura 47a. Se decidió definir la región que contiene a las nasales como un ortoedro definido por valores inferiores y superiores en los tres ejes ($x = \text{energía}$, $y = \text{Densidad}$, $z = \text{sonoridad}$), los valores preliminares de estos límites son aparentemente $X_{\min} = 0.2$, $X_{\max} = 1$, $Y_{\min} = 0$, $Y_{\max} = 1$, $Z_{\min} = 0.8$ y $Z_{\max} = 1$.

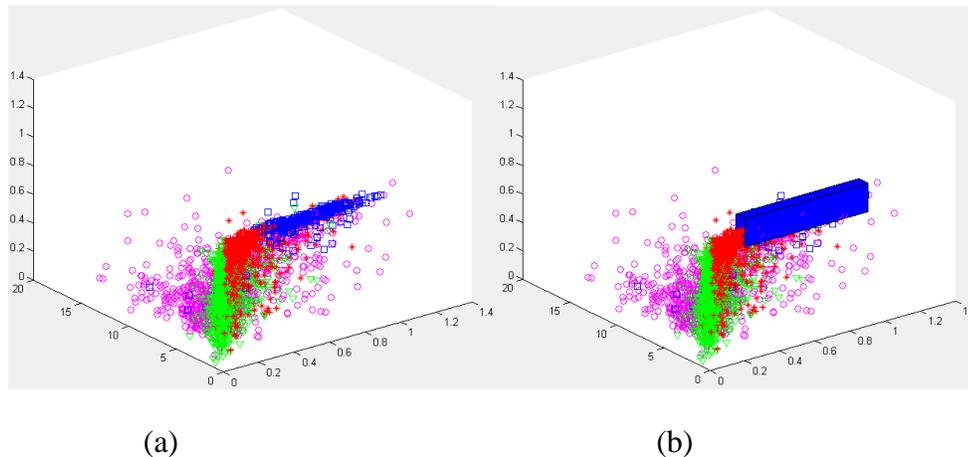


Fig. 47. (a) Ubicación de cada grupo de sílaba en el espacio, (b) región clasificada (ortoedro) correspondiente a las nasales.

En la Figura 48 se presenta la superficie que limita las explosivas sonoras. La separación de estas se realiza mediante el plano de color rojo, inclinado (con respecto al eje z' (*ISS*)), el cual se define por: $z = Ax + By + C$ y se encuentra limitado en la parte inferior por L (límite inferior). Contando como sonoras, las sílabas que se ubiquen a la derecha del plano antes mencionada.

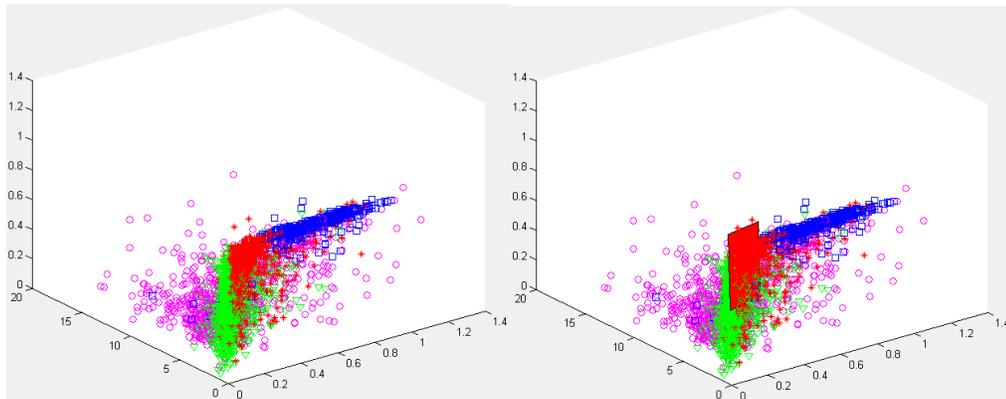


Fig. 48. A la derecha, superficie de separación de las explosivas sonoras.

En la Figura 49 se puede observar la ubicación de las explosivas sonoras según una vista superior y otra frontal, y como quedan separadas de los demás grupos (fricativas y explosivas sordas) por la superficie roja que también limita la zona de las explosivas sonoras situándolas en un área de alta sonoridad y de pocos cruces por cero.

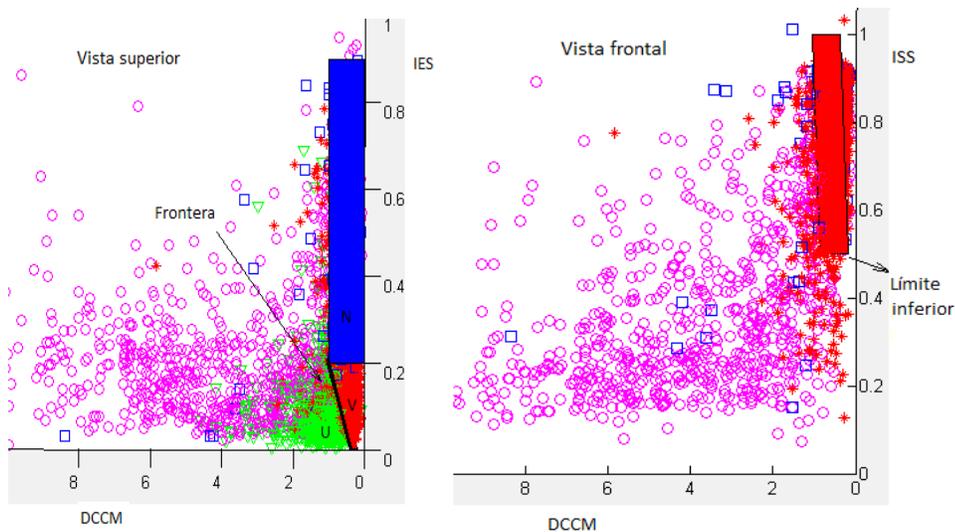


Fig. 49. Ubicación de la región correspondiente a las explosivas sonoras. Valores preliminares: límite inferior = 0.5, $A = -8.6$, $B = 2.6$ y $C = 0$.

En la Figura 50 aparece la superficie (plano verde paralelo con el eje de z' (ISS)) que de conjunto con el plano que se utilizó para confinar las explosivas sonoras, separará a las explosivas sordas. La ubicación de estas superficies se aprecia en la Figura 51, el plano en la “vista superior” representa los segmentos que delimitaran a las explosivas sordas; el de pendiente positiva lo define el plano utilizado para restringir a las sonoras y el segmento de pendiente negativa se definen por una simple ecuación lineal la cual se muestra en la parte

inferior izquierda de la figura. En el plano “vista frontal” se puede apreciar el área de las explosivas sordas ubicada entre el segmento vertical paralelo a *ISS* por la izquierda y limitado por la superficie de separación de las explosivas sonoras por la derecha en la parte superior.

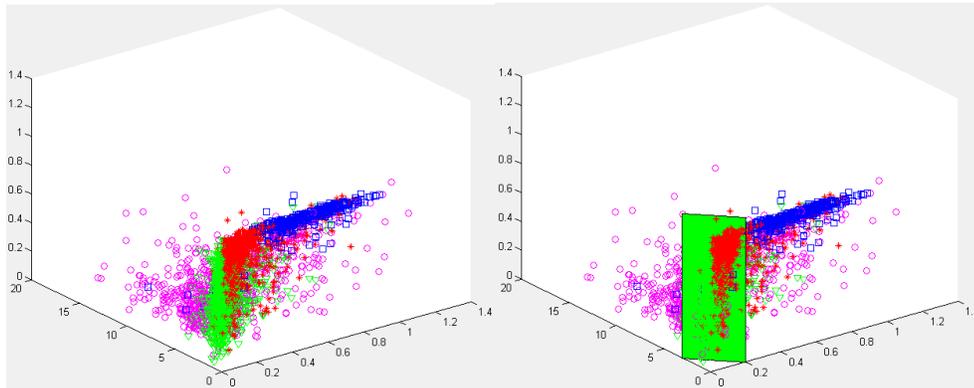


Fig. 50. A la derecha, superficie verde de separación de las explosivas sordas.

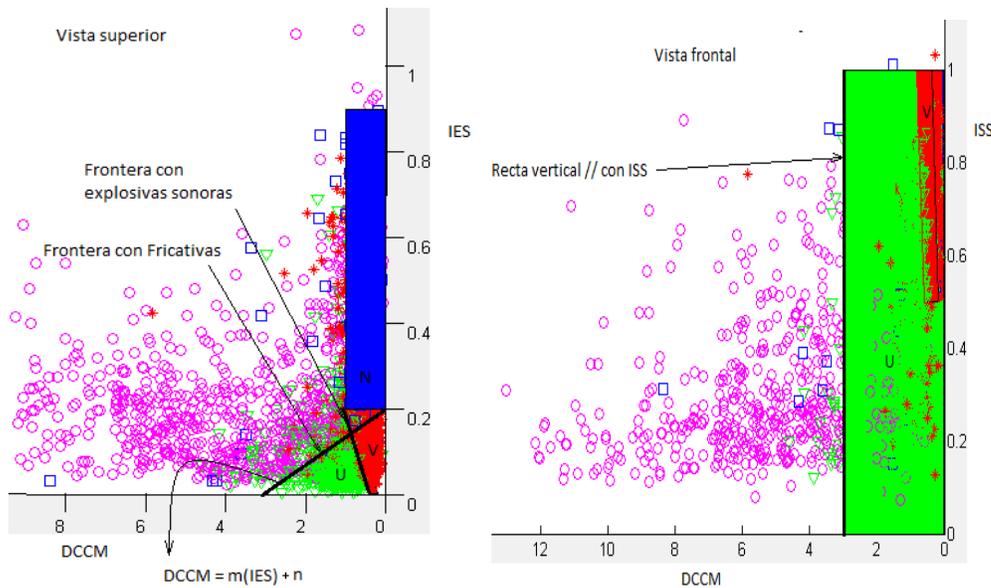


Fig. 51. Vista superior y frontal del plano de separación de las explosivas sordas, valores preliminares: $m = -15$ y $n = 3$.

3.5.3.1 Ajuste fino de los valores y cálculo de la TCT.

Para el ajuste fino de los valores de los parámetros que definen las superficies de separación de las clases se desarrolló una interfaz de usuario en *Matlab* que permite modificar, visualizar y evaluar el efecto en la *TCT*. En la Figura 52 se muestra la interfaz de usuario.

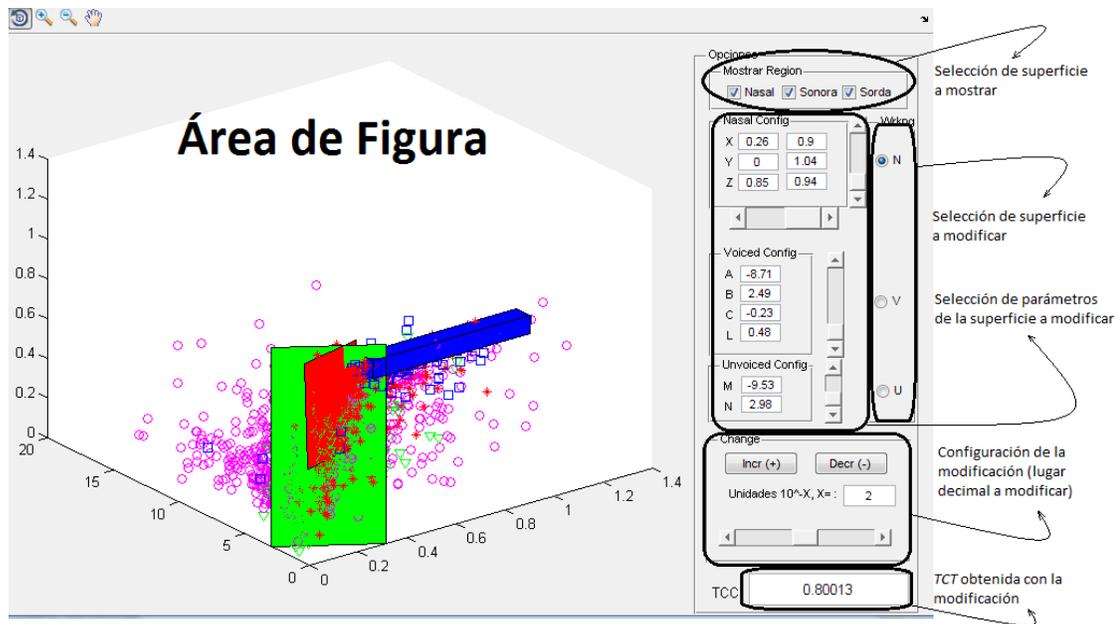


Fig. 52. Interfaz de usuario empleado.

Con los valores preliminares se obtuvo una TCT de 76%. Luego de efectuar los ajustes pertinentes se obtiene una TCT de 80.01% para los valores:

- **Ortoedro:**
 - Xmin = 0.26;
 - Xmax = 0.9;
 - Ymin = 0;
 - Ymax = 1.04;
 - Zmin = 0.85;
 - Zmax = 1;
- **Plano (rojo):**
 - A = -8.71;
 - B = 2.5;
 - C = -0.23;
 - L = 0.48;
- **Plano (verde):**
 - M = -9.53;
 - N = 2.98;

Esta TCT es superior a la obtenida por el mejor método matemático empleado (árbol de decisión), y las regiones tienen una ubicación más lógicas ya que la separación de las regiones empíricas están muy relacionadas con las distribuciones de los histogramas de los índices empleados (ISS, IES y DCCM), y con la lógica lingüística de cada grupo de sílabas procesadas _ver epígrafe 3.1. Por ejemplo se puede observar que la región de las explosivas sonoras en cuanto al índice de “sonoridad” se encuentra entre el límite inferior 0.48 y 1, con respecto a la “energía” entre 0 y 0.31 aproximadamente, y en el indicador “densidad” entre 0 y 1 aproximadamente; por otra parte las explosivas sordas según en el plano ISS vs. DCCM se ubican entre 0 y 1 en cuanto a la “sonoridad”, entre 0 y 2.98 en “densidad”, en cuanto a la “energía”, entre 0 y 0.2. Los resultados adquiridos de manera empírica superan a los que se obtuvieron mediante métodos matemáticos. Además de eliminar los errores de regiones no contiguas, lingüísticamente cuestionables e ilógicos de las explosivas (sonoras y sordas) que aparecen sobre todo en KNN y en TREE.

3.5.4 Conclusiones parciales de los resultados.

Las regiones de clasificación obtenidas mediante los 4 métodos seleccionados no resultan completamente satisfactorias. Muestran incongruencias en cuanto a las características atribuidas a cada uno de los sonidos involucrados, y tasas de clasificación incluso inferiores a una agrupación empírica sencilla definida visualmente.

Esta incapacidad de los clasificadores para producir regiones en correspondencia con las deseadas puede estar motivada por la distribución de los datos y/o por el mismo principio de funcionamiento del clasificador.

CONCLUSIONES Y RECOMENDACIONES

CONCLUSIONES

Al reproducir los experimentos a partir de *ISS vs. IES* y *IES vs. DCCM* se corroboraron los resultados reportados en [10] y [11]. Los resultados son superiores en todos los casos al usar los indicadores *DCCM* y *IES*, a partir de los cuales la *TCT* más alta fue 78.32 % con *TREE*. La mayor *TCT* de la combinación *IES vs. ISS* fue de 71%, inferior a la menor de la combinación *DCCM vs. IES*, de un 73.8% aproximadamente.

Con la implementación de diferentes métodos matemáticos para la clasificación, a partir del espacio tridimensional *ISS*, *IES* y *DCCM* no se logró mejorar de forma sustancial los índices de clasificación de las consonantes presentes en el estudio de las Clínicas Mayo. Las mejoras de la *TCT* fueron de aproximadamente un 1% mediante cada uno de los clasificadores utilizados con respecto a los estudios *IES vs. DCCM*, además de una ubicación no justificada lingüísticamente de determinadas regiones de los grupos pertenecientes a las explosivas. Debido a la imposibilidad de estos métodos matemáticos de separar coherentemente cada uno de los grupos de consonantes involucradas, se creó un interfaz de usuario mediante la cual se realizó la separación manual y empírica a partir de la ubicación visual de los sonidos implicados, método por el cual se obtuvo una *TCT* de un 80.01%, siendo esta la mejor alcanzada hasta en este trabajo.

RECOMENDACIONES

Elaborar una nueva Base de Datos de sujetos sanos, más amplia y con menor presencia de ruido, para evaluar y validar la separación tridimensional aquí propuesta.

Evaluar estos parámetros en la Base de Datos de pacientes disártricos de las Clínicas Mayo.

Considerar nuevos parámetros que puedan mejorar la separación entre estas consonantes.

REFERENCIAS BIBLIOGRÁFICAS

- [1]. 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 (hepatolenticular degenerative)". *Mayo Clinic Proceedings*. 49, pp 405-408, 1974.
- [2]. 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.
- [3]. 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.
- [4]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Differential diagnostic patterns of dysarthria". *Journal of Speech & Hearing Research*. 12, pp 246-269, 1969.
- [5]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Disorders". Philadelphia. Saunders. 1975.
- [6]. Darley, F.L.; Aronson, A.E. & Brown, J.R. "Motor Speech Signs in Neurologic Disease". *Medical Clinics of North America*, 52, 835-844, 1968.
- [7]. Darley, F.L.; Brown, J.R. & Goldstein, N. "Dysarthria in multiple sclerosis". *Journal of Speech & Hearing Research*. 15, pp 229-245, 1972.

- [8]. Bernal, B.J. “Reconocimiento de Voz y fonética acústica”, edición RA-Ma 2000.
- [9]. Llanes, R., Ferrer, C., Díaz, R., Torres D. “Caracterización de consonantes mediante indicadores de energía y sonoridad”. Reconocimiento de Patrones e Ingeniería Biomédica, 2008.
- [10]. Llanes, R. “Separación de consonantes en el plano Energía Sonoridad”. Tesis de Maestría. Universidad Central “Marta Abreu” de Las Villas. 2012.
- [11]. Pérez, Y. “Separación de Consonantes Explosivas de sus Sustituciones Mediante Indicadores en el Dominio del Tiempo”. Tesis de Maestría. Universidad Central “Martha Abreu” de Las Villas 2013.
- [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. A. Cuantificación de parámetros de la voz para aplicaciones médicas. Tesis Doctoral, Universidad Central "Marta Abreu" de las Villas. 2005.
- [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]. Ali, A. M. A., “Acoustic Features of the Automatic Recognition of Fricatives”, Technical Report, TR-CST27AUG97, Center for Sensor Technologies, University of Pennsylvania, 1997.

- [17]. 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*, 87, pp. 820-857, 1990.
- [18]. Stevens, K. N., et all, “Acoustic and perceptual characteristic of voicing in fricatives and fricative clusters”, *Journal of the Acoustical Society of America*, 91, pp. 2979-3000, 1992.
- [19]. González, E., Ferrer, C. y Hernández M. E. “Algoritmo para la Detección de Consonantes Fricativas”. SIE, 2005.
- [20]. Torres, D. “Evaluación de Algoritmos de Detección de Período Fundamental para Voces Patológicas”. Trabajo de Diploma. Universidad Central “Marta Abreu” de Las Villas, 2008.
- [21]. Sharivastav,R Sapienza, C. “Objective measures of breathy voice quality obtained using an auditory model. ” *Journal of the Acoustical Society of America*, 114, pp 2217-2224, 2003.
- [22]. Buder, E.H “Acoustic analysis of vocal quality: a tabulation of algorithms 1902-1990.”. IN KENT, D. R. & BALL, M. J. (Eds.) *Voice Quality Measurement*. San Diego, Singular, 2000.
- [23]. De Cantero, F. J. “Teoría y análisis de la entonación”. Universidad de Barcelona, 2002.
- [24]. Kent, R. D. “Research on speech motor control and its disorders: a review and prospective.” *Journal of Communication Disorders*, 33, 391-428. 2000
- [25]. Ferrer, C.; Hernández, M.E. & González, E. Isolated Syllable Position Detector in Recordings of Patients With Motor Speech Disorders Using Speech Processing Techniques. *Proceedings of the TELECOM'02 International Conference*, Santiago de Cuba, 2002.
- [26]. Rabiner, L. R. “On creating reference templates for speaker independent recognition of isolated words” *IEEE Transactions on Acoustics, Speech and Signal Processing*. Vol. 26, pp 34-42, 1978.

-
- [27]. Ferrer, C., del Toro, A., Gonzales, E. y Hernández M. E. "Using 41 Networks in the Estimation of Consonant Imprecision Ratings." CIARP, 2003.
- [28]. Pérez Y., Ferrer, C., Mújica Y. "Clasificación de consonantes en monosílabos empleando densidad de cruces por cero y energía". UCIENCIA, 2012.
- [29]. Pérez, Y., Ferrer, C., Mújica Y. "Determinación de consonantes fricativas mediante indicadores de densidad de cruces por cero y de energía de las sílabas". Convención Ingeniería Eléctrica, Universidad Central "Marta Abreu" de Las Villas, 2011.
- [30]. Pérez Y., Ferrer, C., Mújica Y. "Clasificación de consonantes mediante indicadores de densidad de cruces por cero y de energía de las sílabas". III Encuentro Regional de Bioingeniería BioVC2012, 2012.
- [31]. Fukunaga, K. "Introduction to statistical pattern recognition". Academic Press. New York, 1972.
- [32]. Clark, P. and F. Evans. "Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations", Ecology, 35, 445-453, 1954.