

**ESTUDIO DEL CAMPO DE DISPERSION DE LAS
VOCALES CASTELLANAS**

ANA M^a FERNÁNDEZ PLANAS
Universidad de Barcelona

RESUMEN

Este artículo estudia el campo de dispersión de las cinco vocales castellanas creadas mediante síntesis de voz desde el punto de vista de la percepción. Los resultados muestran principalmente que el ámbito perceptivo (que guarda una estrecha relación con la atención prestada por el oyente) es mayor que el productivo y que las frecuencias de F1 y de F2 con mayores índices de identificación son más altas que las establecidas en estudios de fonética acústica.

Palabras Clave: campo de dispersión, síntesis de voz, percepción.

ABSTRACT

The aim of this paper is the study of the distribution of the frequencies that are characteristic of the five Spanish vowels. For this purpose we synthesized several vowels and perception tests were carried out. We found that speakers' competence was higher in the field of perception (above all if they payed careful attention) than in production. We also found that the F1 and F2 stimulus with the highest identification rates had higher values than the ones that are usually found in acoustic studies.

Key words: Frequency distribution, speech synthesis, perception.

1. INTRODUCCION. OBJETIVO

El objetivo principal de este experimento consiste en establecer el campo de dispersión de las vocales castellanas desde el punto de vista perceptivo a partir de sonidos creados mediante la síntesis de voz combinando los dos primeros parámetros formánticos (F1 y F2).

2. INSTRUMENTOS UTILIZADOS

Cada estímulo sonoro fue creado con un sintetizador de voz PCF 8200 de Philips del Laboratorio de Fonética de la Universidad de Barcelona conectado a un ordenador personal PC AT. El programa editor de unidades fue el llamado VOXSY en su versión 1.4 confeccionado por Tevox.

Los parámetros que admitía el sintetizador y que había que controlar son los siguientes: cinco formantes (F1-F5) con sus correspondientes anchos de banda (B1-B5), amplitud (AM), entonación (ENT), duración (D), número de tramas y tono fundamental (F0) inicial.

Para establecer el orden aleatorio en el que los estímulos debían aparecer en el test perceptivo se empleó el programa NUMALE creado por E. Martínez Celdrán.

Para la grabación y reproducción en la encuesta de los estímulos una vez creados se usó una grabadora y un amplificador de sonido Marantz modelos CP430 y PM200 respectivamente.

Finalmente, para la elaboración de las cartas de formantes que aparecen en el apartado VI del presente trabajo se usó el conocido programa para gráficos HARVARD GRAPHICS versión 2.30 por ser el que mejor se adaptaba, entre los programas a mi disposición, a las necesidades de manejo de tan gran acopio de datos.

3. MÉTODO

En primer lugar, tuvieron que ser creados los estímulos auditivos. Para ello decidí variar exclusivamente F1 y F2 que son los que dan, fundamentalmente, el timbre vocálico y mantener constantes los restantes parámetros. Mantuve un F0 inicial de 120 Hz. en todos los ficheros y utilicé las cantidades siguientes seleccionadas de entre las posibilidades que el sintetizador ofrecía para cada uno de estos parámetros:

F1	F2	F3	F4	F5	AM
109	500	3063	4500	4600	8
168	646				9
258	754				10
282	835				11
335	925				12
365	974				13
398	1079				14
433	1136				14
472	1258				14
515	1324				14
561	1394				14
612	1467				14
667	1545				14
727	1626				14
793	1711				13
864	1801				12
942	1896				11
1027	1996				10
1119	2101				9
	2211				8
	2328				
	2450				

B1	B2	B3	B4	B5	ENT	D
303	433	60	265	800	0	0

Los datos de los parámetros cuyo valor permanece inalterable en el experimento fueron escogidos tras varias pruebas iniciales por ser los que mejor resaltaban, a mi entender, los dos parámetros que más interesaba: F1 y F2.

El valor en ENT ha permanecido en cero ya que al referirse a aspectos suprasedgmentales no tenía cabida en un estudio de las características del que nos ocupa. También el valor de D pues la duración de cada estímulo ya se había establecido en el número de tramas de cada uno: veinte tramas lo que equivale a un tiempo de 160 ms.

En F1 y F2, los parámetros cambiantes, no se han tenido en cuenta todas las posibilidades que el sintetizador ofrece por dos razones principales:

1) Algunos datos están tan próximos a sus contiguos, ya sean anteriores o posteriores, que la diferencia auditiva que en todo caso conllevarían sería no relevante. Por ejemplo un dato de F1: 307 Hz entre 282 Hz y 335 Hz ; y un dato de F2: 1195 Hz entre 1136 Hz y 1258 Hz.

2) El número de estímulos que habrían resultado (1024) hubiera sido excesivo a la hora de plantear el test perceptivo a los informantes.

Entre las posibilidades escogidas hubo que hacer una nueva selección porque todas aquellas combinaciones en las que F1 tenía una cantidad de Hz superior a la de F2 tuvieron que ser eliminadas ya que no hubiera sido lógico.

Tras todas estas limitaciones quedaron al fin trescientos setenta y un estímulos. Seguían siendo bastantes para un test auditivo pero restringirlo más hubiera implicado olvidar el propósito fundamental del experimento: estudiar el campo de dispersión de las vocales castellanas atendiendo a las combinaciones de F1 y F2 que ofrece el sintetizador de voz empleado, salvo en los casos ya especificados.

Los estímulos fueron creados y almacenados

como archivos de unidades fonéticas (afo) y como archivos procesados de dichas unidades (apr) que son las dos posibilidades que ofrece el programa VOXSY.

Una vez hecho esto y habiendo ya establecido el orden aleatorio que debían mantener entre sí los estímulos auditivos había que grabarlos en una cinta magnetofónica. Lo hice repitiendo cada sonido dos veces con una pequeña pausa intermedia y guardando un silencio algo mayor entre estímulos de timbre diferente.

La encuesta se realizó en el Laboratorio de Fonética de la Universidad de Barcelona a estudiantes de Filología de primer o segundo curso de dicha Universidad, en grupos de cinco a diez personas hasta un total de treinta. No se distinguió entre informantes de lengua materna catalana o castellana porque no se pretendía buscar diferencias de timbre en las vocales medias ya que la diferencia entre ellas no es fonológica en castellano. Es probable, por lo tanto, dado que el mundo de la percepción es aún tan desconocido, que el resultado de este estudio se refiera al castellano de Cataluña y que sea distinto del de otras zonas de la geografía peninsular o americana. La realización de la encuesta se llevó a cabo con dos breves descansos intermedios, pues el cansancio hacía que la atención de los oyentes disminuyera y empezaran a percibir más estímulos del tipo "no".

Dicha encuesta presentaba la siguiente forma:

	SI	QUIZAS	NO
1			
2			
3			

en ella los informantes debían poner en la casilla

del "sí" aquella de las cinco vocales que oían sin duda acerca de su timbre; si la duda existía, debían poner la vocal que creían en la casilla del "quizás" y en caso de percibir el estímulo como un ruido no asimilable a ninguna de las cinco vocales debían colocar una cruz en la casilla del "no".

Al hacer el primer recuento de resultados observé que varias combinaciones de F1 y F2 que debían entenderse como vocales velares tenían un gran número de respuestas "no". Se podía deber a la amplitud que era alta para tales vocales pero que había permanecido inalterada para no imponer a priori un criterio subjetivo, pues en el ordenador utilizado sucede que siendo el resto de los parámetros constantes las vocales posteriores tienen una intensidad mucho mayor que las demás. Por ello efectué una segunda encuesta compuesta por los treinta y cuatro sonidos que habían tenido como mínimo diez respuestas "no" cambiándoles su amplitud inicial por la siguiente: 3, 3, 4, 4, 5, 5, 6, 6, 7, 8, 8, 7, 7, 6, 6, 5, 5, 4, 4, 3.

Quedaron veinticinco encuestas válidas en este segundo test y los resultados para estos estímulos sustituyeron a los del primero en el recuento general.

4. RESULTADOS

Las encuestas ofrecían once tipos posibles de respuesta a cada estímulo: cinco para "sí", cinco para "quizás" y una para "no". Hubo que hacer un primer recuento de cada una de dichas posibilidades. Seguidamente había que restringirlo a cinco tipos que correspondieran a cada una de las vocales castellanas, (las respuestas "no" aparecían en la mayoría de estímulos pero en cantidades muy bajas y por lo tanto desde este punto no las tuve en cuenta). Para que la diferencia de calidad entre "sí" y "quizás" siguiera manteniéndose establecí que los valores del primer grupo valieran dos puntos y los del segundo uno. Así, el máximo de puntos que un sonido podía tener era cincuenta y equivaldría al 100%.

Los puntos resultantes para cada estímulo variaban entre sí en cantidad. Por esa razón establecí tres niveles de análisis:

1) **Primer nivel.** En él aparecen todos aquellos estímulos que han obtenido un mínimo de cinco puntos en el recuento, es decir, a partir del 10% de puntos posibles o de identificación auditiva. Menos de cinco puntos lo creí insignificante para ser tenido en consideración.

2) **Segundo nivel.** Contiene aquellos estímulos que han obtenido dieciséis puntos como mínimo en el recuento, es decir, a partir de un 32%.

3) **Tercer nivel.** Incluye aquellos estímulos que han obtenido treinta y tres puntos como mínimo en el recuento, es decir, a partir de un 66% lo que se acerca a la llamada "zona estrecha de dispersión" para cada vocal.

Para cada nivel presento carta de formantes de las vocales aisladas y de ellas dos a dos para su comparación.

Primer nivel

Todos los estímulos muestran un amplio campo o zona "ancha de dispersión" -Romero (1988): 183,205- especialmente los identificados con [i] en la zona palatal del triángulo vocálico del español y los identificados con [o] en la zona velar del mismo. Estos dos cubren casi completamente dichas zonas. Los restantes estímulos identificados con [e], [a], [u] se ciñen más exactamente a la zona que teóricamente, según estudios de fonética acústica, les corresponde con algunas incursiones hacia la parte superior del triángulo [e], [a] y hacia la parte inferior [u].

Así, hay zonas de intersección entre todas las vocales entre sí como muestran las cartas de formantes de vocales comparadas. De todas ellas llaman poderosamente la atención las que aparecen entre [i]-[u], [i]-[a] y [u]-[a] ya que son posiciones extremas del triángulo.

Este hecho indica que los límites en las frecuencias a la hora de percibir sonidos del lenguaje, por lo menos las vocales, son mucho más amplios que a la hora de articularlos. Parece también, a juzgar por los límites en que se mueve cada vocal que pesa más su F2 que su F1 a la hora de discriminar en ellas el timbre pues, por lo menos, por F2 se distinguirían tres zonas: palatal central y velar, distinción difícil de alcanzar dando primacía a F1 porque sólo experimenta una ligera variación en las vocales velares.

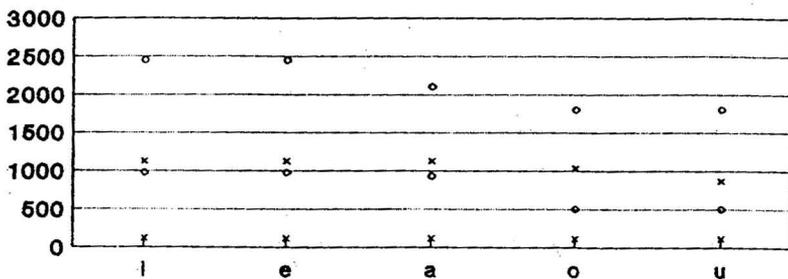
Los límites tanto inferior como superior de ambos formantes son los que aparecen en los siguientes esquemas y gráfico:

F1	i	e	a	o	u
l.inf.	109	109	109	109	109
l.sup.	1119	1119	1119	1027	864

F2	i	e	a	o	u
l.inf.	974	974	925	500	500
l.sup.	2450	2450	2101	1801	1801

l.inf. = límite inferior del formante

l.sup. = límite superior del formante



x = límites del primer formante

o = límites del segundo formante

Segundo nivel

En este nivel las vocales van ciñéndose a la zona que les corresponde. Ya no hallamos zonas de intersección entre [i]-[a] ni entre [u]-[a] lo cual es lógico por ser ubicaciones extremas, sin embargo persiste una pequeña franja común entre [i]-[u], hecho sorprendente pues se trata de posiciones del mismo tipo.

Tampoco hay confluencia de frecuencias entre [e]-[a] ni entre [i]-[o] aunque ambas parejas presentan campos muy próximos.

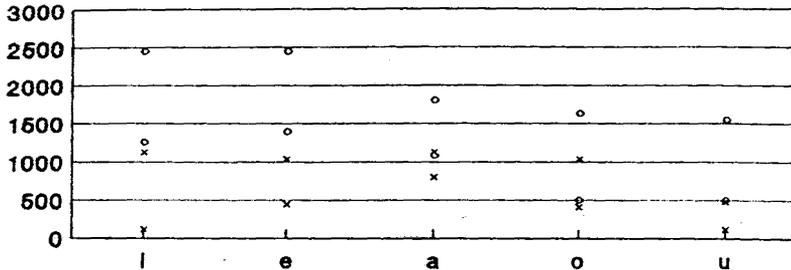
Los campos de dispersión de las parejas restantes muestran zonas de intersección más reducidas que las que aparecían en el primer nivel. Destaca la comparación entre [i]-[e] porque se mantiene en este segundo análisis un grupo de estímulos percibidos como [i] cuyo F1 está cercano a los 1000 Hz. Ello parece volver a indicar la supremacía de F2 a la hora de identificar [i]. Sin embargo, vistos globalmente los límites en que se mueven los F1 y F2 de las cinco vocales parece que es fundamentalmente el primer formante el que delimita el timbre de las mismas pues el segundo es prácticamente el mismo en las dos vocales de la serie palatal y también el mismo en las de la serie velar.

Los esquemas y el gráfico que ilustran estos límites son los siguientes:

F1	i	e	a	o	u
l.inf.	109	433	793	398	109
l.sup.	1119	1027	1119	1027	472

F2	i	e	a	o	u
l.inf.	1258	1394	1079	500	500
l.sup.	2450	2450	1801	1626	1545

l.inf. = límite inferior del formante
l.sup. = límite superior del formante



x = límites del primer formante
o = límites del segundo formante

Tercer nivel

En este nivel de análisis ya no se dan intersecciones entre las percepciones de las distintas vocales. Cada una de ellas tiene su campo de dispersión bien delimitado por claros márgenes de seguridad entre ellos.

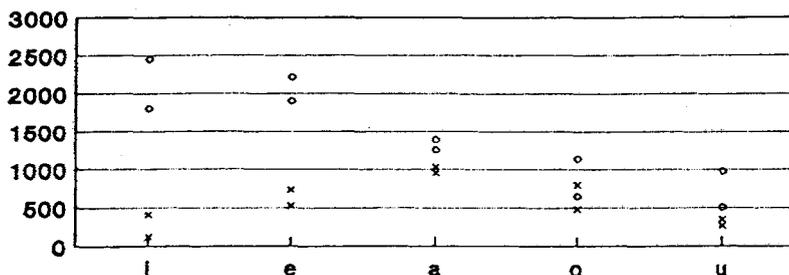
Observando los límites de F1 y de F2 parece que sea F1, en general, más relevante a la hora de establecer diferencias tímbricas entre ellas pues la diferencia en Hz entre el límite inferior y el superior es menor que en F2; sin embargo, los primeros formantes son muy parecidos entre las vocales medias y entre las vocales cerradas, por eso cabe decir que es, en realidad, la combinación de F1 y F2 quien provoca distinción de timbre entre los sonidos vocálicos, por lo menos en un campo o zona "estrecha de dispersión" -Romero (1988): 183-205.

Los esquemas y el gráfico que siguen ilustran este hecho:

F1	i	e	a	o	u
l.inf.	109	515	942	472	258
l.sup.	398	727	1027	793	335

F2	i	e	a	o	u
l.inf.	1801	1896	1258	646	500
l.sup.	2450	2211	1394	1136	974

l.inf. = límite inferior de los formantes
 l.sup. = límite superior de los formantes



x = límites del primer formante
 o = límites del segundo formante

5. DISCUSION

En general cabe destacar que el campo de dispersión de las vocales castellanas es más amplio desde el punto perceptivo que desde el articulatorio, estoy de acuerdo con J.Romero (1988): en este punto. En el terreno de la audición se dan grandes diferencias entre distintas personas, a veces en grado suficiente para ser representadas en los gráficos. Ejemplos de ello los tenemos en el primer nivel de cada vocal, especialmente en [e], [a], [o], [u] pues se pueden observar puntos aislados del bloque en sus cartas. De todas formas, prevalece una estructura auditiva común que es la que permite establecer

niveles de análisis tal y como están planteados en este experimento.

Centrándonos en el tercero hay que subrayar la escasa identificación con [a] de los estímulos presentados, constituyen sólo un 3,48% de los que integran este nivel, y su inclinación hacia la zona velar más que hacia la palatal. Por el contrario, la identificación con [i] ha sido la mayor, representa un 39,53%, y presumiblemente si el sintetizador hubiera permitido algún dato más para F2 desde 2450 Hz el porcentaje hubiera sido mayor. Este factor corrobora la idea expuesta al inicio de este apartado acerca de la percepción.

Respecto a los valores que da E.Martínez Celdrán (1986): para las vocales castellanas no contextuales hallo diferencias especialmente en [a], en [i] y también en [e]. Las citadas vocales, en el tercer nivel de este estudio, el de mayor especificación, presentan tanto en F1 como en F2 frecuencias más altas con la salvedad de la extensión del primer formante de [i] hacia frecuencias de 109 Hz y 168 Hz considerablemente más bajas.

Comparando los resultados obtenidos con los de J.Romero (1988): hay que señalar también la mayor frecuencia en este estudio, salvo en [i], del límite inferior de F1; no así en el límite superior donde son sus resultados ligeramente más altos aunque la diferencia es mínima. Respecto a F2 los datos del presente trabajo siguen siendo más altos en el límite inferior para [i], [e], [a] y algo menores para [o], [u]. En el límite superior mis resultados son más bajos para todas las vocales y hay diferencias significativas en [a], [o], [u]. Por lo tanto, los resultados que nos ocupan presentan mayor especificidad, mayor restricción, del campo estrecho de dispersión de las vocales del castellano que los que ofreció J.Romero.

Los valores que presenta R.Monroy (1980): tienen mayor frecuencia que los de este estudio para las vocales cerradas en su F1 y para la vocal media velar en su F2; por el contrario, las frecuencias que establece Monroy para el F1 de la vocal abierta y para el F2 de la vocal media

palatal son menores que las que resultan del presente experimento, siempre tomando en consideración su tercer nivel de análisis.

6. CONCLUSIONES

1) El campo de dispersión desde el punto de vista perceptivo es mayor que desde el punto de vista de la producción.

2) En el ámbito de la percepción se dan diferencias personales que pueden llegar a ser relevantes, sin embargo, prevalecen valores específicos comunes.

3) El cansancio influye en la percepción de los sonidos.

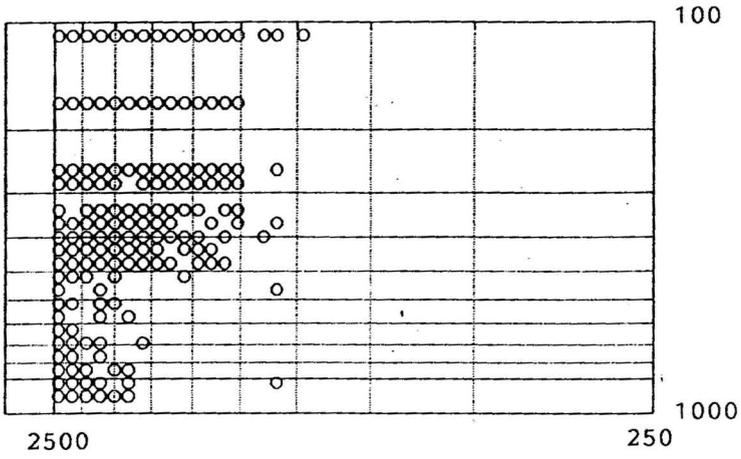
4) Es la combinación de los dos primeros formantes la que establece el timbre vocálico.

5) Las frecuencias de las vocales sintéticas no contextuales son, generalmente, más altas, en los casos de mayor identificación, que los establecidos en estudios de fonética acústica.

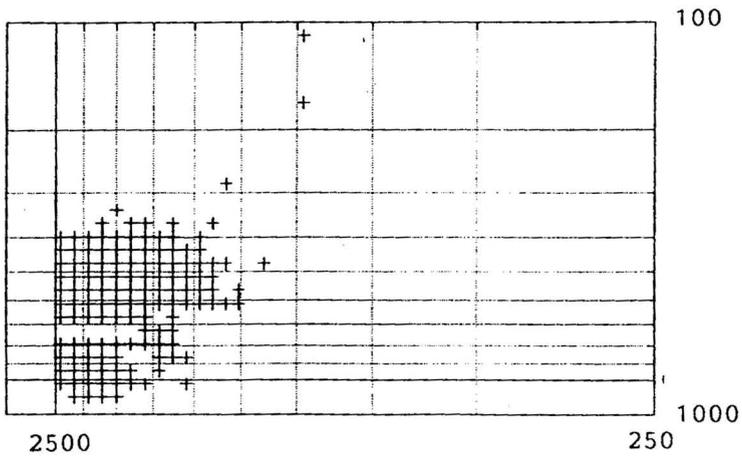
7. GRAFICOS

1. Primer nivel

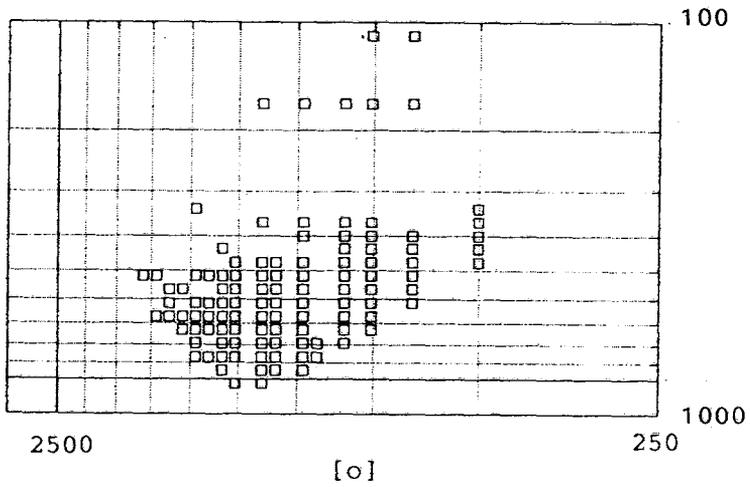
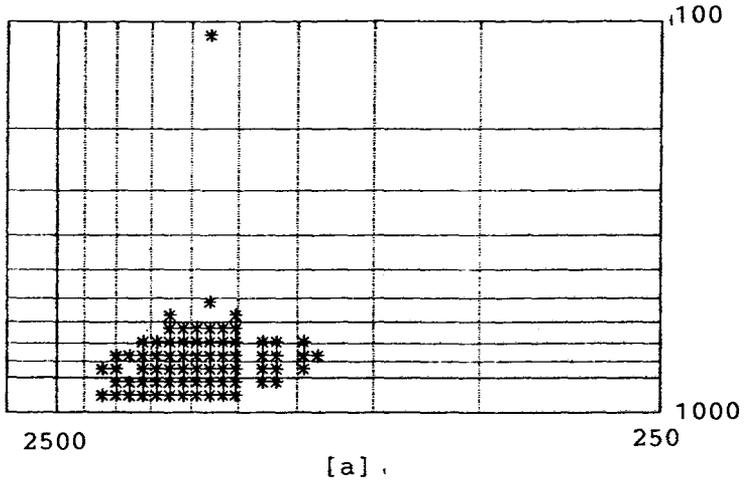
1.1. Vocales aisladas

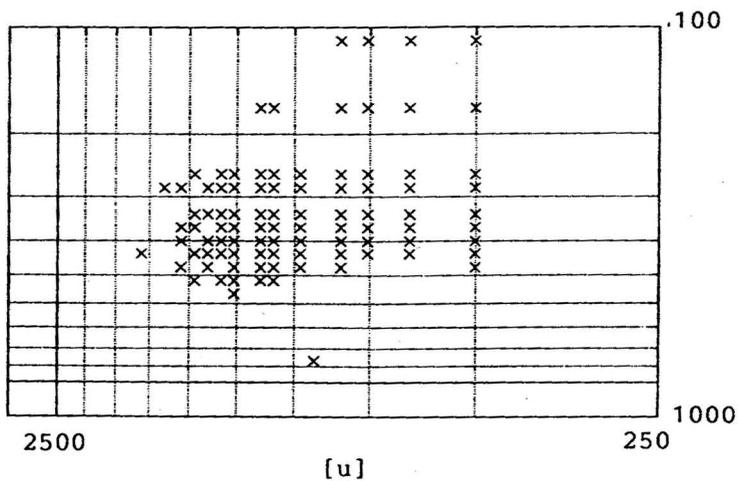


[i]

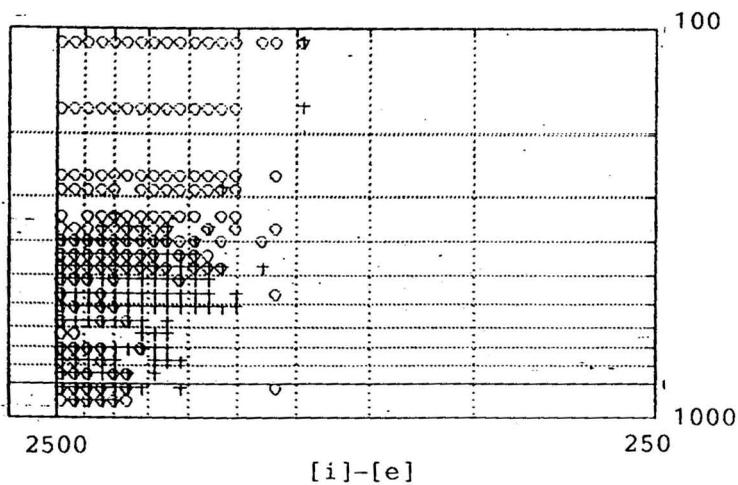


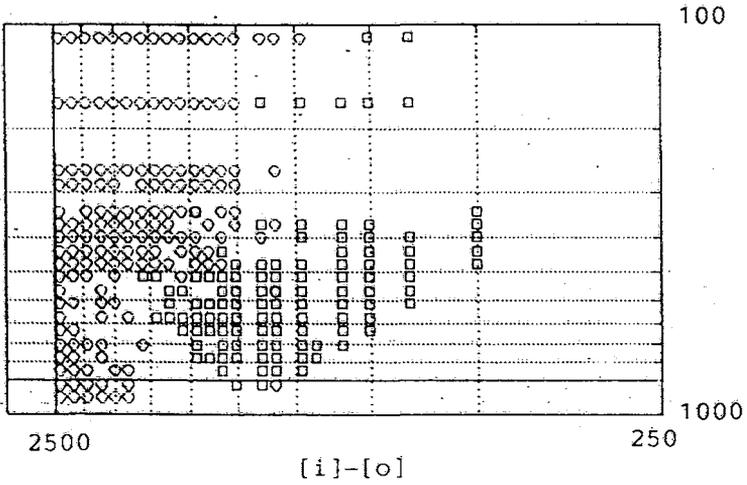
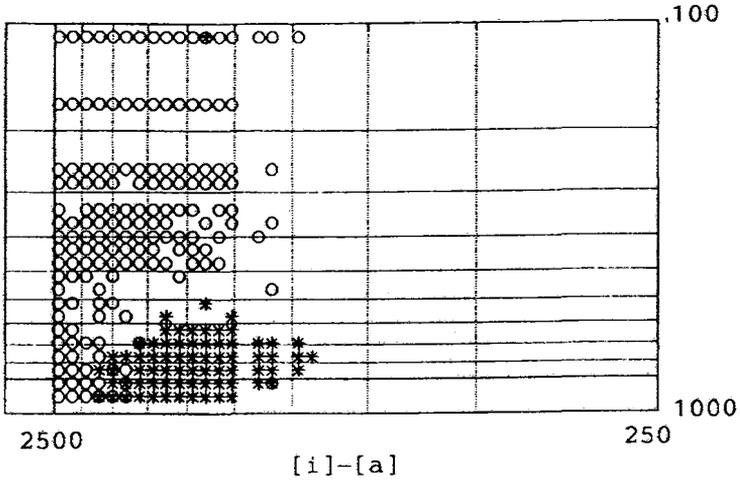
[e]

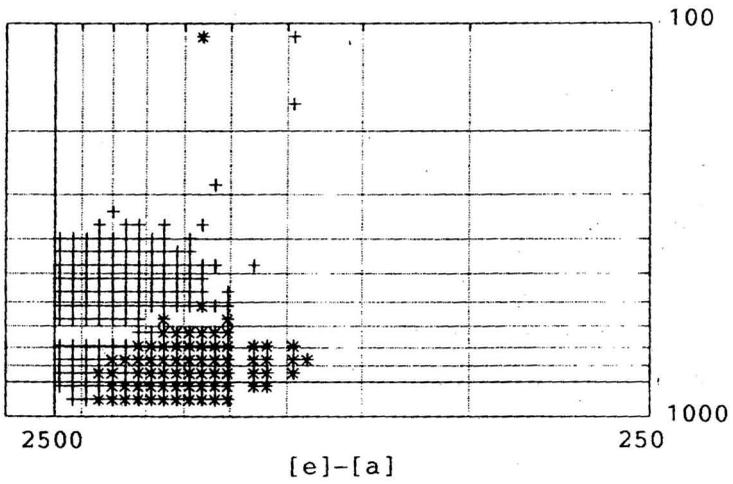
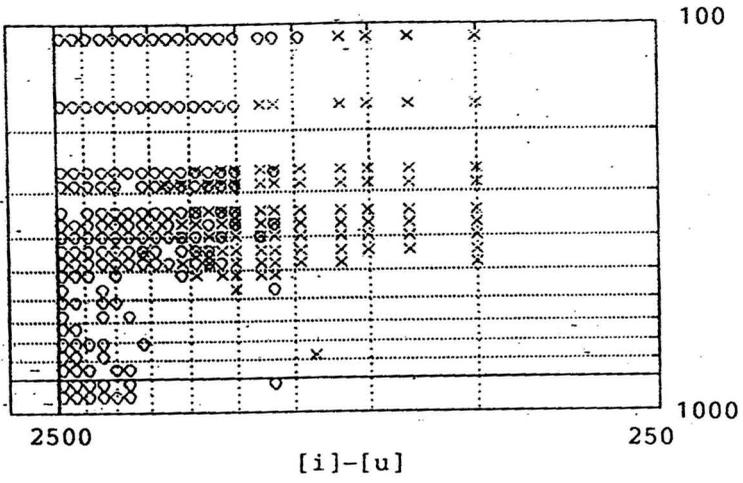


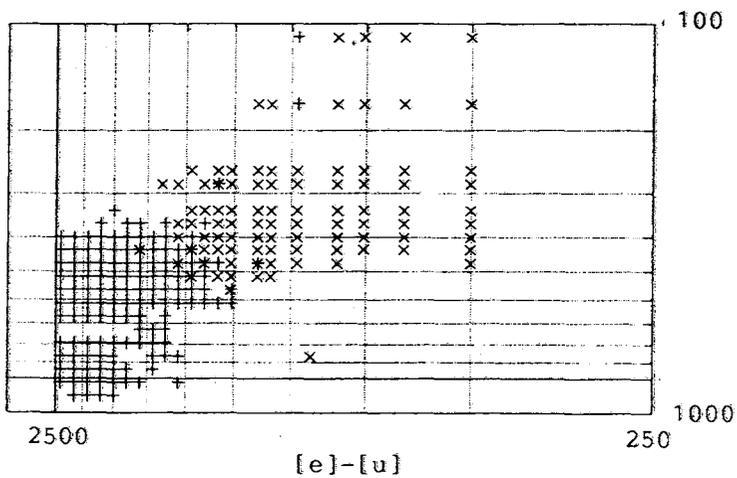
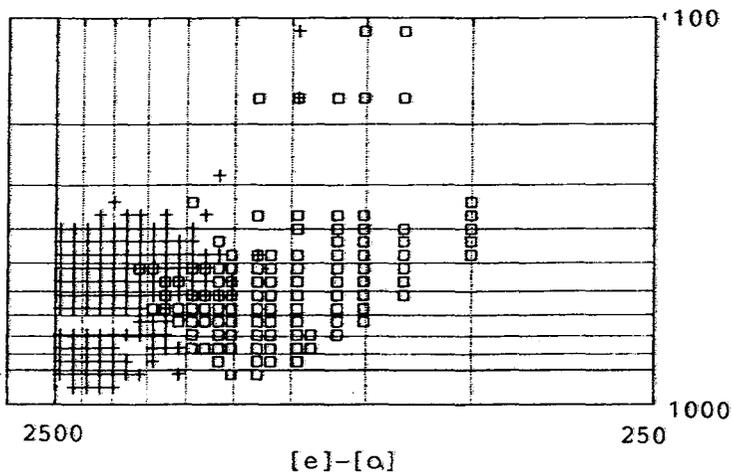


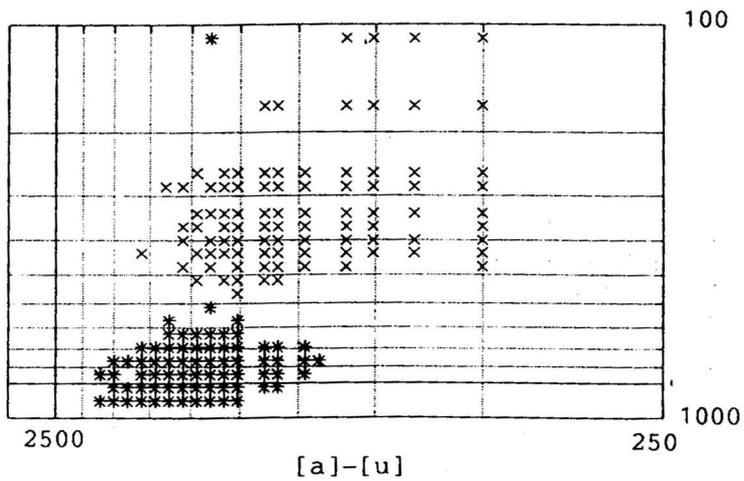
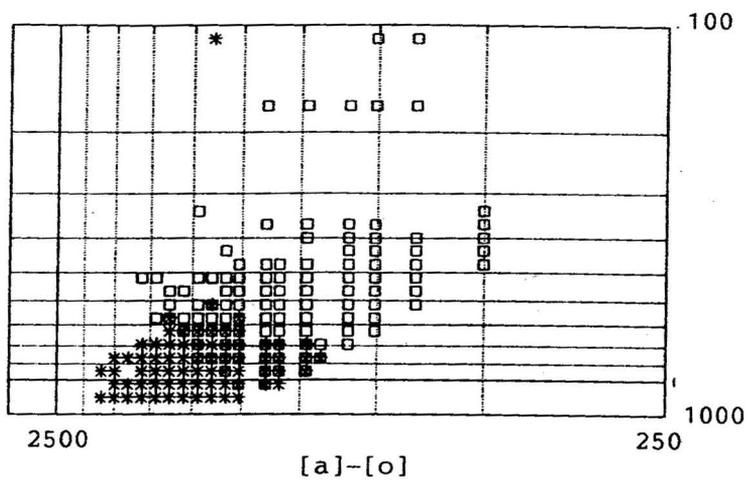
1.2. Comparación entre vocales

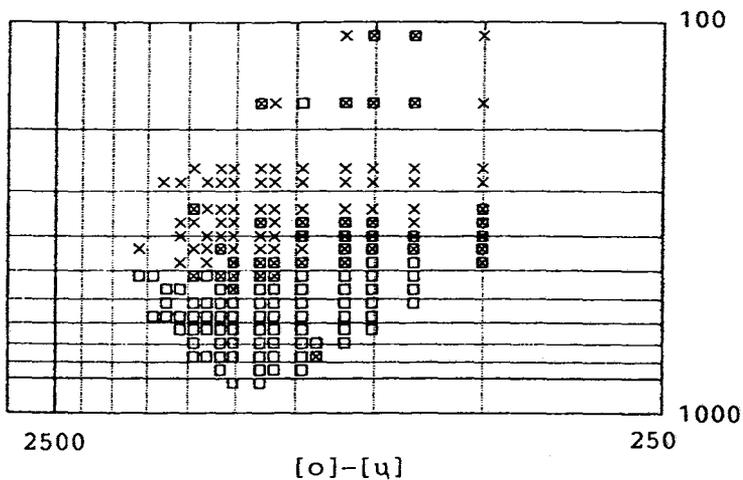






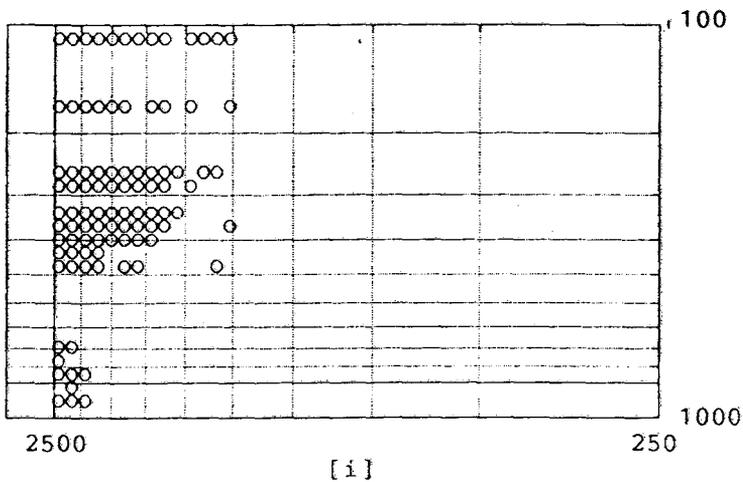


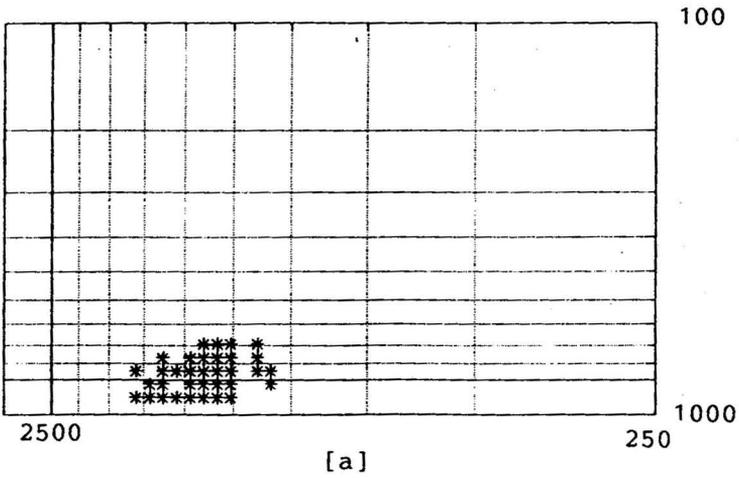
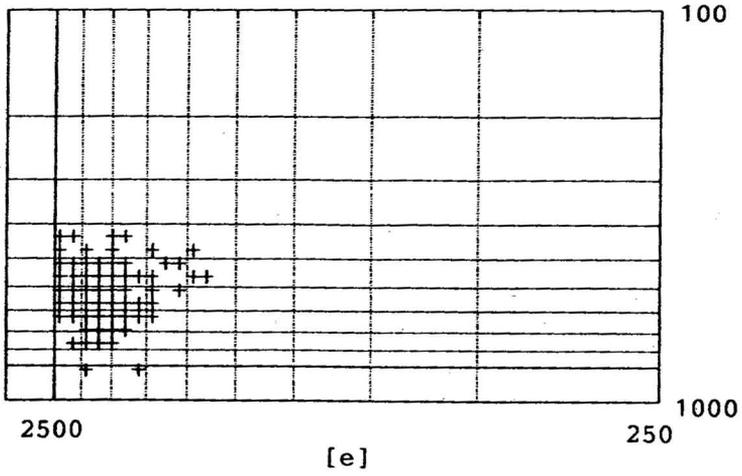


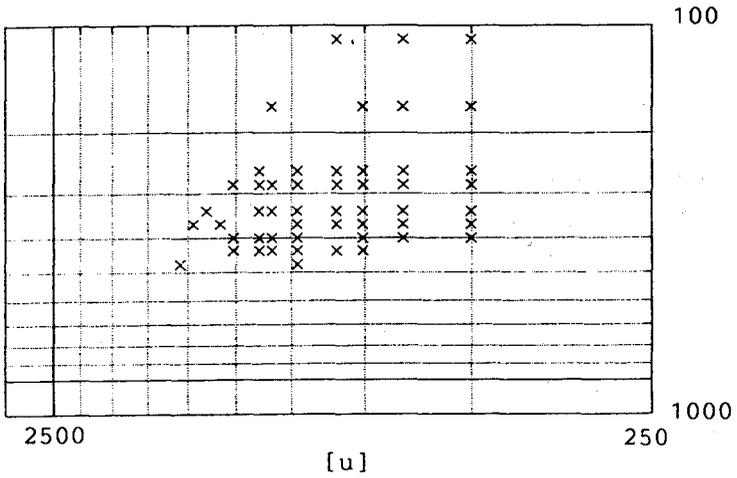
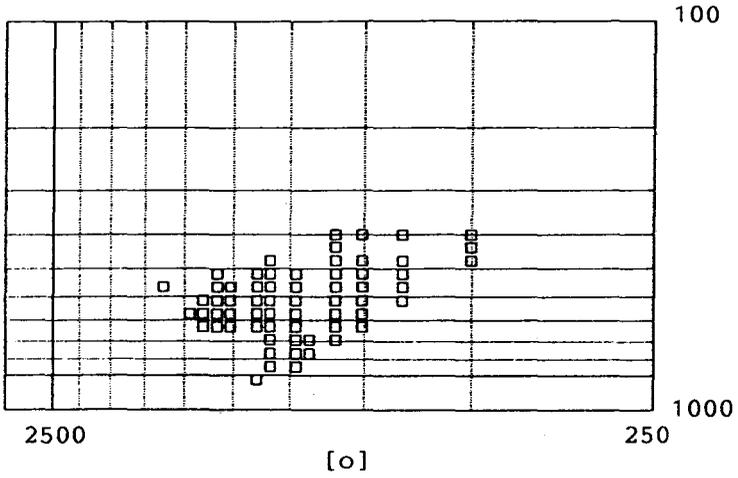


2. Segundo nivel

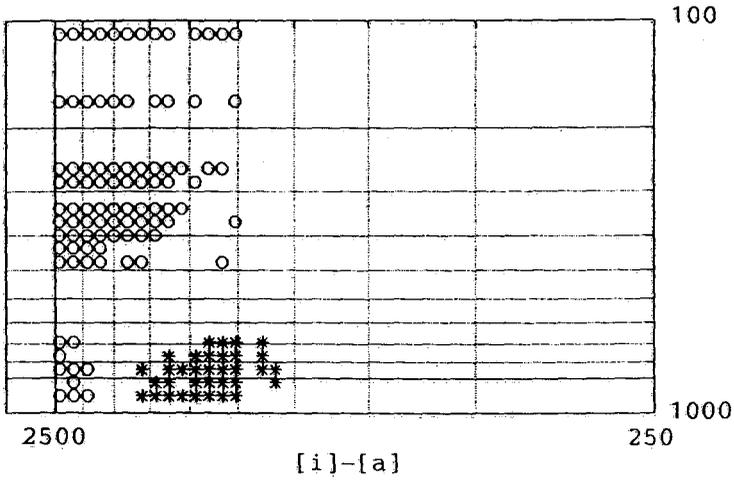
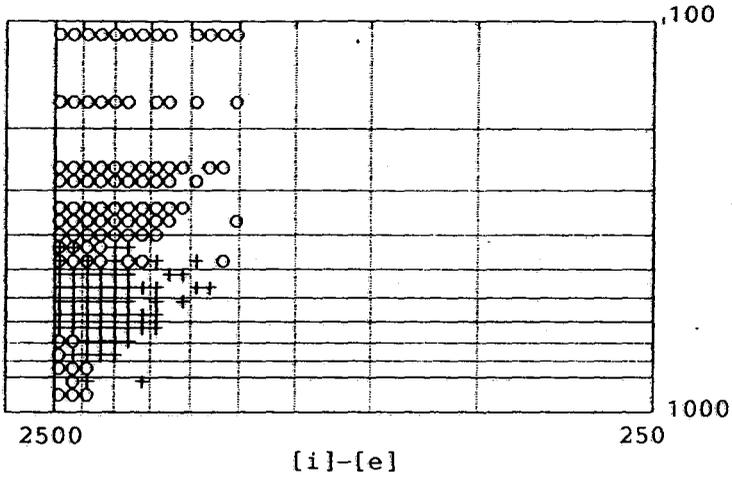
2.1. Vocales aisladas

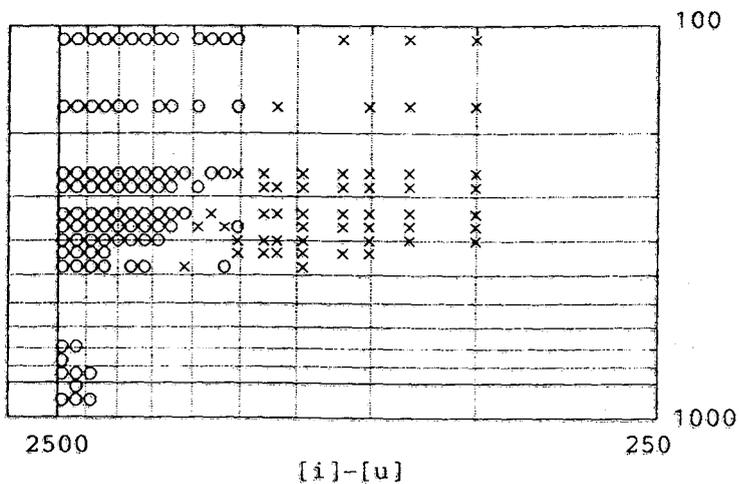
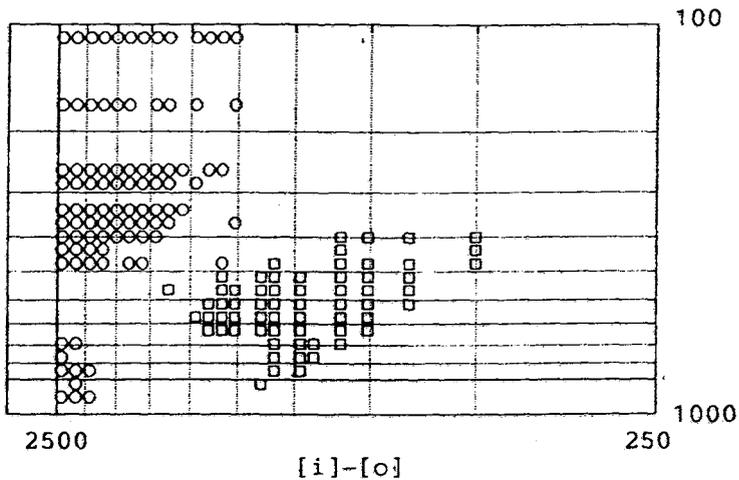


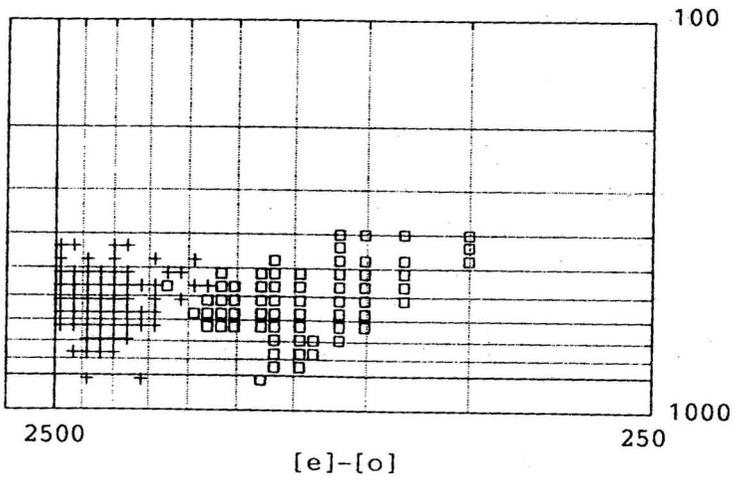
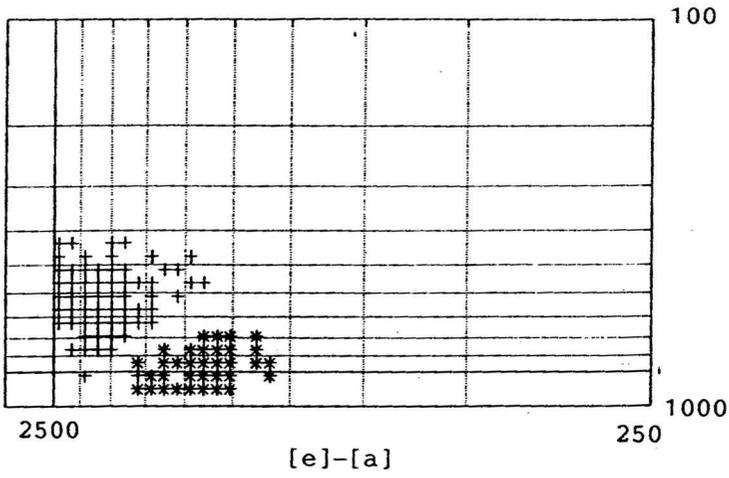


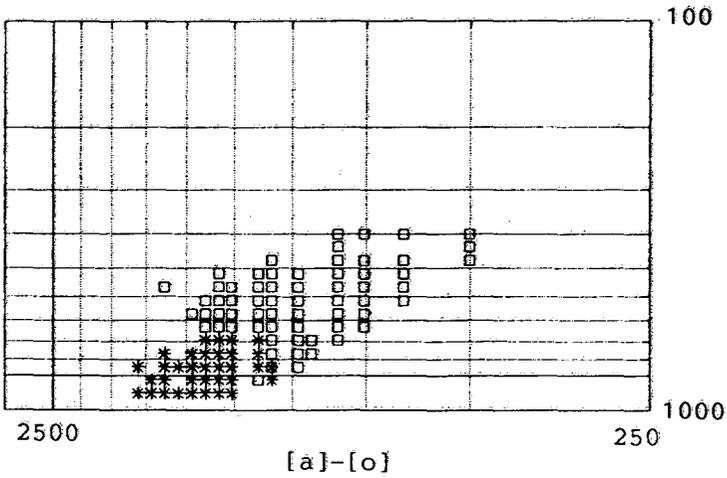
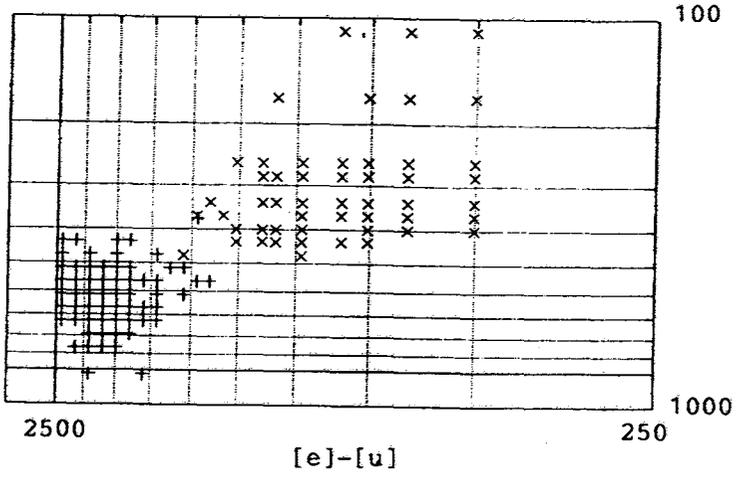


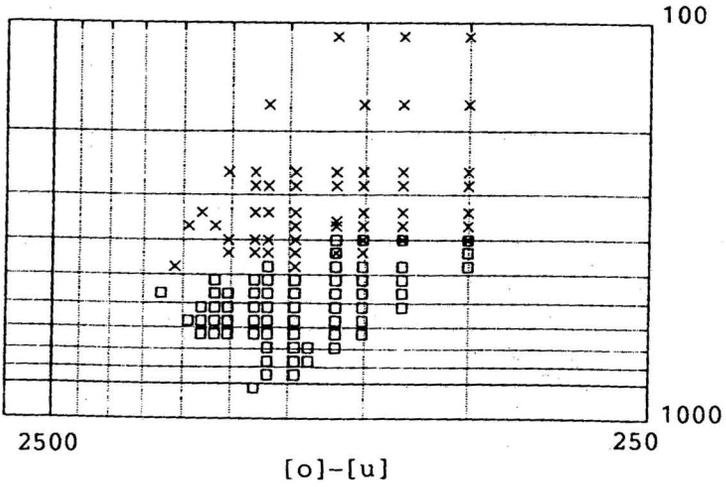
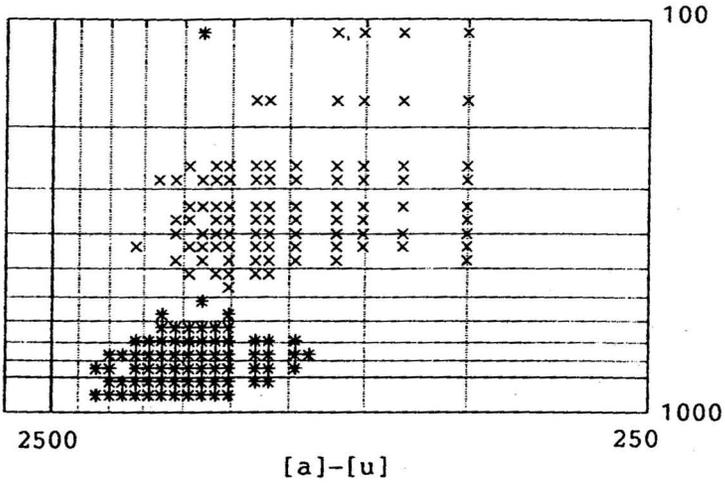
2.1. Comparación entre vocales





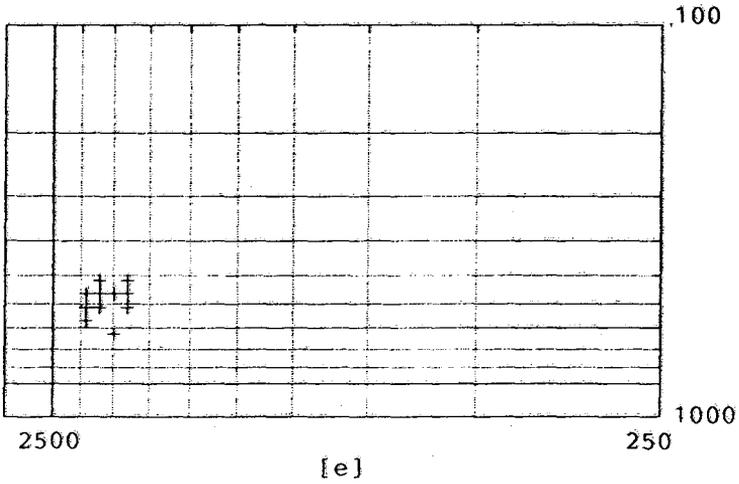
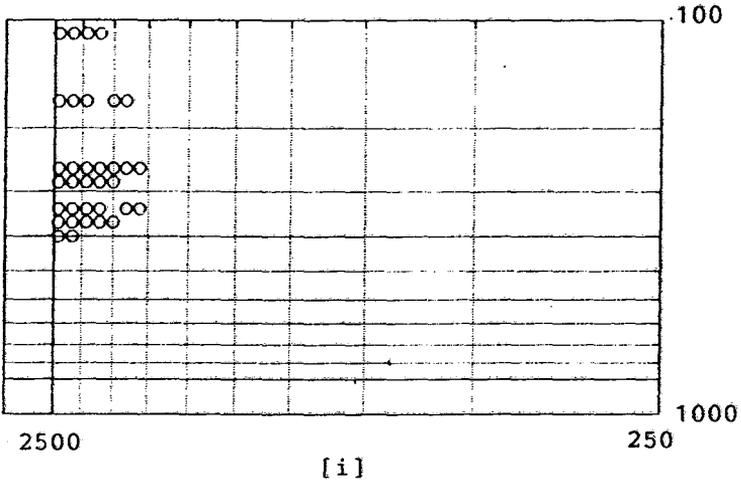


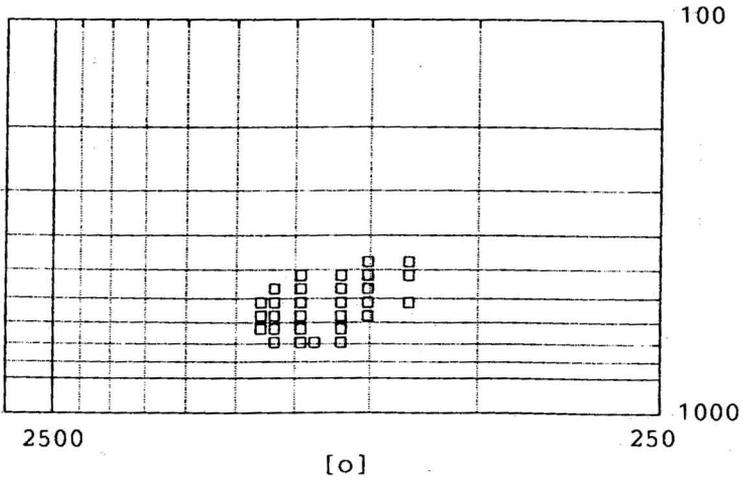
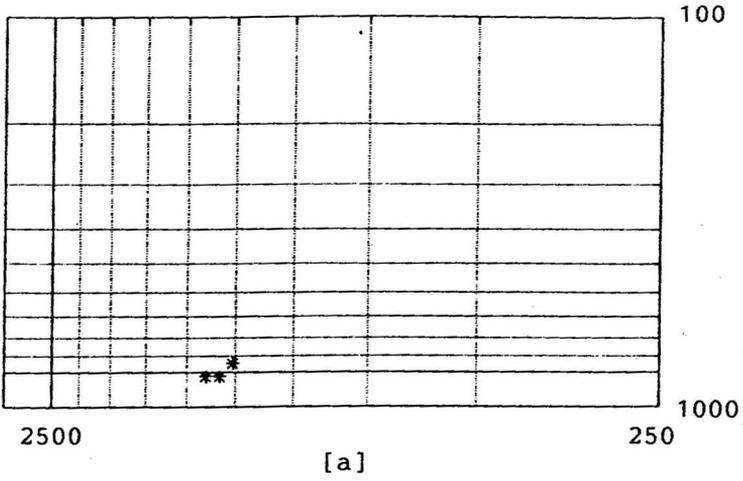


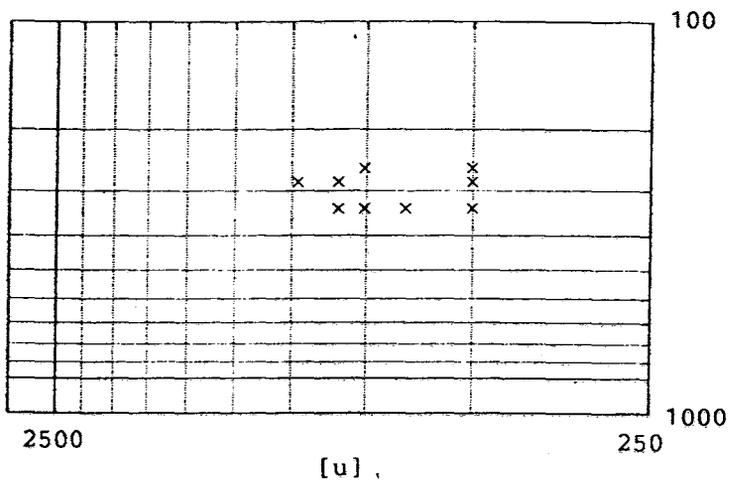


3. Tercer nivel

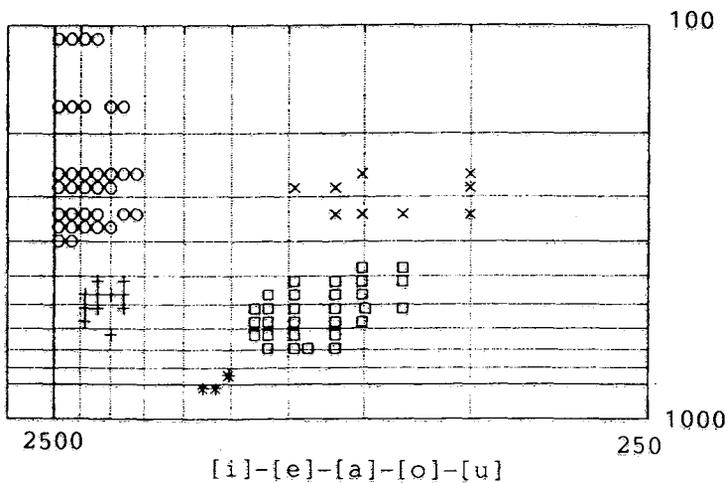
3.1. Vocales aisladas







3.2. Comparación entre vocales



8. BIBLIOGRAFIA

- MARTINEZ CELDRAN, E. 1986. *Fonética*, Teide, Barcelona.
- MARTINEZ CELDRAN, E. 1991. *Fonética experimental: Teoría y práctica*, Síntesis, Madrid.
- MONROY CASAS, R. 1980. *Aspectos fonéticos de las vocales españolas*, SGEL, Madrid.
- PUIG I RIERA, J. y FREIXA I AYMERICH, J. 1990. "El camp de dispersió de les vocals catalanes des del punt de vista de la percepció", *Estudios de Fonética Experimental*, IV, Promociones Publicaciones Universitarias, Barcelona.
- ROMERO, J. 1988. "Campos de dispersión auditivos de las vocales del castellano. Percepción de las vocales", *Estudios de Fonética Experimental*, III, Promociones Publicaciones Universitarias, Barcelona.
- SOLÉ SABATER, M.J. 1984. "La experimentación en fonética y fonología", *Estudios de Fonética Experimental*, I, Promociones Publicaciones Universitarias, Barcelona.