EN TORNO A LAS VOCALES DEL ESPAÑOL: ANÁLISIS Y RECONOCIMIENTO

EUGENIO MARTÍNEZ CELDRÁN
Laboratori de Fonètica, Facultat de Filologia
Universitat de Barcelona

RESUMEN

Con este estudio se pretendía realizar un análisis de los formantes vocálicos masculinos y femeninos para obtener unos valores que sirvieran de límites entre las distintas clases vocálicas y, así, que pudieran ser utilizados en el Reconocimiento Automático de las Vocales mediante la asociación de cualquier dato que cayera dentro de los alímites establecidos con rasgos fónicos.

Hemos conseguido demostrar que bastan los dos primeros formantes para caracterizar cualquiera de las cinco vocales españolas, que el F1 sirve para determinar los rasgos alto-medio-bajo, que el F2 está implicado en la identificación de los rasgos anterior-central-posterior. Hemos obtenido dos fórmulas que relacionan de forma sistemática los formantes masculinos y los femeninos. Hemos conseguido los límites que hay que utilizar en las reglas y, por último, hemos confeccionado un programa de reconocimiento y hemos verificado que reconoce por encima del 90% las vocales presentadas en palabras diversas y por hablantes diferentes.

ABSTRACT

In this study we intended to carry out an analysis of male and female vowel formants to obtain the data that would serve us to draw the boundaries between the different vowel categories, and could also be used for automatic recognition of vowels through association of any data outside the boundaries established by phonetic features. We have succeeded in proving that the first two vowel formants sufficed to characterize any of the five Spanish vowels, that F1 serves to distinguish between the features: high, middle and low, and also that F2 plays an important role in the identification of front, central and back vowels. We have also obtained two formuli that relate in a systematic way male vowel formants to female vowel formants. We have found the limits necessary for the rules to work, and, finally we created a speech recognition programme and verified that 90% of the vowels of various words uttered by different speakers, were accurately recognised.

1. PRESENTACIÓN

1.1. Análisis

Hace tiempo que se sabe que las cinco vocales del español pueden caracterizarse perfectamente a través de sus dos primeros formantes. Se han proporcionado muchos datos sobre los valores de estos formantes (Martínez Celdrán, 1984:288 y ss), pero quizás no se ha demostrado que esos valores sean los adecuados y, por otra parte, que realmente sean suficientes para discriminar el timbre de cada una de ellas. Además, hay que manifestar que los valores puntuales sólo sirven como valores de referencia y que lo que realmente tiene valor es el **campo de dispersión** y, sobre todo, los límites de este campo para cada una de las vocales. Desde el punto de vista de la invariación, el campo de dispersión es el hecho invariante. Dentro de sus límites los datos pueden ser muy variables, pero se trata de un fenómeno meramente físico, sin repercusiones fonético-fonológicas (M.Tatham, 1990). En la producción de los sonidos sucede la denominada coarticulación que provoca una modificación de la frecuencia canónica del formante al adaptarse al contexto; pero este es un hecho mecánico en la articulación, que no tiene consecuencias en la percepción. La explicación puede basarse en las dos teorías expuestas en la introducción. En la teoría de los "quanta" de Stevens (1972, 1989) se indica que ciertas variaciones en la articulación pueden no causar diferencias acústicas notables, aunque sí pequeños cambios alrededor de un centroide, pero a su vez estos cambios pueden no tener ninguna repercusión perceptiva. Por tanto, es una variación física despreciable totalmente desde el punto de vista perceptivo. Por otra parte, la percepción categorial nos indica que tampoco se perciben las diferencias físicas producidas dentro de una misma categoría determinada. Todo esto viene a decirnos que una vocal, desde la perspectiva acústico-perceptiva, no es un punto en el espacio, sino un dominio con unos límites amplios. Lo que verdaderamente importa es el conocimiento de cada dominio y de sus límites. Este estudio pretende establecer estos dominios y sus límites.

1.2. El reconocimiento automático de las vocales basado en rasgos.

La determinación de los límites tiene también un objetivo clave: establecer las reglas que sirvan para el reconocimiento automático de las vocales. Es decir, nos proponemos confeccionar un programa computacional que reconozca vocales españolas. Haremos que un grupo de personas pronuncien unas palabras. Luego las digitalizaremos. Lo

cual nos permitirá tener unos ficheros con los datos físicos con que se han pronunciado. El programa ha de leer los datos de los dos primeros formantes; obtener un media de las tramas centrales de la vocal para conseguir un valor único de formante, ya que la digitalización proporciona un valor cada 10 ms. Tomaremos los valores centrales para despreciar las transiciones. A continuación se aplicarán unas reglas condicionales que establecen una relación del dato obtenido con el límite de un dominio. Si entra dentro del dominio, entonces tendrá el rasgo correspondiente a ese dominio, de lo contrario tendrá el rasgo opuesto. Por ejemplo, si los rasgos son [alto] y [bajo] y el límite lo situamos en 375 Hz para las voces masculinas, todo valor físico de F1 que se sitúe por debajo de ese límite permitirá que al segmento correspondiente se le asigne el rasgo [+alto], de lo contrario será [-alto]. Luego, se hace lo mismo con el F2 hasta tener el haz de rasgos que caracteriza a un segmento determinado. Se compara el haz obtenido con los que hay en la base de datos y habrá reconocido en el momento que encuentre en la base de datos un conjunto de rasgos igual al obtenido por los reglas. Este procedimiento demuestra la importancia de encontrar esos límites.

2. MÉTODO.

2.1. Análisis

Hemos tomado cinco hablantes masculinos y cinco femeninos universitarios con un español estándar. Sus edades se sitúan entre los 20 y los 30 años. Se les hizo pronunciar cinco veces cada logatomo que seguía el siguiente esquema: *pamp/bVna, tant/dVna, kank/gVna,* donde V = {i,e,a,o,u}; cada hablante ha realizado 30 emisiones de cada vocal (seis logatomos por cinco repeticiones); un total de 300 emisiones al tener diez informantes. Les hemos grabado en una cabina insonorizada en un cassette Marantz, modelo CP430 y un micrófono Shure SM58.

El análisis de los formantes se llevó a cabo en el CSL 4300B de Kay Elemetrics. En primer lugar, se hacía un espectrograma de banda ancha y sobre él se efectuaba el análisis de LPC, con lo cual se trazaba una línea roja que cruzaba por el medio del formante. Colocado el cursor sobre esta línea en el centro aproximado de la longitud de la vocal se obtenía el valor del formante.

2.2. Reconocimiento

Hemos tomado cinco hablantes masculinos que han pronunciado cada uno cien palabras con la estructura CVCV o VCCV, donde C era

siempre un oclusiva o fricativa. Estas palabras se digitalizaron con la estación de desarrollo de voz Philips OM8210 y el programa SP2 que proporciona un archivo de cada palabra con los datos. Véase la estructura de la palabra *piso* de un hablante:

F1	F2	F3	F4	F5	B1	B2	B3	B4	B5	AM	ENT
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
433	1394	2414	3528	4600	3000	800	600	700	3000	0	16
335	1801	2143	3253	4600	153	433	600	265	800	6	31
307	1996	2143	3253	3900	39	234	190	265	140	10	31
307	1996	2414	3253	3900	39	68	190	100	140	11	0
307	2101	2414	3253	3900	39	126	600	100	140	11	1
282	2101	2719	3253	3900	39	68	600	100	140	11	0
282	2101	2719	3253	3900	39	68	600	100	140	11	2
307	2101	2719	3253	3900	39	126	600	100	140	11	3
307	2101	2719	3523	3900	39	126	600	100	140	11	0
307	2101	2719	3253	3900	77	126	190	265	800	11	31
307	2101	2719	3253	4600	77	126	190	100	800	11	31
282	2101	2719	3528	4600	77	433	600	265	800	10	0
168	2211	2719	3528	4600	303	433	600	265	335	9	16
217	2101	2719	3826	4600	600	433	190	100	800	9	16
1119	2101	3063	3826	4600	600	433	600	265	800	9	16
1119	1896	3063	3528	4600	3000	433	3000	265	800	9	16
1220	1996	3063	3528	4600	600	433	600	265	800	10	16
1119	1996	3063	3528	4600	600	433	600	700	335	10	16
1450	1801	3063	3528	4600	600	433	600	265	335	8	16
1450	2211	3063	3528	4600	303	433	600	700	800	9	16
1450	1896	2719	3528	4600	600	433	600	265	335	9	16
398	1626	2719	3528	4600	600	234	190	700	800	9	1
398	1545	2414	2999	4600	153	126	600	700	800	9	0
433	1467	2143	2999	3900	153	234	190	265	800	9	0
433	1394	2143	2999	3900	153	234	600	265	800	9	0
472	1324	2143	2999	3900	77	234	190	100	335	9	0
472	1258	2143	2999	3900	77	126	190	100	140	9	0
472	1136	2143	2999	3900	77	126	60	100	335	8	0
472	1136	2414	2999	3900	77	126	190	265	335	9	0
472	1079	2414	3253	3900	153	234	190	265	800	8	0
472	1025	2414	2999	4600	153	126	190	265	800	7	0
515	1025	2414	3253	4600	153	234	190	265	800	8	0
472	925	2414	2999	4600	303	433	190	265	335	8	0

Como se ve, la estructura consiste en una matriz donde las columnas son los datos de formantes (Fn), anchos de banda (Bn), amplitud (AM) y entonación (ENT). Y las filas representan tiempo. Cada fila posee aproximadamente 10 ms. El SP2 coloca un código (16) en la ENT para indicar sonido inarmónico; por tanto la segmentación de esta estructura es fácil. Todas las filas que posean 16 en ENT corresponden a consonante, si no, es una vocal. Obsérvese que las nueve primeras filas tienen 16, por tanto se trata de una consonante [p], las 11 siguientes no lo tienen porque es la voal [i], las nueve siguientes lo tienen [s] y acaba con 12 filas sin el 16 [o]. La obtención de los cuatro segmentos es fácil limitándose a este dato y aplicando una regla que diga:

IF ENT = 16 THEN S = [+cons] ELSE S = [-cons]

Por este hecho hemos utilizado exclusivamente palabras con oclusivas y fricativas. Si tomamos en este ejemplo las cinco filas centrales de la primera vocal despreciando las tres primeras y las tres últimas, se obtienen las medias de F1 = 297 Hz y el F2 = 2101.

Por tanto, se ha confeccionado un programa (RAVE "Reconocimiento Automático de Vocales Españolas) que pueda leer un fichero como el anterior, obtenga los segmentos y asocie las medias de los datos con rasgos para, por último, poder comparar con el conjunto de rasgos que define cada segmento, con lo cual habrá reconocido el segmento (S) pronunciado. Hemos tomado cinco hablantes masculinos y hemos hecho que pronuncien un conjunto de palabras con la estructura CVCV o VCCV. Tras digitalizarlas, las hemos sometido al programa RAVE y hemos cuantificado los aciertos y errores cometidos.

3. RESULTADOS

3.1. Análisis de la voz masculina

3.1.1. Formante primero (F1)

F1	i	e	a	o	u
media	313	457	699	495	349
sd	29	40	83	56	38
mínimo	241	381	571	393	277
máximo	414	587	1002	656	449

Las mediciones se refieren a frecuencia (Hz). Sd significa desviación estándar. Realizado el test de Student, se demuestra que todas las medias son diferentes significativamente con nivel de significación de $p < 0.01$.

A continuación queríamos saber si el primer formante es suficiente para diferenciar las cinco vocales; para ello utilizamos una prueba estadística de métodos multivariantes que hace un análisis discriminante; es decir, teniendo en cuenta la distribución de los datos de cada media comprueba la posibilidad de agruparlos en torno a un centroide y mide el grado de pertenencia al grupo.

Real ----------Predicho (porcentajes)----------

F1	i	e	a	o	u
i	79	1	0	0	20
e	0	67	0	29	4
a	0	0	95	5	0
o	0	42	5	52	1
u	38	9	0	0	53

Como se ve, la única vocal que se predice correctamente es [a], todas las demás se confunden en porcentajes elevados; sobre todo [o] y [u]; pero las confusiones nos sugieren otro criterio discriminante: agrupar [i,u] en [altas], [e,o] en [medias] y [a] en [baja]; pues, según ponen de manifiesto los datos, las confusiones por el primer formante se dan entre las parejas expuestas. La confusión entre i/e es mínima, así como entre a/o y u/e. Nuevo cuadro discriminante:

Real ----------Predicho (porcentajes)----------

F1	altas	medias	baja
i,u	95	5	0
e,o	3	94	3
a	0	3	97

Los datos ahora concuerdan perfectamente al tener porcentajes de predicción muy elevados. Esto significa que el F1 es el responsable de la clasificación de las vocales en *altas, medias y baja.*

3.1.2. Formante segundo (F2).

F2	i	e	a	o	u
media	2200	1926	1471	1070	877
sd	153	117	84	114	128
mínimo	1832	1676	1296	793	622
máximo	2523	2212	1642	1313	1175

Realizada la prueba de la T de Student todas las medias resultan diferentes significativamente con un nivel de $p < 0.01$.

Nuevamente se busca si el F2 por sí solo es suficiente para discriminar las cinco vocales. Realizada la prueba estadística resulta lo siguiente:

Real ----------Predicho (porcentajes)----------

F2	i	e	a	o	u
i	83	17	0	0	0
e	12	85	3	0	0
a	0	0	100	0	0
o	0	0	1	79	20
u	0	0	0	19	81

Los porcentajes son bastante elevados, aunque en verdad el F2 sólo discrimina perfectamente la vocal [a]. Obsérvese, por otra parte, que ahora las confusiones nos sugieren que debemos agrupar [i,e] en [anteriores] y [o,u] en [posteriores] y realizar una nueva discriminación con estas agrupaciones.

Real ----------Predicho (porcentajes)---------

F2	anteriores	posteriores	central
i,e	96	0	4
o,u	0	95	5
a	0	0	100

Está claro, pues, que el F2 clasifica las vocales perfectamente dentro de los rasgos anterior/posterior.

3.1.3. La acción conjunta de F1 y F2 masculinos.

La figura 1 presenta una visión de conjunto de las medias de los formantes primero y segundo de las diferentes vocales. La figura 2 presenta los mínimos y máximos encontrados; por tanto, muestra la variación existente en cada media.

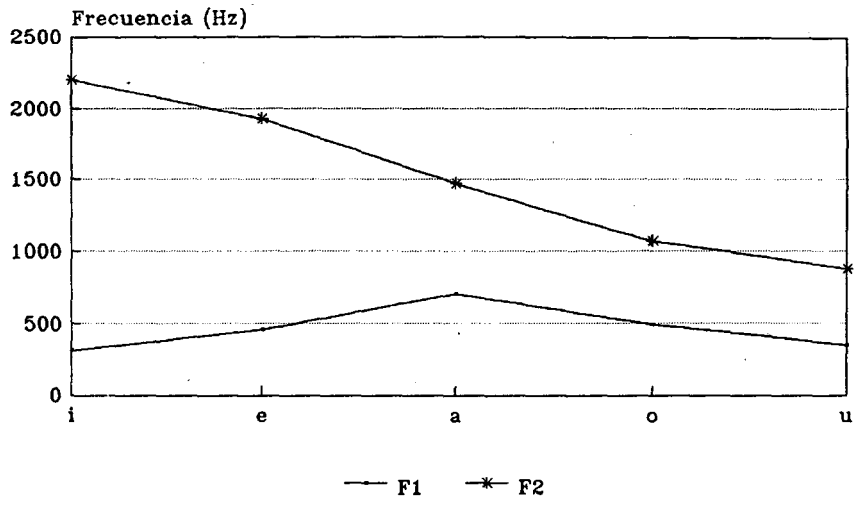


Fig. 1

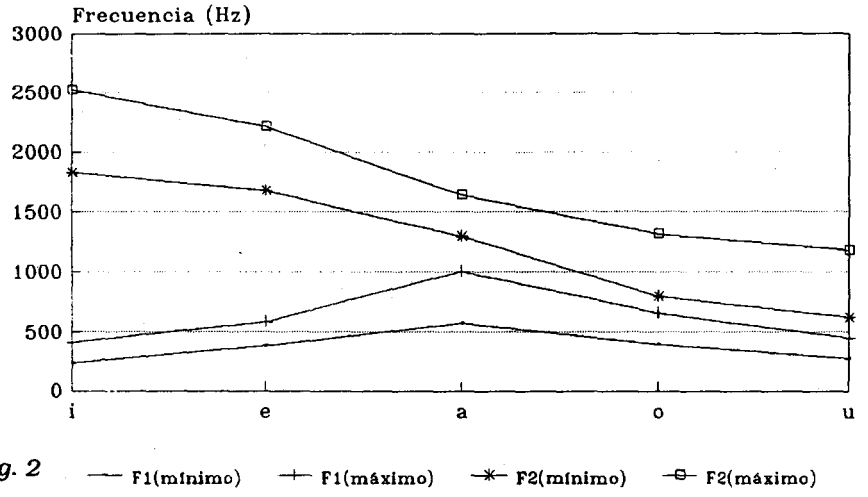


Fig. 2

Parece evidente que si el primer formante clasifica mejor las vocales agupándolas en altas-medias-bajas y que si el segundo las clasifica mejor agrupándolas en anteriores-central-posteriores entonces la acción conjunta de los dos formantes ha de discriminar satisfactoriamente esas cinco vocales:

Real ----------Predicho (porcentajes)----------

F1 y F2	i	e	a	o	u
i	97	3	0	0	0
e	2	98	0	0	0
a	0	0	99	1	0
o	0	0	0	93	7
u	0	0	0	4	96

Ahora, la acción conjunta de los dos formantes determina perfectamente, dados los altos porcentajes de predicción, las cinco vocales. Existen todavía algunas confusiones, aunque son mínimas. Además, teniendo en cuenta la acción de ambos formantes el análisis discriminante determina unos centroides y presenta en un gráfico las agrupaciones que se originan alrededor de los centroides. Esas agrupaciones son los campos de dispersión de las vocales. Obsérvese cómo se distribuyen las cinco vocales en el gráfico de la figura 3 (1 significa vocales altas; 2, medias y 3, bajas). Las confusiones presentadas en el cuadro anterior corresponden a las intersecciones que se observan en los grupos de vocales. Esas intersecciones se ven más entre [o] y [u]. Los valores de X e Y no se corresponden con frecuencia (Hz), sino con los valores de los centroides determinados estadísticamente por la función discriminante. No obstante, hay un gran parecido entre este gráfico y otros de cartas de formantes donde el F2 se sitúa en la ordenada y F1 en la abscisa (P. Lieberman, 1977:152-3).

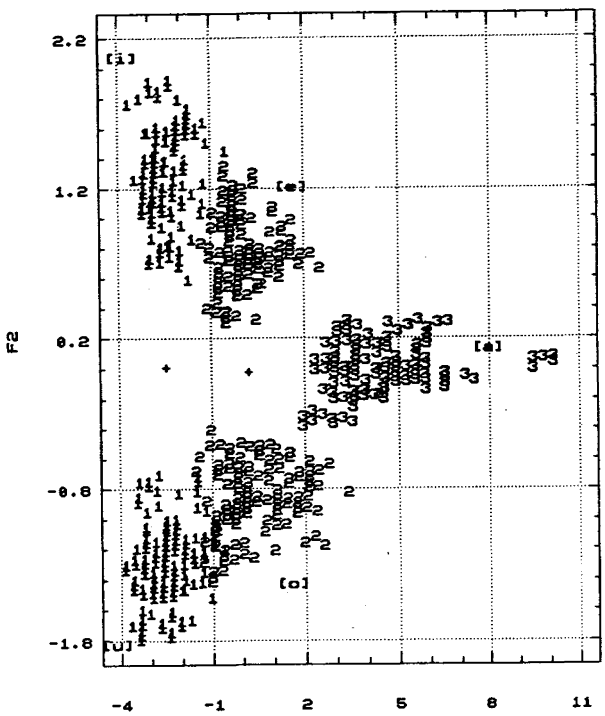


Fig. 3

3.1.4. Determinación de los límites y las reglas: probabilidades de las áreas de cola ("Tail areas probabilities")

Tratamos de encontrar los límites de los campos de dispersión. Para ello hemos partido de la curva normal y hemos buscado qué probabilidad posee un dato concreto que sirva de límite.

Si partimos de la siguiente clasificación de las vocales teniendo en cuenta que el F2 es más relevante que el F1 según la función discriminante, veremos qué datos son los pertinentes para nuestro propósito de tener unas reglas de reconocimiento de vocales:

	i	e	a	o	u
altas	+	-	-	-	+
anteriores	+	+	-	-	-
centrales			+	-	-

Cada columna es diferente en la composición de sus signos positivos y negativos. Necesitamos encontrar tres datos: el primero debe diferenciar entre vocales altas y no altas; el segundo debe distinguir entre central y no centrales, y el tercero deberá captar la diferencia entre anteriores y no anteriores. Bastarán estos tres datos para poder obtener las reglas que reconozcan las cinco vocales españolas.

Exploramos la probabilidad de los datos iniciales y finales de la distribución de cada vocal tanto del F1 como del F2 y llegamos a la siguiente conclusión:

375 Hz es el dato de F1 que puede servir como límite en el mejor de los casos entre altas y no altas. Este dato deja por debajo el 98% del área de [i], y por otra parte, cubre sólo un 2% del área de las vocales medias, lo cual proporciona una alta seguridad de que puede ser un buen dato para el límite de estas vocales; aunque [u], según nuestros datos, deja por debajo un área de sólo un 75% . Esto significa que hay un 2% de probabilidad de que nuestra regla clasifique mal una vocal que ha de reconocer. O dicho de otro modo: hay un 2% de probabilidad de que una vocal media sea reconocida como alta y una alta como no alta, en el caso de [i, e, o]. El mejor dato entre central y no centrales es 1200 Hz en F2, como límite inferior. Este dato puede hacer que [a] se confunda con la vocal [o] en un 0.1%. Es decir, con este dato hay cierta probabilidad de que [a] y [o] puedan confundirse mutuamente. Por último, 1650 Hz de F2 es el mejor dato para separar anteriores y no anteriores; puesto que este dato representa en [e] una probabilidad menor que 0.01% y en [a] 1,7%. Probabilidades realmente muy bajas.

A partir de estos datos se pueden obtener las tres reglas siguientes:

1. IF F1 < 375 THEN V = [+alta] ELSE V = [-alta]
2. IF F2 > 1650 THEN V = [+ant] ELSE V = [-ant]
3. IF F2 > 1200 AND F2 < 1650 THEN V = [+central]
   ELSE V = [-central]

Las reglas están formuladas con el formalismo del lenguaje de programación BASIC que utilizaremos en nuestro programa de reconocimiento. Son reglas de tipo condicional: Si (IF) ... Entonces (THEN) ... De lo contrario (ELSE) ... Si la condición se cumple entonces vale la primera igualdad; si no se cumple, vale la segunda.

Con estas tres reglas el reconocimiento automático de vocales españolas masculinas se ha de producir en un porcentaje superior al

90%, dados los porcentajes de los límites y el análisis discriminante surgido de los datos analizados.

3.2. Reconocimiento de la voz masculina.

A. Aciertos

---Resultados del Reconocimiento(porcentajes)---

INFOR.	i	e	a	o	u	TOTAL
1º	100	88	77	80.5	100	89
2º	100	94	95.4	98.8	89	95.4
3º	100	88	78	97	100	92.6
4º	100	79	100	100	100	95.8
5º	100	100	100	100	100	100
TOTAL	100	89.9	90	95.2	97.8	94.5

B. Errores

	i	e	a	o	u
i	-	0	0	0	0
e	8.4	-	1.7	0	0
a	0	0	-	10	0
o	0	0	0.2	-	4.6
u	0	0	0	2.2	-

3.3. Análisis de la voz femenina

3.3.1 Formante primero (F1)

F1	i	e	a	o	u
media	369	576	886	586	390
sd	50	105	90	80	48
mínimo	276	380	640	398	293
máximo	483	795	1088	795	500

3.3.2 Formante segundo (F2)

F2	i	e	a	o	u
media	2685	2367	1712	1201	937
sd	85	96	92	148	158
mínimo	2471	2108	1503	950	518
	2852	2713	1918	1607	1279

La figura 4 presenta una comparación de las diferencias formánticas de las distintas vocales femeninas. La figura 5 muestra la variación entre mínimos y máximos.

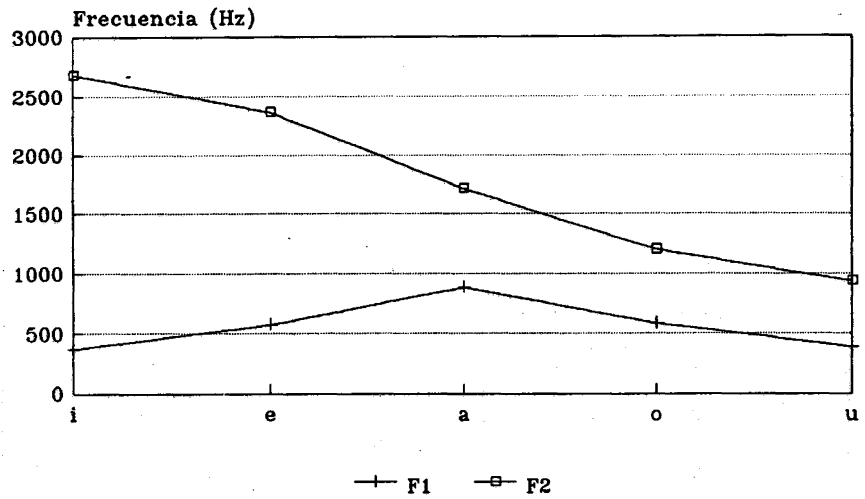


Fig. 4

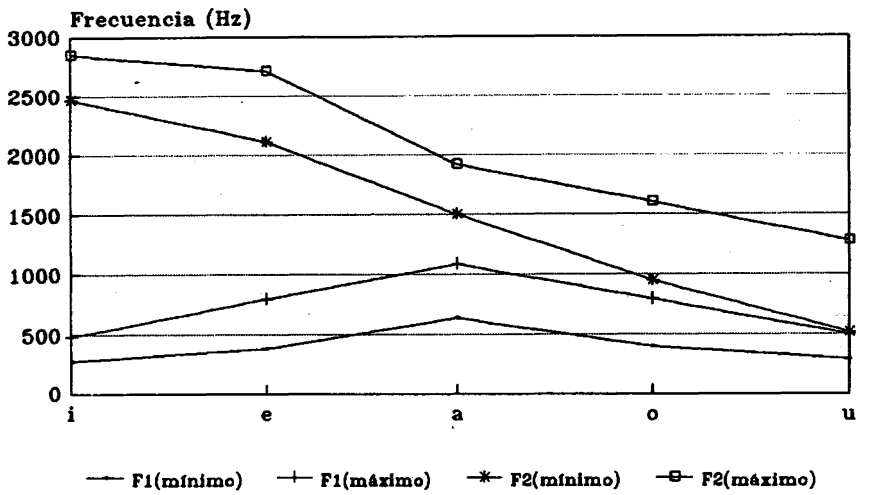


Fig. 5

3.3.3. Función discriminante por el F1

F1	altas	medias	bajas
i,u	97.67	2.33	0
e,o	15.67	81.67	2.67
a	0	8.67	91.33

El F1 discrimina bien el grupo de las vocales altas, pero comete muchos fallos con medias y baja.

3.3.4. Función discriminante con el F2

F2	anteriores	posteriores	centrales
i,e	99.67	0	0.33
u,o	0	94.33	5.67
a	0	0	100

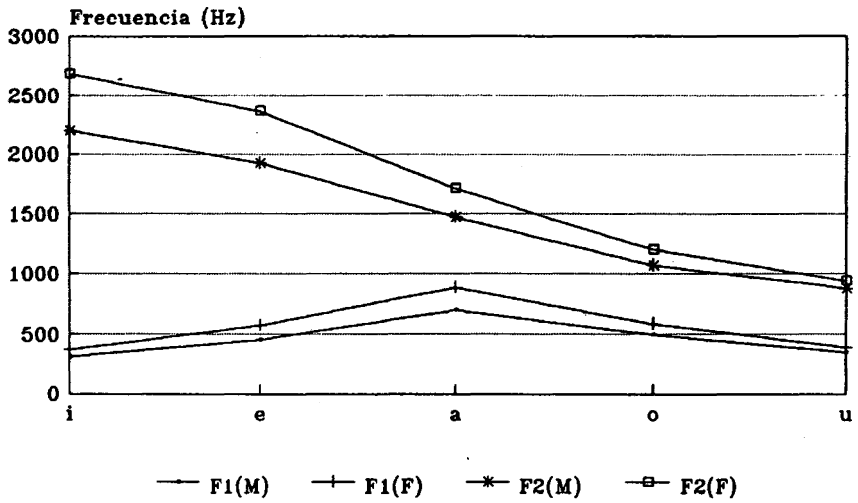
El F2 es un buen discriminante de los tres grupos propuestos. Nuevamente el F2 discrimina mejor que el F1.

3.3.5. Función discriminante de los dos formantes para las cinco vocales.

F1 y F2	i	e	a	o	u
i	100	0	0	0	0
e	5.33	94.67	0	0	0
a	0	0	100	0	0
o	0	0	2	94	4
u	0	0	0	7.33	92.67

3.4. Diferencias entre voz masculina y femenina

La fig. 6 muestra las diferencias entre los formantes masculinos y femeninos para cada vocal.



M=masculino; F=Femenino

Fig. 6

Aplicando la prueba de la t de Student se comprueba que existen diferencias significativas ($p < 0.01$) entre la voz masculina y femenina tanto en el F1 como en el F2.

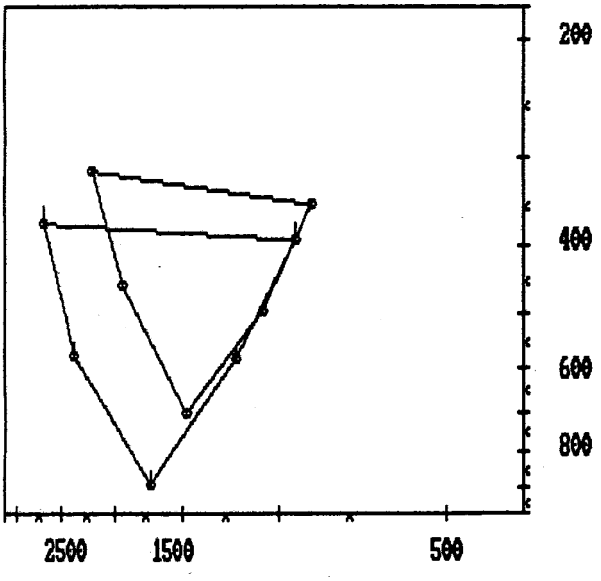


Fig. 7 Carta de formantes con las medias de las vocales masculinas y femeninas.

Función discriminante de los sexos tanto por F1, como por F2:

F1	M	F	F2	M	F
M	96	4	M	97.33	2.67
F	5.33	94.67	F	0	100

El F2 es mejor discriminante de los sexos que el F1, aunque ambos poseen porcentajes de clasificación muy altos.

Vista la diferencia significativa entre ambos sexos quisimos comprobar si existe una relación mutua entre F1 y F2 masculino-femenino y si podríamos predecir uno a partir del otro. Para lo primero se lleva a cabo una correlación estadística y para lo segundo se obtiene la regresión, que proporciona una fórmula que predice los valores formánticos de un sexo a partir de los datos del otro.

Coeficiente de Correlación de los F2: R = 0.97; 94.17%
Regresión: y = 1.278x - 148

Se comprueba nuevamente que F2 está más altamente correlacionado que F1. La fórmula de la regresión permite la predicción o estimación; por ejemplo, en la tabla siguiente se presenta la media real de los F2 masculinos, la frecuencia estimada de los F2 femeninos obtenidos con la aplicación de la fórmula y se compara con la media real de los F2 femeninos: el error de estimación en Hz no es demasiado elevado en ningún caso. Son errores asumibles.

F2

media real M	estimada F	media real F	error est.
[i] 2200	2664	2685	-21
[e] 1926	2313	2367	-54
[a] 1471	1732	1712	20
[o] 1070	1219	1201	18
[u] 877	973	937	36

Si se desea establecer la estimación a partir de los valores femeninos tendríamos exactamente lo mismo despejando x: x = y/1.278 + 148; por ejemplo, 2685/1.278 + 148 = 2249.

4. DISCUSIÓN

4.1. Sobre el análisis.

Podemos asegurar que nuestros datos de las vocales masculinas coinciden de forma sensible con los presentados para el español con anterioridad por Delattre (1965), Quilis y Esgueva (1983) y Martínez Celdrán (1984); aunque existen diferencias en los datos puntuales, que son sólo medias; está claro que los valores presentados por esos autores entran dentro de los campos de dispersión que hemos establecido. Si a tales medias les aplicamos las tres reglas de reconocimiento en todos los casos serían reconocidas las vocales correspondientes. No obstante, el F2 de [a] es el más distante de todos los datos, ya que hemos encontrado una media de 1471 Hz frente a la de Quilis y Esgueva 1220 Hz; Delattre,

1300 Hz y Martínez Celdrán, 1230 Hz (datos de vocales tónicas fundamentalmente). Es probable que nuestra media actual esté sesgada hacia la zona alta del campo de dispersión. Los datos ofrecidos por R. Monroy (1980), por el contrario, no coinciden con todos los proporcionados por los autores citados, ni con los actuales que nosotros hemos presentado, sobre todo en los formantes primeros de las vocales altas y medias. 482 Hz para [i] y 490 Hz para [u] son realmente imposibles. Pienso que aquí hubo un error de análisis en el sentido de que el primer formante se midió no en el centro sino en la parte final. Hay que tener en cuenta que los instrumentos se han perfeccionado y ahora podemos estar más seguros de nuestros datos gracias a técnicas complementarias como la de LPC. Si comparamos con datos proporcionados para otras lenguas próximas veremos que coinciden con los nuestros, no con los de Monroy:

francés: 240 para [i] y [u] (Malmberg, 1974).
catalán: 258 para [i] y 338 para [u] (Martí, 1984)
italiano: 250 para [i] y [u] (Canepari, 1979).

Hablamos de coincidencia siempre que los valores entren dentro de nuestro campo de dispersión. Los datos de Monroy no entran dentro de esos campos. Tampoco los valores de F1 de [e] y [o] son demasiado acertados. Son más elevados que los proporcionados para las abiertas del francés, catálán e italiano, aunque nuestras vocales [e] y [o] son medias comparadas con las de esas lenguas que distinguen entre abiertas y cerradas.

Aunque nuestro estudio se ha basado sólo en vocales tónicas, se puede comprobar que los valores de átonas proporcionados por Quilis y Esgueva (1983) y por Martínez Celdrán (1984) entran dentro de los campos de dispersión establecidos gracias a las reglas de reconocimiento que hemos formulado.

Casi todos los autores que proporcionan datos de formantes de forma general se basan en hablantes masculinos. Para el español peninsular, que sepamos, sólo Quilis y Esgueva proporcionaron valores para los formantes femeninos. En este caso las diferencias con los nuestros son considerables: todos los formantes primeros son mucho más bajos que los nuestros. Es curioso pero los F1 de [i] y [u], incluso [a], son más bajos en la voz femenina que en la masculina que ellos mismos proporcionan. En los F2 que presentan, en el caso de [a] y [u] son más bajos en la voz femenina que en la masculina. Nuestros datos son más sistemáticos, pues los de la voz femenina son siempre un poco más elevados que los de la voz masculina, lo cual parece lógico dada la mayor

altura del F0 en la voz femenina y supuesto que las cavidades del tracto vocal femenino suelen ser un poco más reducidas, por regla general.

4.2. Sobre el reconocimiento

Aunque los ingenieros e informáticos están utilizando otras metodologías en el Reconocimiento Automático del Habla (Casacuberta y Vidal, 1987), que tienen poco en cuenta las teorías lingüísticas, nosotros pensamos que es posible el reconocimiento a partir de lo que sabemos por la lingüística. No faltan autores que siguen métodos semejantes a los nuestros (É. Giraud, 1992; Espy-Wilson, 1994). Hemos comenzado por el estudio de las vocales, pero nuestra pretensión es extenderlo a las consonantes también. Como se ha visto con las vocales el éxito ha sido muy considerable al superar el 90% de aciertos el reconocimiento efectuado.

Quizás merezca la pena destacar la diferencia en el uso de los rasgos para clasificar las vocales que hemos utilizado. Los generativistas (Harris, 1969) utilizan por regla general los rasgos: **alto, bajo, posterior y redondeado.** Nosotros hemos cambiado y hemos utilizado el central en vez del bajo y el anterior en vez del posterior y hemos prescindido del redondeado, pues era innecesario. Esto se ha debido a que nuestro objetivo era el reconocimiento y al hecho de que el F2 se ha mostrado más relevante que el F1. El rasgo bajo viene determinado por el F1, en cambio el central está determinado por el F2. En nuestro caso ha parecido más discriminante el F2 y, por tanto, el uso del rasgo central. Las vocales anteriores se han mostrado estadísticamente como un grupo mejor definido que las posteriores. En el uso generativista [a] queda incluida en las posteriores, pero los datos no favorecen esta asunción. Para el reconocimiento ha sido mucho más discriminante utilizar el rasgo central y el anterior. Obsérvese que así se obvian también las incoherencias al especificar la vocal [a] como [+central] y [-anterior]. Incluso si se hace una ordenación crítica de las reglas de manera que la regla de las anteriores (nº 2) se aplique antes que la de las centrales (nº3), ésta no será necesario aplicarla cuando la vocal sea [+ant], puesto que toda [+ant] es necesariamente [-central], con lo cual se elimina redundancia y se economiza.

5. CONCLUSIONES

Es evidente que con este estudio hemos conseguido probar objetivamente mediante pruebas estadísticas que los dos primeros formantes son suficientes para caracterizar el timbre de las cinco vocales españolas. Además también se ha probado que el F1 es el responsable de la agrupación de las vocales en **altas, medias y bajas** y el F2 en **anteriores, centrales y posteriores**. Hemos comprobado que el F2 tiene mayor importancia en la discriminación de las vocales que el F1. Hemos probado también que las vocales femeninas son ligeramente más altas que las masculinas de forma sistemática y hemos hallado una fórmula para cada formante que puede estimar los valores de los formantes de un sexo a partir de los datos formánticos del sexo opuesto. Además, se ha mostrado también una posibilidad de efectuar reconocimiento de habla más acorde con las teorías lingüísticas, al basarse en rasgos fónicos.

6. REFERENCIAS

CANEPARI, L. (1979): *Introduzione alla fonetica*, Turín, Einaudi.

CASACUBERTA, F. y VIDAL, E. (1987): *Reconocimiento automático del habla*, Barcelona, Marcombo.

DELATTRE, P. (1965): *Comparing the Phonetic Features of English, French German and Spanish*, Heidelberg, Chilton, Groos.

ESPY-WILSON, C. Y. (1994): "A feature-based semivowel recognition system", *JASA*, 96(1), 65-72.

GIRAUD, É. (1992): "Système expérimental en Reconnaissance Automatique de la Parole", *Revue de Phonétique Appliquée*, 105, 299-317.

HARRIS, J. W. (1969): *Fonología generativa del español*, Barcelona, Planeta, 1975.

LIEBERMAN, P. (1977): *Speech Physiology and Acoustics Phonetics*, Nueva York, MacMillan Publishing.

MALMBERG, B. (1974): *Manuel de Phonétique Générale*, París, Picard.

MARTÍ, J. (1984): "Paràmetres vocàlics del català", *Folia Phonetica, 1*, Lleida, Laboratori de Fonètica Pere Barnils, 23-43.

MARTÍNEZ CELDRÁN, E. (1984): *Fonética*, Barcelona, Teide.

MONROY, R. (1980): *Aspectos fonéticos de las vocales españolas*, Madrid, SGEL.

QUILIS, A. Y ESGUEVA M. (1983): "Realización de los fonemas vocálicos españoles en posición fonética normal", en M. Esgueva y M. Cantarero, *Estudios de Fonética I*, Madrid, CSIC.

STEVENS, K. N. (1972): "Quantal nature of speech", en E.E. David Jr. & P. B. Denes (Eds.), *Human comunication: A unified view*, Nueva York, McGraw-Hill, 51-66.

STEVENS, K.N. (1989): "On the quantal nature of speech", *Journal of phonetics*, 17, 3-45.

TATHAM, M. (1990), "Cognitive phonetics", en *Advances in Speech, Hearing and Language Processing*, vol. 1, Londres, JAI Press, 193-218.