

*Ariel Practicum*

Eugenio Martínez Celdrán

Análisis  
espectrográfico  
de los sonidos  
del habla

*Editorial Ariel, S.A.*  
Barcelona

Diseño cubierta: Vicente Morales

1.ª edición: junio 1998

© 1998: Eugenio Martínez Celdrán

Derechos exclusivos de edición en español  
reservados para todo el mundo:

© 1998: Editorial Ariel, S. A.  
Córcega, 270 - 08008 Barcelona

ISBN: 84-344-2823-7

Depósito legal: B. 18.723 - 1998

Impreso en España

Ninguna parte de esta publicación, incluido el diseño  
de la cubierta, puede ser reproducida, almacenada o transmitida  
en manera alguna ni por ningún medio, ya sea eléctrico,  
químico, mecánico, óptico, de grabación o de fotocopia,  
sin permiso previo del editor.

## PRÓLOGO

Este libro pretende, como indica su nombre, enseñar los conceptos y criterios que permiten analizar los gráficos más habituales que se manejan en fonética acústica y que proporcionan todos los analizadores de habla. Nosotros hemos utilizado el CSL (Computerized Speech Lab) de Kay Elem. S. A. de New Jersey, pero los gráficos son semejantes a los proporcionados por cualquier otro programa de los que existen en el mercado actualmente.

Se ofrecen ejercicios múltiples en cada capítulo para practicar lo que se estudia en cada lección. Los profesores que estén interesados en la clave de esos ejercicios deben ponerse en contacto con el autor. En definitiva, lo importante es superar los posibles problemas que puedan ofrecer.

Esperamos que el material que presentamos sea suficientemente práctico y que ayude al dominio de la fonética en esta rama tan útil para la investigación y para otros objetivos como son los logopédicos, forenses, tecnológicos, etc.

EUGENIO MARTÍNEZ CELDRÁN  
Universidad de Barcelona  
Facultad de Filología  
Gran Vía, 585  
08007 Barcelona  
*e-mail*: emartine@fil.ub.es

*Barcelona, 6 de febrero de 1998*

### CAPÍTULO 3

## PROPIEDADES ACÚSTICAS DE LAS VOCALES

Como hemos visto, las vocales están caracterizadas por sus dos primeros formantes, siendo sólo las lenguas que distinguen fonéticamente entre vocales redondeadas / no redondeadas y/o postnasalizadas / no postnasalizadas se cuentan de verdad la referencia de tercer formante. A las lenguas, como el español, cuyas vocales sólo se distinguen por su redondeamiento y las postnasalizadas sólo por la presencia de nasalización en los dos primeros formantes. Vamos a presentar los datos de los primeros formantes de las vocales de español. Los valores de los primeros formantes de las vocales de español se muestran en la tabla 3.1. Los datos de los segundos formantes se muestran en la tabla 3.2. Los datos de los terceros formantes se muestran en la tabla 3.3. Los datos de los cuartos formantes se muestran en la tabla 3.4. Los datos de los quintos formantes se muestran en la tabla 3.5. Cada punto de datos que aparece en estas tablas es el resultado de un análisis de los datos de los primeros formantes de las vocales de español. Los datos corresponden a la voz masculina.

Vocal	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)	F5 (Hz)
a	314	477	628	805	949
á	230	426	571	700	811

Los datos de los segundos formantes de las vocales de español se muestran en la tabla 3.2. Los datos de los segundos formantes se muestran en la tabla 3.2. Los datos de los terceros formantes se muestran en la tabla 3.3. Los datos de los cuartos formantes se muestran en la tabla 3.4. Los datos de los quintos formantes se muestran en la tabla 3.5. Cada punto de datos que aparece en estas tablas es el resultado de un análisis de los datos de los segundos formantes de las vocales de español. Los datos corresponden a la voz masculina.

Capítulo 3  
PROPIEDADES ACÚSTICAS  
DE LAS VOCALES

Como hemos visto en el apartado anterior, las vocales están caracterizadas por sus tres primeros formantes, aunque sólo las lenguas que distinguen fonológicamente entre anteriores redondeadas / no redondeadas y/o posteriores redondeadas / no redondeadas necesitan de verdad la referencia del tercer formante. A las lenguas, como el español, cuyas vocales anteriores son no redondeadas y las posteriores son redondeadas sólo les basta con mencionar los dos primeros formantes. Vamos a presentar las vocales a través del español. Los valores son totalmente relativos, en el sentido de que se trata de medias aritméticas obtenidas a partir de los datos de cinco individuos varones (Martínez Celdrán, 1995). En realidad, no existen los valores absolutos. Los datos de cada formante cubren un amplio abanico de valores que se denomina *campo de dispersión*. Cualquier valor que demos es uno de los posibles dentro de ese campo y sólo lo ofrecemos como referencia. Estos datos corresponden a la voz masculina.

DATOS:

Datos en Hz	<i>i</i>	<i>e</i>	<i>a</i>	<i>o</i>	<i>u</i>
F1	313	457	699	495	349
F2	2.200	1.926	1.471	1.070	877

Estas medias pueden llevarse a una carta de formantes (fig. 3.1); esto es, a un gráfico en el que en la ordenada se coloca el F1 de forma descendente y el F2 en la abscisa y ordenado de derecha a izquierda según

se mira. Además los dos ejes de coordenadas se suelen expresar con escalas logarítmicas para que reflejen el modo especial de percepción de la realidad.

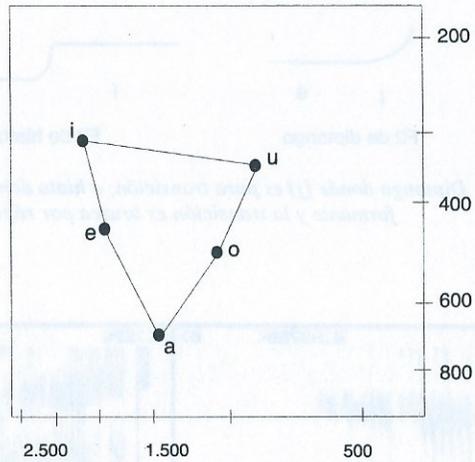


FIG. 3.1. Carta de formantes.

Como se puede apreciar en la figura 3.2, el F2 desciende de frecuencia según la lengua va retrocediendo desde el paladar duro. La vocal [i] es la vocal más adelantada; la [u] es la más atrasada. El F2 hace que la vocal sea más aguda cuanto más alto esté y más grave cuanto más bajo. En el triángulo vocálico, las vocales anteriores son agudas y, las posteriores, graves. La vocal [a] no está definida en cuanto a la acuidad. El F1 tiene relación con la elevación de la masa de la lengua; si ésta se eleva el F1 desciende y viceversa; esto sucede porque si se levanta se amplía la zona faríngea; de lo contrario, se estrecha. Si el F1 sube aproximándose al F2 en el centro del espectro vocálico, la vocal será densa, si el F1 se hace extremo, entonces la vocal será difusa. Así, el rasgo acústico agudo/grave está en relación estrecha con el articulario anterior/posterior y, el difuso/denso, con el alto/bajo. Todo lo dicho se puede comprobar en la figura 3.3, donde se muestra ese descenso progresivo de los segundos formantes desde la vocal más anterior a la más posterior y la uve invertida de los primeros por el descenso de las difusas y el ascenso de las densas. Se ve, además, la diferencia entre voz masculina y femenina; esta última posee, de forma sistemática, valores de formantes más altos que los de la voz masculina.

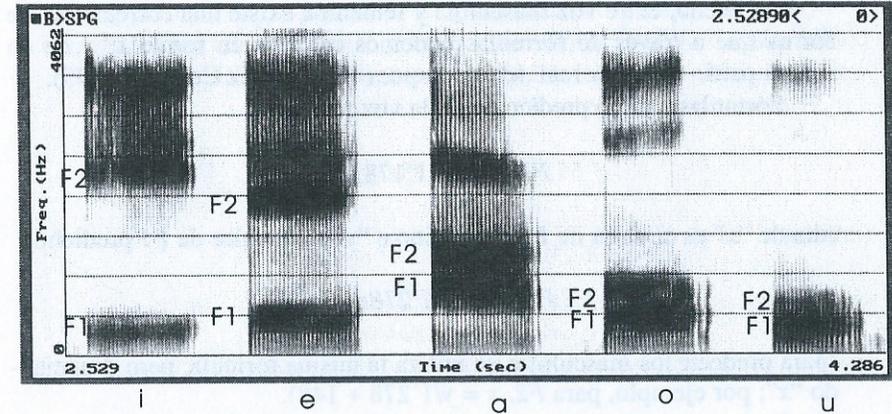
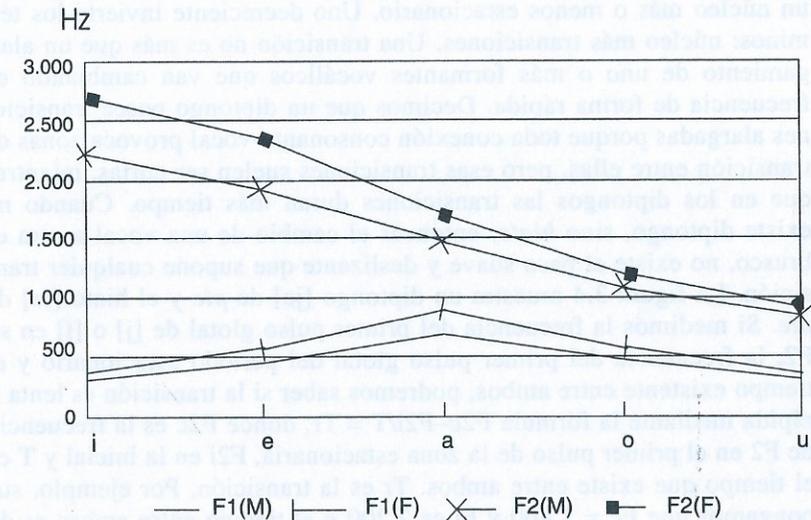


FIG. 3.2. Una muestra de los formantes vocálicos de las cinco vocales españolas.



M = masculino; F = femenino.

FIG. 3.3. Relación de los valores formánticos de la voz masculina y la femenina.

De hecho, entre voz masculina y femenina existe una correlación, de forma que a través de fórmulas podemos calcular un formante  $x$  de un sexo a partir del valor real del sexo opuesto (Martínez Celdrán, 1995).

Fórmulas para la predicción de la voz femenina:

$$F1: y = 1'178x + 16$$

(donde "x" es el valor de  $F1$  masculino e "y" es el valor de  $F1$  predicho)

$$F2: y = 1'278x - 148$$

(para predecir los masculinos se utiliza la misma fórmula, pero despejando "x"; por ejemplo, para  $F2$ ,  $x = y/1'278 + 148$ ).

Por otra parte, si se prefiere simplificar, se podría multiplicar siempre por el coeficiente 1'2, que es bastante aproximado a las cantidades de las fórmulas anteriores. Por ejemplo, un  $F2$  masculino de 2.250 Hz, aplicando la fórmula se transforma en 2.727 Hz; y multiplicando por el coeficiente 1'2 nos da 2.700 Hz.

Un *diptongo* creciente se compone de unas transiciones alargadas y un núcleo más o menos estacionario. Uno decreciente invierte los términos: núcleo más transiciones. Una transición no es más que un alargamiento de uno o más formantes vocálicos que van cambiando de frecuencia de forma rápida. Decimos que un diptongo posee transiciones alargadas porque toda conexión consonante-vocal provoca zonas de transición entre ellas, pero esas transiciones suelen ser cortas, mientras que en los diptongos las transiciones duran más tiempo. Cuando no existe diptongo, sino *hiato*, entonces el cambio de una vocal a otra es brusco, no existe el paso suave y deslizante que supone cualquier transición. La figura 3.4 muestra un diptongo [je] de *pie* y el hiato [íe] de *pie*. Si medimos la frecuencia del primer pulso glotal de [j] o [i] en su  $F2$ , la frecuencia del primer pulso glotal del periodo estacionario y el tiempo existente entre ambos, podremos saber si la transición es lenta o rápida mediante la fórmula  $F2c - F2i / T = Tr$ , donde  $F2c$  es la frecuencia de  $F2$  en el primer pulso de la zona estacionaria,  $F2i$  en la inicial y  $T$  es el tiempo que existe entre ambos.  $Tr$  es la transición. Por ejemplo, supongamos que  $Fc = 1.800$  y  $Fi$  es 2.200 y el tiempo entre ambas es de 30 ms, entonces  $1.800 - 2.200 / 30 = -13.33$  Hz/ms; es decir, se trata de una transición que va descendiendo (porque es negativa) 13'13 Hz cada milésima de segundo. Si obtuviéramos otros datos de frecuencia para un hiato cuya transición durase 10 ms entonces  $1.800 - 2.100 / 10 = -30$ ; es decir, el descenso sería más rápido pues en cada milésima de segun-

do se descenderían 30 Hz. Está claro que tendríamos un diptongo en el primer caso y un hiato en el segundo. Estos datos podrían muy bien ser los de *pie/pie*.

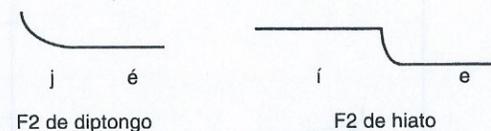


FIG. 3.4. Diptongo donde [j] es pura transición; e hiato donde [í] posee su propio formante y la transición es brusca por rápida.

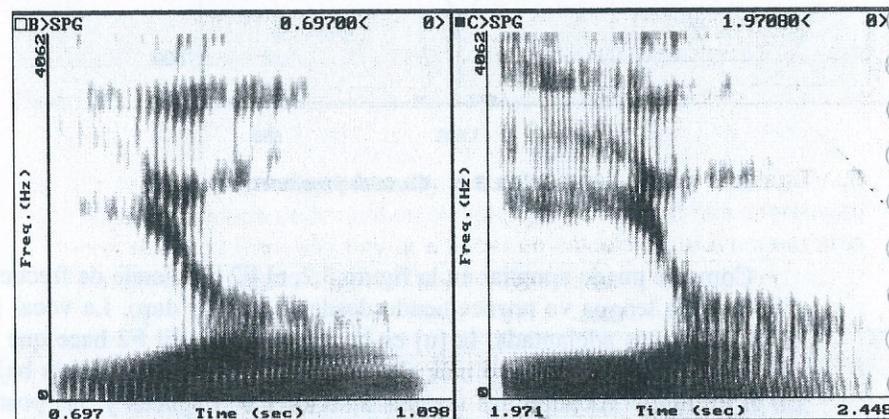
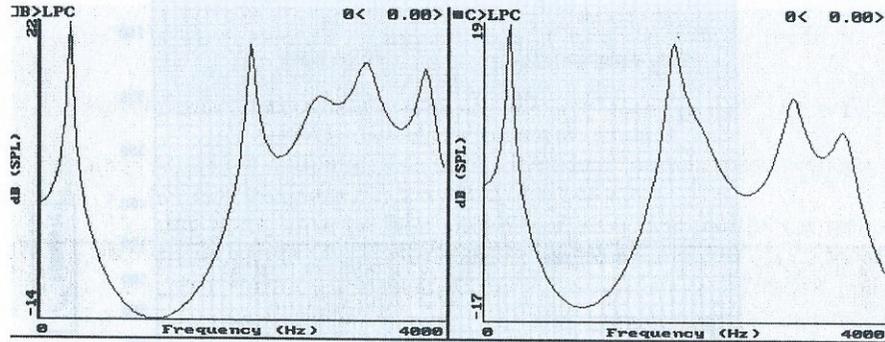


FIG. 3.5. Espectrograma real del diptongo y del hiato: [jó] - [to].

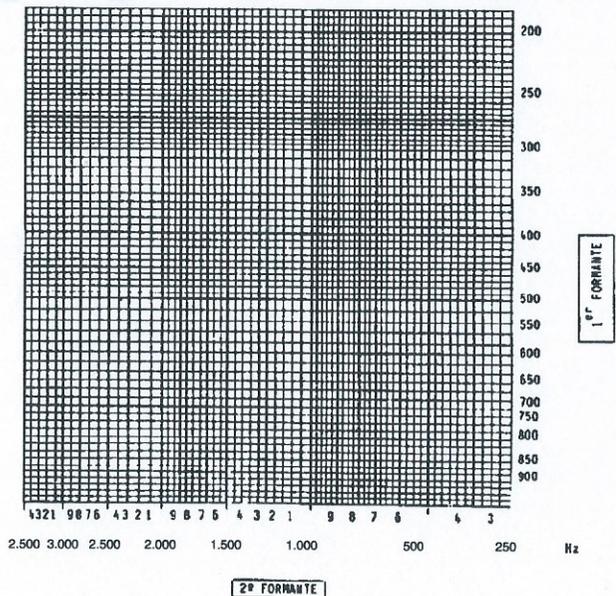
Ejercicios

1. Analizar los tres primeros formantes (frecuencia central, ancho de banda y amplitud) de las dos vocales siguientes [i-y] y comentar las diferencias existentes entre ellas. Proporcionar una explicación de cavidad articularia para esas diferencias.

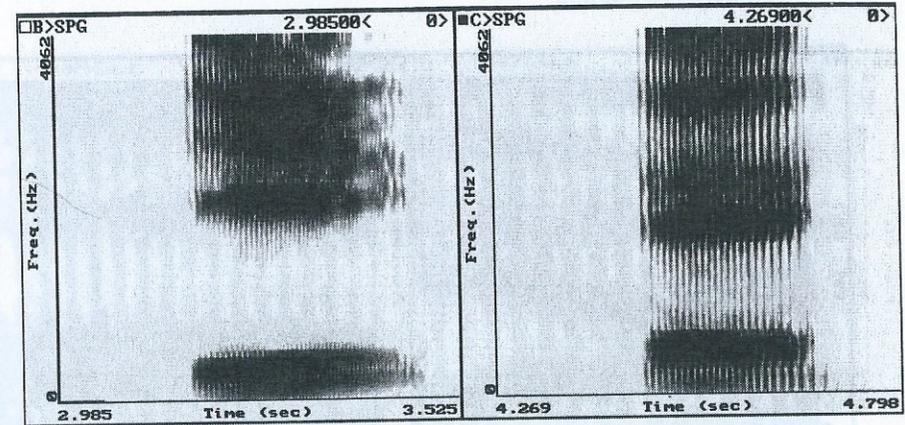


2. Trasladar los siguientes datos a una carta de formantes:

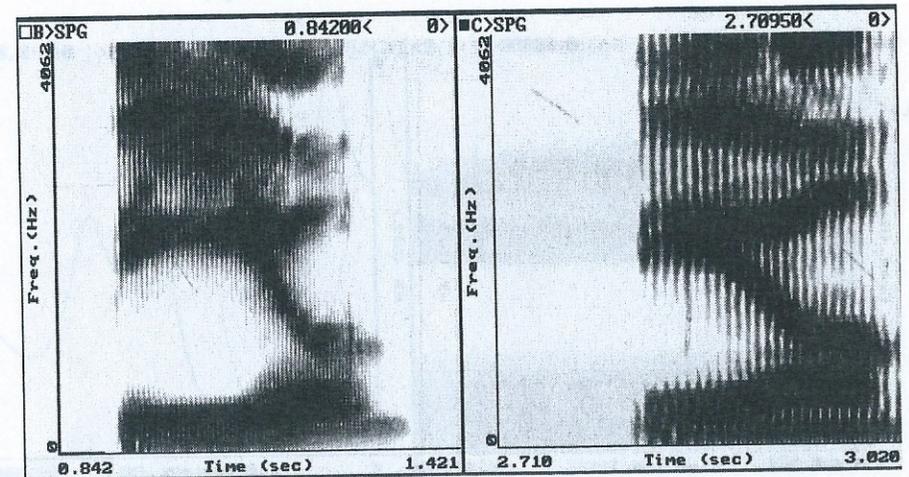
	i	e	ɛ	a	ɔ	o	u
F1	300	450	550	690	520	470	270
F2	2.300	1.960	1.710	1.380	1.128	980	800



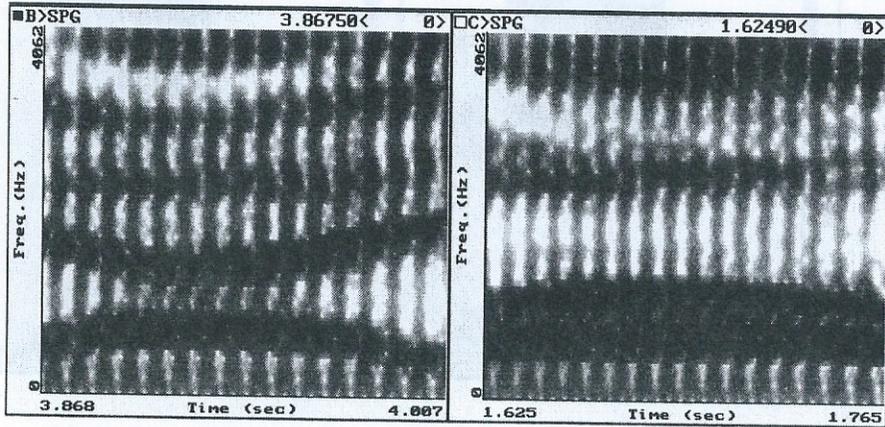
3. Indicar si en los espectrogramas siguientes hay dos muestras de una misma vocal o dos vocales diferentes.



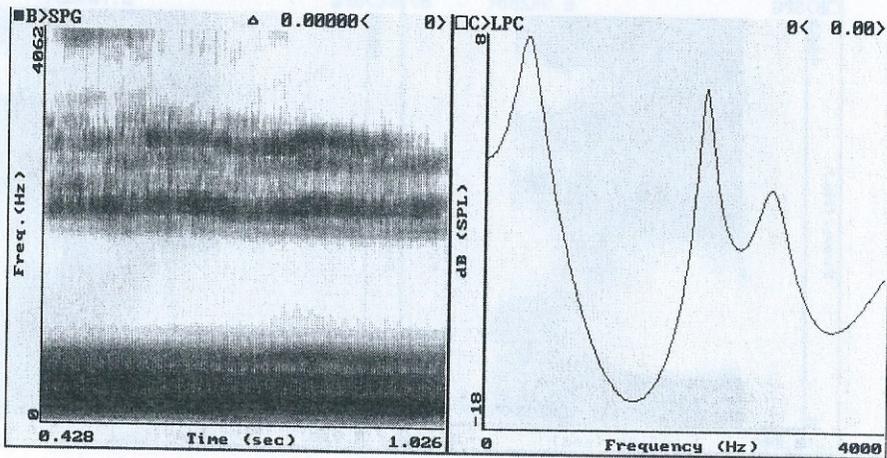
4. En los dos espectrogramas siguientes se presenta un diptongo y un hiato, ¿se puede decir cuál es el diptongo y comentar sus diferencias con el hiato? Para ello hay que hacer un cálculo para averiguar si la transición es rápida o lenta en cada caso.



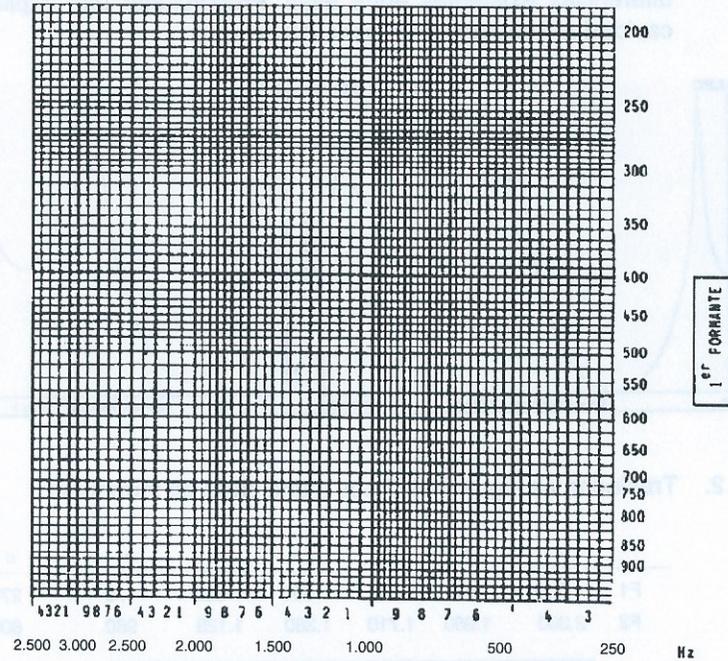
5. Ahora te ofrecemos dos espectrogramas con sendas vocales y hay que dilucidar si se trata de la misma vocal o no, tras medir su tres primeros formantes.



6. Indicar si el espectrograma y espectro siguientes son de voz masculina o femenina por los valores vocálicos: vocal [e].



7. Obtener los valores de F1 y F2 de las vocales de la figura 3.2 y colocarlos en una carta de formantes.



## CAPÍTULO 4

### PROPIEDADES ACÚSTICAS DE LAS CONSONANTES OCLUSIVAS

#### 4.1. Sonoridad, tensión y aspiración

La sonoridad hace referencia a la vibración de las cuerdas vocales. El sonido cuya fuente principal sea la vibración laringal será sonoro, mientras que si no existe ésta será sordo. La vibración laringal es muy rica en armónicos. El tono fundamental depende del número de veces que las cuerdas se abren y se cierran en un segundo; en el hombre, entre 100 y 200 veces en la mujer, entre 150 y 200. En los espectrogramas de banda ancha se manifiesta por un formante de baja frecuencia (unos 200 Hz) que denota la vibración laringal y que se suele denominar para de voz-tensión (fig. 4.1).

La tensión se manifiesta en los sonidos tensores por una mayor presión aérea detrás del punto de articulación y una rigidez de los segmentos por lo cual conlleva una mayor duración del sonido. Los tensores consonánticos tienen una reducida presión, rigidez y duración.

Esta clase de muchos fonemas consonánticos que la duración es uno de los índices de la tensión de los sonidos, no tanto porque la duración en sí misma produce tensión sino porque toda tensión conlleva tensión. Como una mayor duración, la tensión exige un mayor esfuerzo muscular y para realizarlo se necesita más tiempo. Existen sonidos largos que no son tensores, de hecho, todos los sonidos que se forman en el grupo laringal se suelen alargar y, en cambio, precisamente por estar en la fase final del grupo se truncan mucho. Además, cuando se encuentran dos sonidos iguales en la cadena hablada, por ejemplo entre final de una palabra y principio de otra, por ej. los sabores rientes o con naturalidad, pueden alargarlos sin que la tensión aumente. Incluso, si esa tensión se encuentra en el interior de una palabra, por ej. *lanche*, no

#### 4.1. Sonoridad, tensión y aspiración

La sonoridad hace referencia a la vibración de las cuerdas vocales. El sonido cuya fuente principal sea la vibración laríngea será *sonoro*, mientras que si no existe ésta será *sordo*. La vibración laríngea es muy rica en armónicos. El tono fundamental depende del número de veces que las cuerdas se abran y se cierren en un segundo; en el hombre, entre 100 y 200 veces; en la mujer, entre 150 y 300. En los espectrogramas de banda ancha se manifiesta por un formante de baja frecuencia (unos 500 Hz), que denota la vibración laríngea y que se suele denominar barra de sonoridad (fig. 4.1).

La tensión se manifiesta en los sonidos *tensos* por una mayor presión aérea detrás del punto de articulación y una rigidez de los músculos, lo cual conlleva una mayor duración del sonido. Los *laxos* correspondientes tienen una reducida presión, rigidez y duración.

Está claro que muchos fonetistas consideran que la duración es uno de los índices de la tensión de los sonidos, no tanto porque la duración en sí misma produzca tensión sino porque toda tensión conlleva forzosamente una mayor duración: la tensión exige un mayor esfuerzo muscular y para realizarlo se necesita más tiempo. Existen sonidos largos que no son tensos; de hecho, todos los sonidos que acaban un grupo fónico se suelen alargar y, en cambio, precisamente por estar en la fase final del grupo se relajan mucho. Además, cuando se encuentran dos sonidos iguales en la cadena hablada, por ejemplo entre final de una palabra y principio de otra, por ej. *los sabores salados* o *con naturalidad*, pueden alargarse sin que la tensión aumente. Incluso, si esa secuencia se encuentra en el interior de una palabra, por ej. *innoble*, no

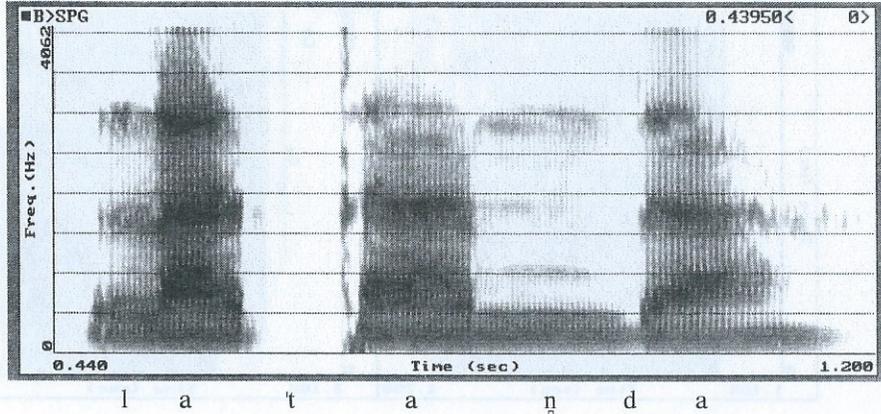


FIG. 4.1. En el espectrograma de la palabra "la tnda", se observa la barra de sonoridad en la [d] y su ausencia en la [t].

tiene por qué necesitar mayor tensión que cuando sólo se da una consonante sola: y *noble*. En este caso se trata de sonidos largos o breves, no de tensos o laxos.

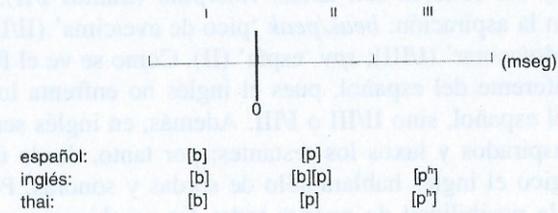
En el caso de los sonidos tensos, frente a los laxos correspondientes, se produce una mayor duración porque en su articulación los músculos están enérgicamente enervados. Un gesto tenso exige por sí mismo un mayor tiempo de permanencia de los órganos en su posición que un gesto relajado. Por ejemplo, Fisher-Jørgensen (1968:108) caracterizaba las oclusivas tensas de la siguiente forma: «acoustically, they are characterized by a longer closure period and a shortening of the preceding vowel and, to some extent, of the following vowel, perhaps also by a stronger intensity of the explosion».

Este fenómeno se observa también en otros sonidos como los fricativos (fig. 4.2).

La aspiración se manifiesta sobre todo en los sonidos *aspirados* de varias lenguas como un soplo laríngeo que sucede a la explosión por un gesto de abertura de la glotis, ausente en los sonidos *no aspirados*; desde un punto de vista acústico, ese soplo constituye un ruido turbulento (fig. 4.3).

Por las explicaciones que hemos dado podría parecer que sonoridad, tensión y aspiración son rasgos que no tienen nada en común; pero recientemente se ha defendido que los tres pueden explicarse conjuntamente en las oclusivas en relación con el momento en que

comienza la vibración laríngea respecto de la relajación de la oclusiva, fenómeno conocido con la sigla VOT (VOICE ONSET TIME) (Lisker y Abramson, 1964) (en algunos libros se ha traducido por «tiempo de emisión de la voz» TEV). Tomadas tres lenguas como el español, el inglés y el thai, cada una de ellas segmenta el *continuum* del VOT de diferente manera.



La barra vertical representa el momento de la explosión. Por tanto, en el segmento I el comienzo de la vibración laríngea ocurre antes de la explosión. En el segmento II el comienzo de la vibración laríngea coincide o está muy próximo a la explosión. Por último, en el segmento III el comienzo es bastante posterior a la explosión. (En la representación del segmento, la línea horizontal representa la dispersión de los valores temporales de comienzo de la vibración laríngea o VOT.)

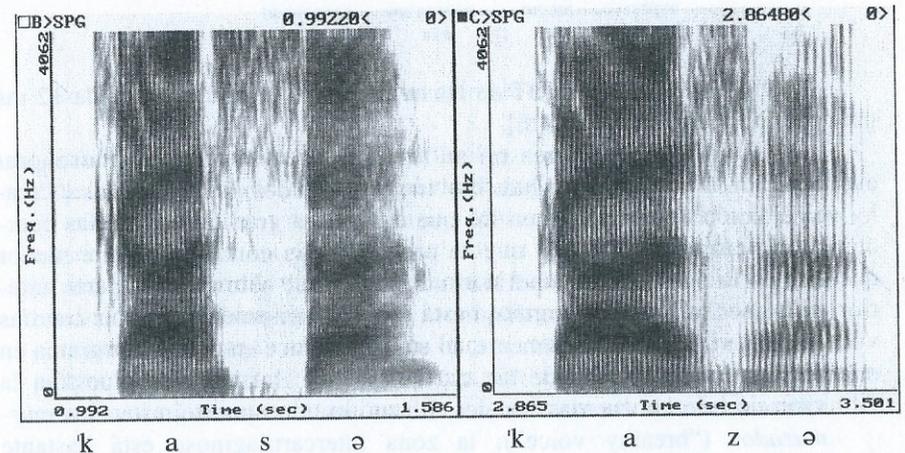
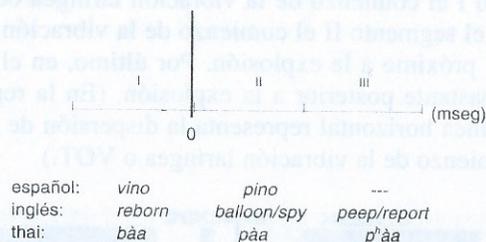


FIG. 4.2. La oposición [kasə]/[kazə] del catalán muestra claramente las diferencias de duración no sólo de las [s/z] sino también de las vocales precedentes respectivas, mayor en la sonora-laxa que en la sorda-tensa, como compensación de la menor duración de la fricativa sonora frente a la mayor de la sorda.

El *continuum* queda segmentado en tres partes. El español sólo utiliza las dos primeras, pues carece de oclusivas aspiradas. El inglés utiliza los tres, pero fonológicamente sólo distingue dos, pues en el momento intermedio realiza las sonoras iniciales y las sordas tras [s] inicial. El thai, por último, utiliza las tres segmentaciones fonológicamente. Es decir, el español sólo tiene sonidos sordos y sonoros y, en todo caso, los sordos son tensos y los sonoros son laxos: *vino/pino* (tramos I/II). El inglés utiliza también la aspiración: *beak/peak* 'pico de ave/cima' (II/III); *reborn/report* 'renacido/contar' (I/III); *spy* 'espía' (II). Como se ve el funcionamiento es muy diferente del español, pues el inglés no enfrenta los segmentos I/II, como el español, sino II/III o I/III. Además, en inglés serán tensos los sonidos aspirados y laxos los restantes; por tanto, desde un punto de vista fonológico el inglés hablará sólo de sordas y sonoras. Por último, el thai ofrece la posibilidad de oponer todas las combinaciones entre sí, por lo que en thai habrá que hablar de dos rasgos fonológicos: sonoro y aspirado; ej. *bàa/pàa/p<sup>h</sup>àa* 'hombro/bosque/partir'.

Ejemplos:



Obsérvese que el VOT de las oclusivas de la figura 4.3 es de 12 ms para [t] y de 93 ms para [th].

La actividad laríngea no se limita a la comentada: sorda/sonora/aspirada. En muchas lenguas hay también sonidos *laringalizados* ("creaky voice"): son aquellos en los que existe una gran tensión en las cuerdas vocales lo cual hace que se produzca una corriente de aire menor que en la posición de sonoridad normal ya que sólo vibra la parte anterior, pues la zona intercartilaginosa está fuertemente unida y las cuerdas vocales vibran más lentamente, lo cual se traduce en el espectrograma en una mayor separación de las estrías (pulsos glotales) que muestran la vibración de las cuerdas vocales. En un sentido opuesto, están los *murmurados* ("breathy voice"): la zona intercartilaginosa está bastante abierta y la corriente de aire es mayor que en los sonoros normales, lo cual hace que en el espectrograma se vea ruido mezclado con las estrías de la vibración (fig. 4.4).

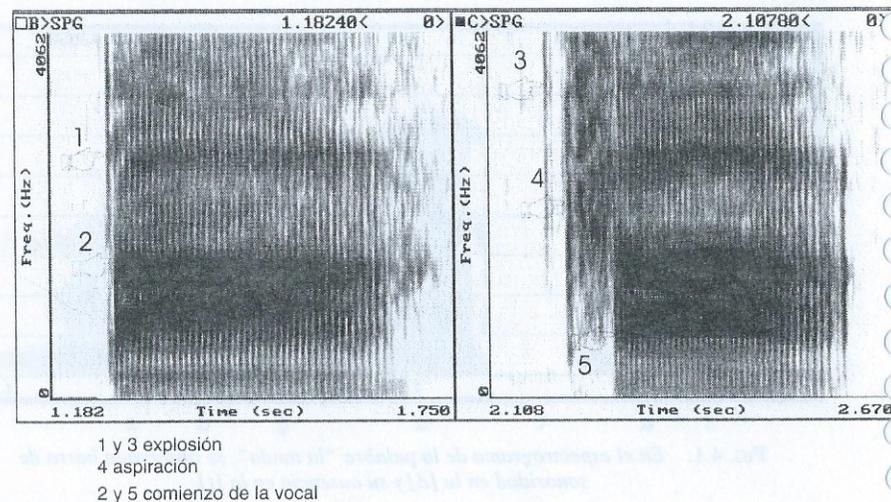


FIG. 4.3. Diferencias espectrográficas entre la oclusiva sorda no aspirada y la sorda aspirada [ta-tha].

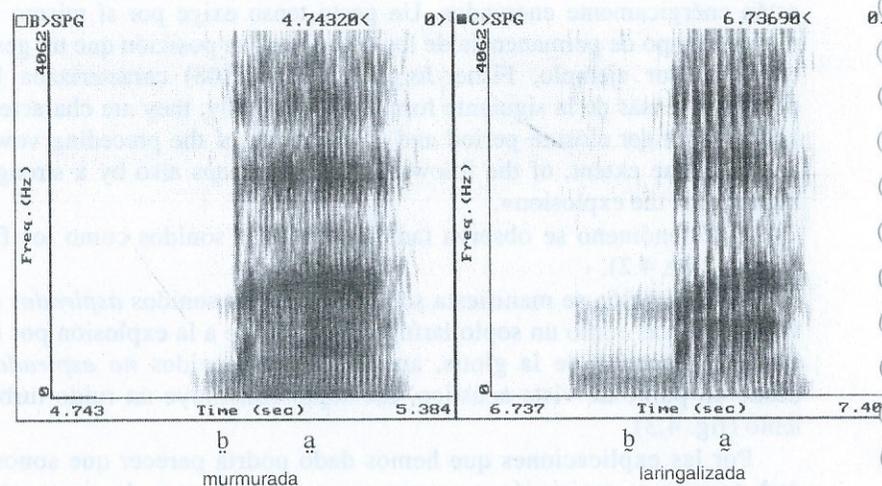


FIG. 4.4. Sonidos murmurados y laringalizados.

#### 4.2. Consonantes no pulmonares

Hay una distinción mayor entre consonantes pulmonares, la mayoría, que se caracterizan por ser realizadas con el aire que sale de los pulmones y no pulmonares, que aprovechan el aire que se queda atrapado en la boca entre dos cierres, caso de las eyectivas, o hay una entrada de aire exterior, como sucede con inyectivas y clics. Como todas estas consonantes poseen dos cierres, se caracterizan por tener dos explosiones: la primera corresponde a la abertura de la zona más anterior: labial, dentalalveolar, etc.; la segunda, a la abertura de la zona glotal, en las eyectivas o inyectivas, o de la zona velar en los clics. Entre la primera explosión y la segunda hay un tiempo de silencio que en algunas de estas consonantes es muy largo, por encima de las 150 ms. Además, la primera explosión es también muy larga: alrededor de las 20 ms.

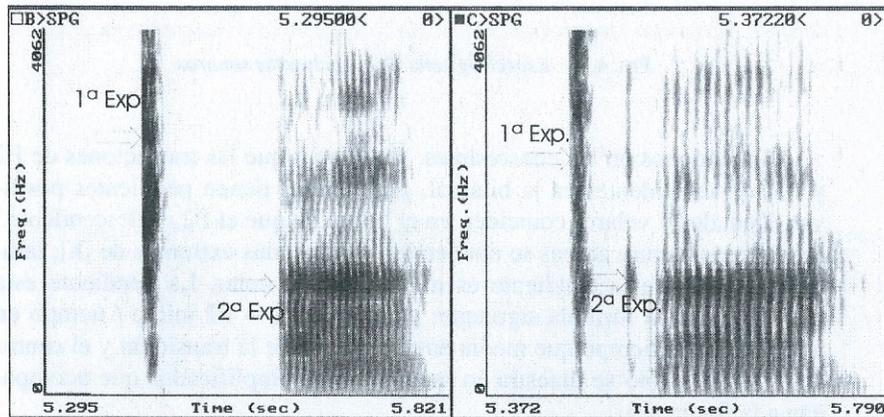


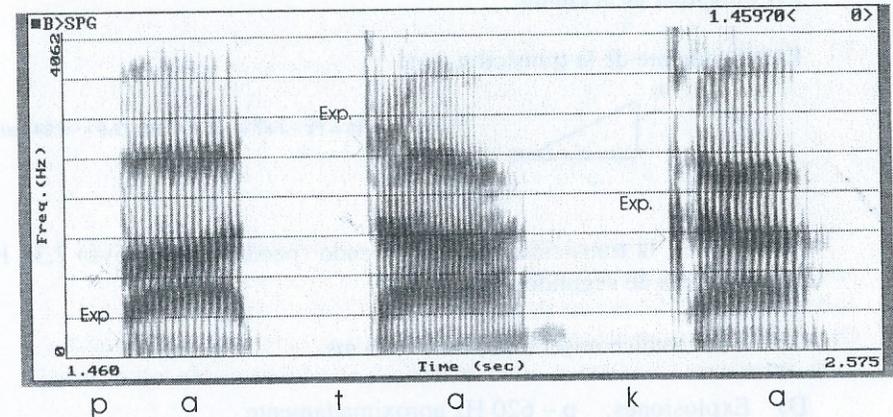
FIG. 4.5. Consonante eyectiva alveolar y clic dental.

#### 4.3. Las oclusivas del español

Se caracterizan acústicamente por una zona de silencio, unas 90 ms para las sordas y unas 60 ms para las sonoras, y una barra de explosión en el momento de la relajación, esto por lo que se refiere al modo de articulación. Como las oclusivas españolas no son aspiradas, veremos

que el VOT no alcanza valores altos en las sordas, aunque el contraste entre sonoras y sordas está suficientemente representado por las diferencias de VOT. Por último, el punto de articulación viene dado tanto por la frecuencia donde se sitúa la máxima intensidad de la explosión, como por las transiciones vocálicas. Resulta que los formantes de las vocales no son rectos; es decir, no se mantienen en la misma frecuencia, sino que sufren variaciones en los extremos debido a la influencia del punto de articulación de las consonantes que la circundan. Estas variaciones se denominan transiciones.

Como siempre, los valores y los espectrogramas expuestos para el español no son absolutos, son sólo valores de referencia.



Estas flechas indican la dirección de las transiciones de los formantes segundos.

Esta otra indica la máxima intensidad de la explosión a una frecuencia determinada.

Este símbolo indica el comienzo de la transición y el centro de la vocal según el F2.

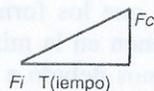
FIG. 4.6. Espectrograma de las oclusivas sordas.

A partir de un espectrograma como éste podemos efectuar diversos cálculos como los que siguen:

Cálculos previos: 1.115 ms -115 mm; por tanto 1 mm = 9.7 ms (1.115/115)

500 Hz – 5.5 ms; por tanto 1ms = 91 Hz (500/5.5)

A) Pendiente de la transición de p:

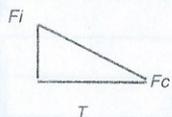


$$\text{Pendiente}(p) = F_c - F_i/T = 1.245 - 1.079/76 = 2.18 \text{ Hz/ms}$$

Donde  $F_c$  = Frecuencia del centro del formante;  $F_i$  = Frecuencia inicial o del primer pulso glotal; y  $T$  = tiempo existente entre esos dos puntos.

Es decir, la transición va ascendiendo (pendiente positiva) 2.18 Hz por milésima de segundo.

B) Pendiente de la transición de t:



$$\text{Pendiente}(t) = F_c - F_i/T = 1.328 - 1.500/73.6 = -2,34 \text{ Hz/ms}$$

Esto es, la transición va descendiendo (pendiente negativa) 2,34 Hz por milésima de segundo.

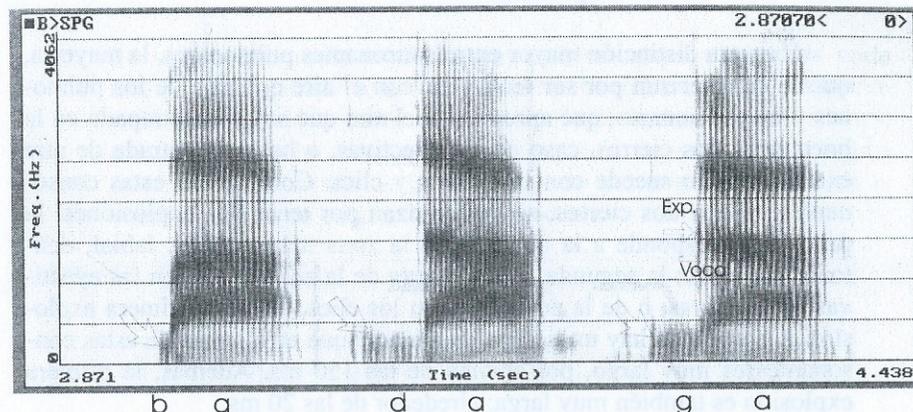
C) VOT (voice onset time) de  $k = 36$  ms.

D) Explosiones: p – 620 Hz aproximadamente  
t – por encima de los 2.500 Hz  
k – 1.500 Hz aproximadamente

Podemos ofrecer unos valores generales de VOT para el español peninsular, siguiendo los datos de M.<sup>a</sup> L. Castañeda (1986) (datos en milésimas de segundo):<sup>1</sup>

[p] = 6.5	[b] = -69.8
[t] = 10.4	[d] = -77.7
[k] = 25.7	[g] = -58

1. Para medir el VOT se considera que la barra de explosión posee 0 ms; cuando los pulsos glotales comienzan antes, entonces el VOT será negativo, de lo contrario será positivo.



Barras de sonoridad.

FIG. 4.7. Espectrograma de las oclusivas sonoras.

Situándonos en las consonantes, se observa que las transiciones de F2 y F3 son ascendentes en la bilabial, pues ambos tienen pendientes positivas. Dentales y velares coinciden en el hecho de que el F2 es descendente.<sup>2</sup> La diferencia entre ambas se nota en los valores más extremos de [k]; también el valor de la pendiente es mayor para la velar. La pendiente está calculada con la fórmula siguiente:  $p = F2 \text{ centro} - F2 \text{ inicio} / \text{tiempo en ms}$ ; se toma el tiempo que media entre el inicio de la transición y el centro de la vocal, como se muestra en los cálculos ejemplificados que acompañan a la figura 4.6.

En la figura 4.8 hemos resumido los índices más importantes que caracterizan a las oclusivas en general y a cada una en particular: transiciones, barras de explosión y VOT. Hemos añadido un concepto discutido: el locus. Desde un punto de vista acústico no existe ningún elemento que podamos identificar como locus. Es sólo un punto teórico a donde se dirigen las transiciones vocálicas. En la síntesis se determinaron tres loci: 700 Hz para las bilabiales; 1.800 Hz para las alveolares y 3.000 Hz para las velares, aunque posteriormente se vio que las velares poseían dos: 3.000 para las vocales anteriores y 800 para las posteriores.

2. Aunque se ha de tener en cuenta el valor del F2 y F3 de la vocal, ya que la transición variará en función de la altura de estos formantes (véase fig. 4.8).

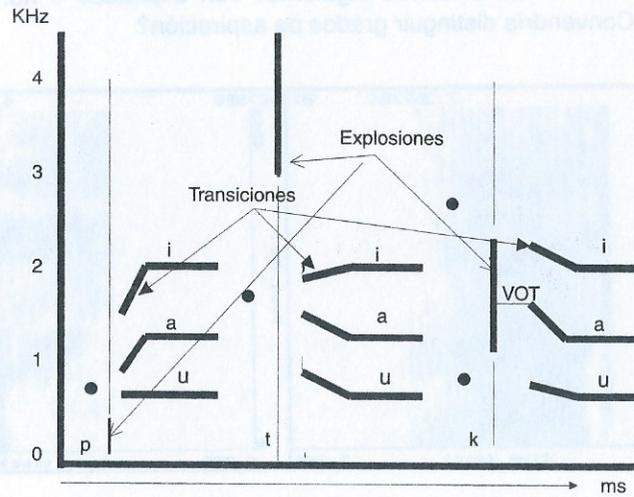
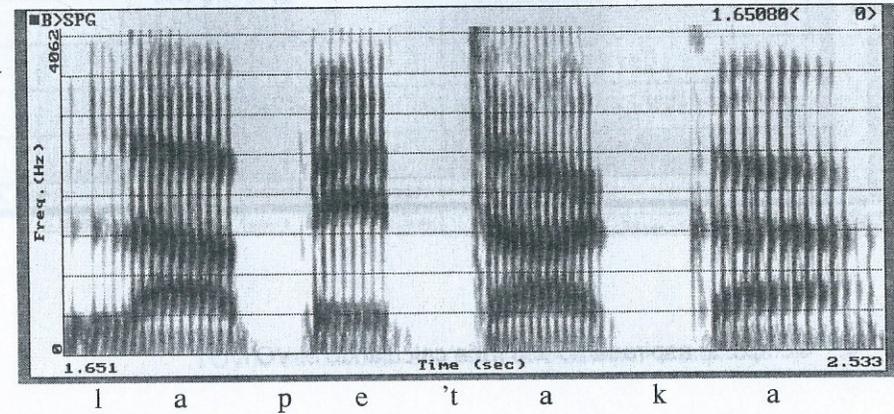


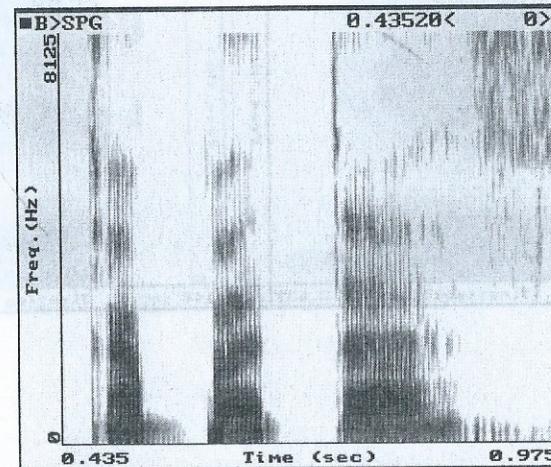
FIG. 4.8. Dibujo esquemático de las transiciones de F2 de [i, a, u] y de las barras de explosión (se señala el VOT al tener en cuenta la distancia de la explosión al formante vocálico y también mostramos con mayor grosor en qué frecuencia se da la mayor intensidad), según las oclusivas sordas. El punto negro indica la dirección que toman las transiciones. Se conoce con el nombre de "locus".

### Ejercicios

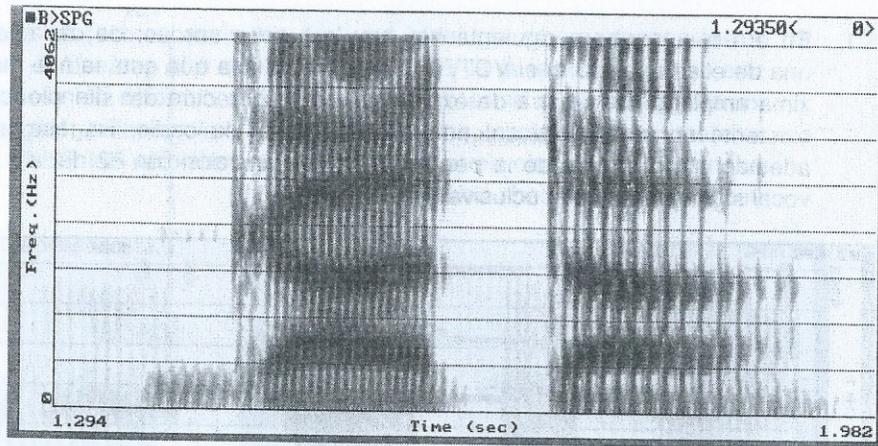
- En el espectrograma siguiente hay tres oclusivas sordas; de cada una de ellas analizar a) el VOT, b) la frecuencia a la que está la máxima amplitud de la barra de explosión y c) la duración del silencio existente entre la vocal anterior y la barra de explosión. Hágase además d) un cálculo de la pendiente de la transición del F2 de la vocal siguiente de cada oclusiva.



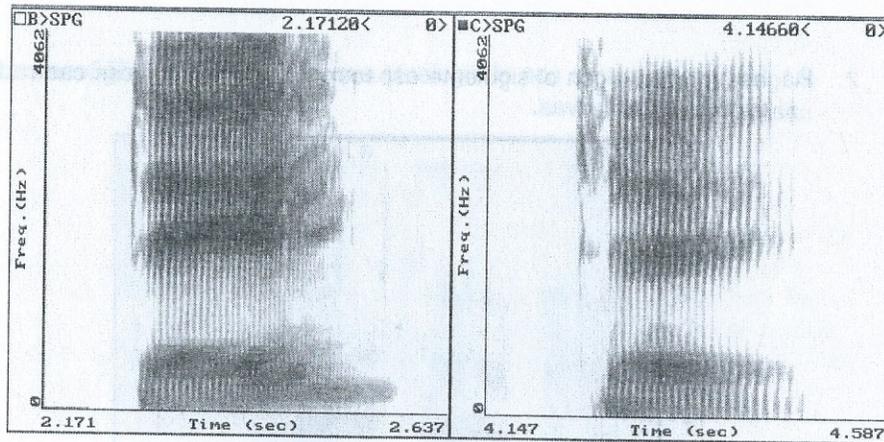
- Hágase lo mismo con el siguiente espectrograma y averígüese cada una de las oclusivas.



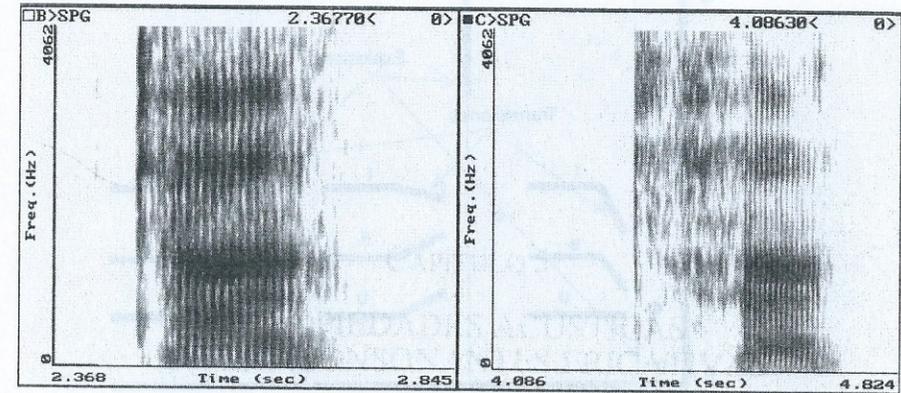
3. Comparar sorda/sonora calculando el VOT en la palabra "gata".



4. Comparar aspirada/no aspirada calculando el VOT.



5. Indicar si las oclusivas siguientes son aspiradas o no. ¿Por qué?  
¿Convenría distinguir grados de aspiración?



## CAPÍTULO 5

### PROPIEDADES ACÚSTICAS DE LAS CONSONANTES FRICATIVAS Y APROXIMANTES

### 5.1. Fricativas

Las consonantes fricativas se caracterizan por tener ruido turbulento; es decir, inarmónicos. Obsérvese en los espectrogramas la armonicidad de las vocales vecinas, que se manifiesta a través de las estrías de los pulsos glotales perfectamente regulares, frente a la amalgama aleatoria que es la inarmonicidad del ruido. Esto es lo que las caracteriza frente a los demás sonidos. Entre sí, además de las transiciones son evidentes dos diferencias: la frecuencia donde se da la máxima intensidad (mayor negror) y esa misma intensidad, como se ve en el siguiente cuadro.

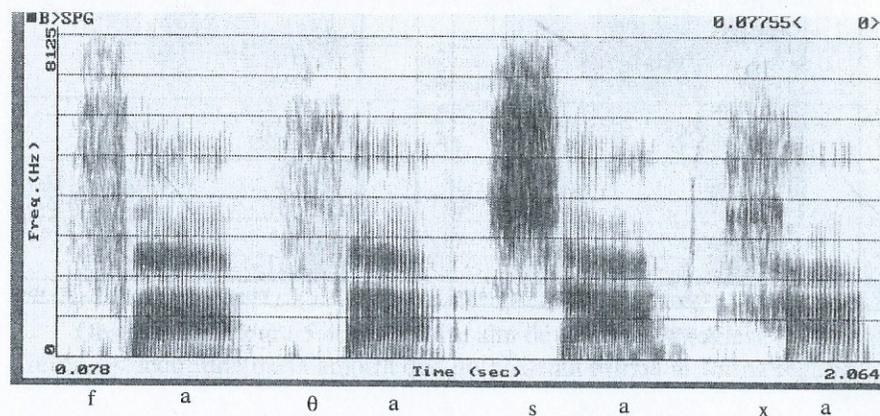


FIG. 5.1. Espectrograma de las fricativas.

	Frecuencia comienzo Hz	Frecuencia máxima intensidad Hz	Intensidad de pico dB	Intensidad global dB
[f]	2.000	5.000	4	7
[θ]	3.000	5.500	13	9
[s]	3.000	3.680	26	24
[x]	1.000	3.700	20	14

(La frecuencia de comienzo es sólo aproximada; obsérvese en los espectrogramas. Intensidad de pico quiere decir la intensidad que tiene el pico más alto del espectro. En el espectrograma, las intensidades sólo se pueden apreciar por el mayor o menor negror.)

Una diferencia que queremos destacar es la existente entre la [s] apical típica de las zonas castellanas y la predorsal que predomina en andaluz, canario, latinoamericano, etc. Obsérvese que la apical posee gran intensidad de ruido a partir de los 3.000 Hz, mientras que la predorsal posee una intensidad menor y comienza a partir de los 4.000 Hz.

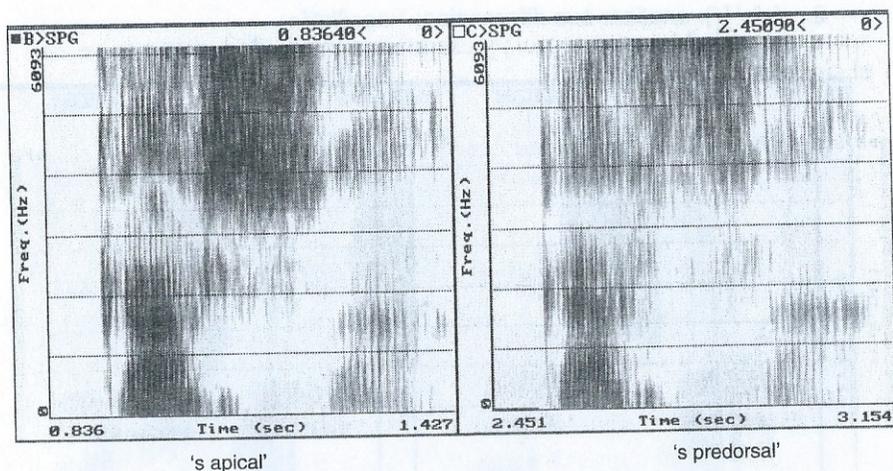


FIG. 5.2. Espectrogramas de las dos variedades principales de "s" pronunciadas en la palabra "casa".

## 5.2. Aproximantes

Carecen de ruido. Tienen las estrías típicas de los sonidos armónicos. Se caracterizan por un descenso considerable de la intensidad respecto de las vocales vecinas, apreciable por el menor negror de sus zonas formánticas. Los formantes son meras transiciones entre los formantes vocálicos; su frecuencia determina su distinto punto de articulación, sobre todo a través de las transiciones del F2 y de la frecuencia de ese mismo F2.

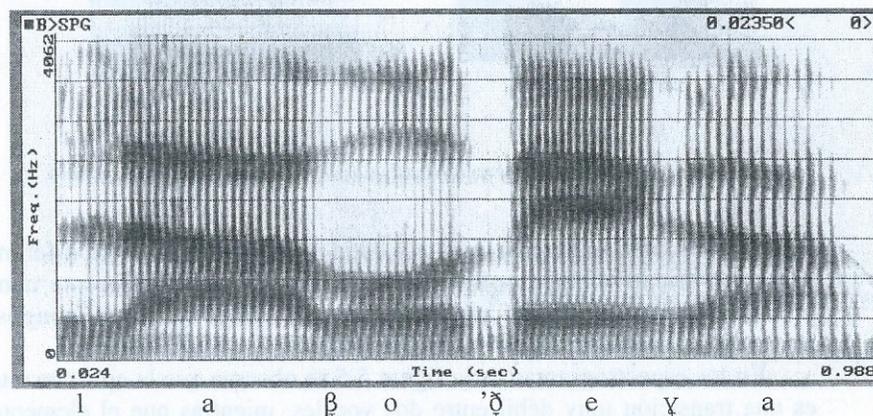


FIG. 5.3. Espectrograma de las aproximantes.

	Intensidad vocal precedente	Intensidad consonante aproximante	Diferencia de intensidad	Frecuencia F2 aproximante
[β]	32	14	18	900
[ø]:	27	10	17	1.280
[j]:	25	16	9	2.080
[γ]:	34	13	21	1.640

Obsérvese la figura 5.4; en la parte alta de la fricativa inglesa [θ] hay ruido, es decir, una masa amorfa que no presenta estrías ni forma regular alguna. Compárese con el sonido español [θ] y se verá que este último no posee nada de ruido y todo en él son estrías: ésta es la diferencia esencial entre fricativas y aproximantes (Martínez Celdrán, 1991).

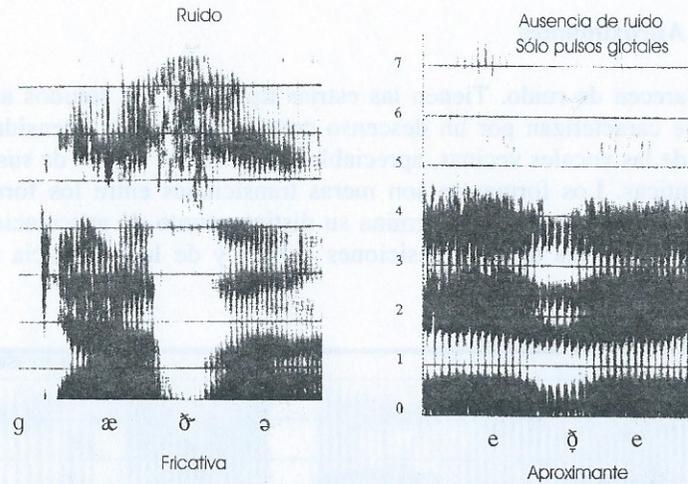


Fig. 5.4. Comparación entre la fricativa inglesa y la aproximante española.

Hay también una aproximante palatal que no sólo se ha de diferenciar de la fricativa correspondiente, sino que además es diferente también de los elementos pre y posnucleares palatales de los diptongos; por ej. [j] de *mayo* y la [j] de *piojo*.

En los espectrogramas de la figura 5.5 se observa que la aproximante es una transición muy débil entre dos vocales, mientras que el elemento prenuclear del diptongo no es nada débil, es en realidad un elemento vocálico muy breve y una gran transición.

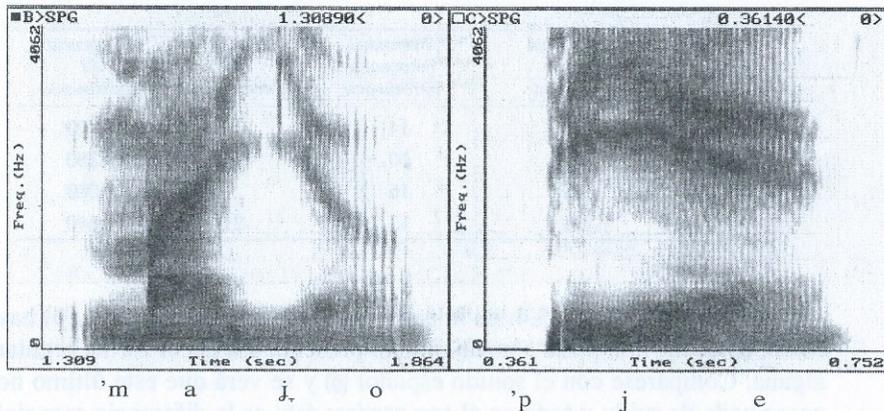
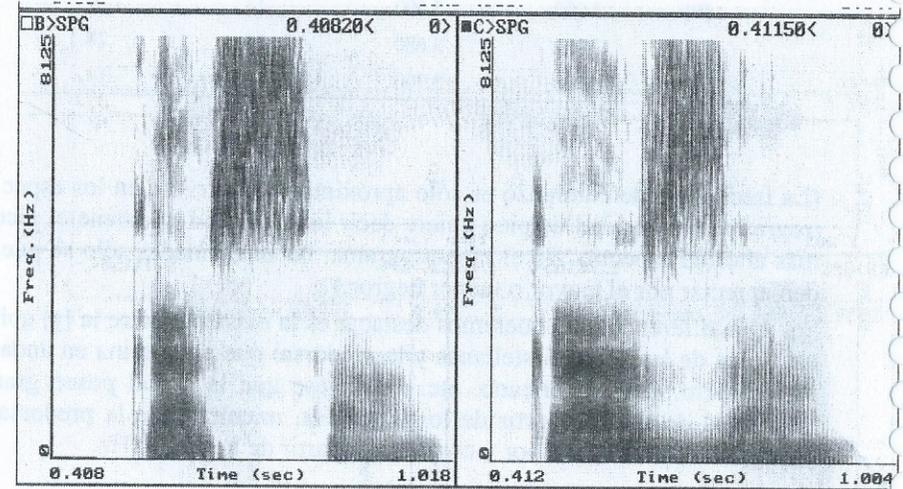


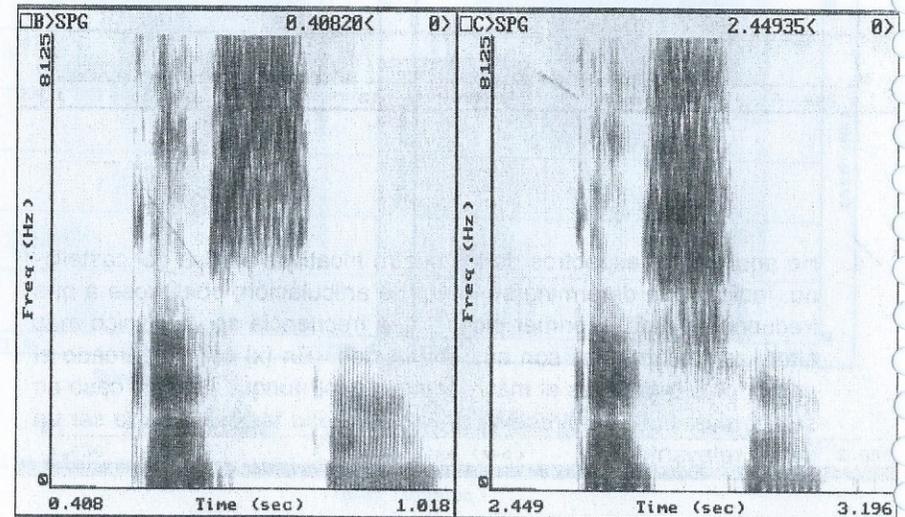
FIG. 5.5. Aproximante vs semiconsonante de diptongo.

Ejercicios

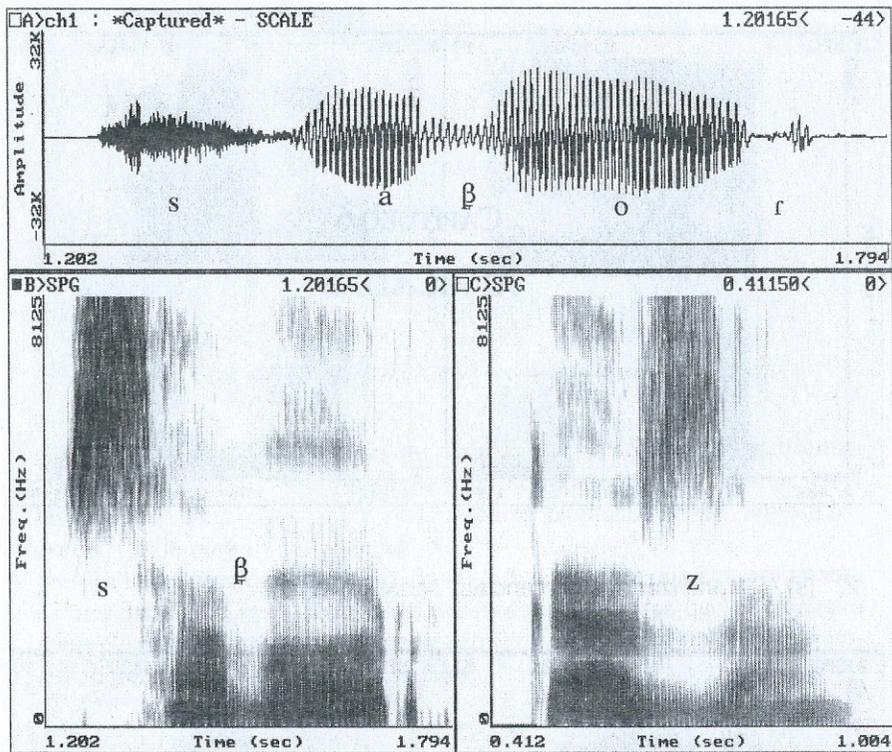
1. Analizar las diferencias existentes entre estas dos fricativas: sorda/sonora: 'kasə/'kazə.



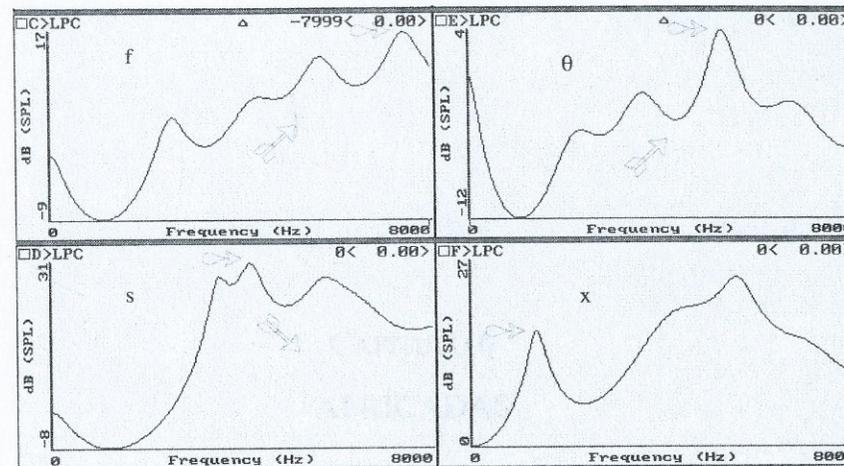
2. [s] / [ʃ]: analizar sus diferencias: 'kasə/'kaʃə.



3. En el oscilograma y en el espectrograma siguientes hay una aproximante y una fricativa; analizar sus diferencias. Comparar también la aproximante con la fricativa sonora.

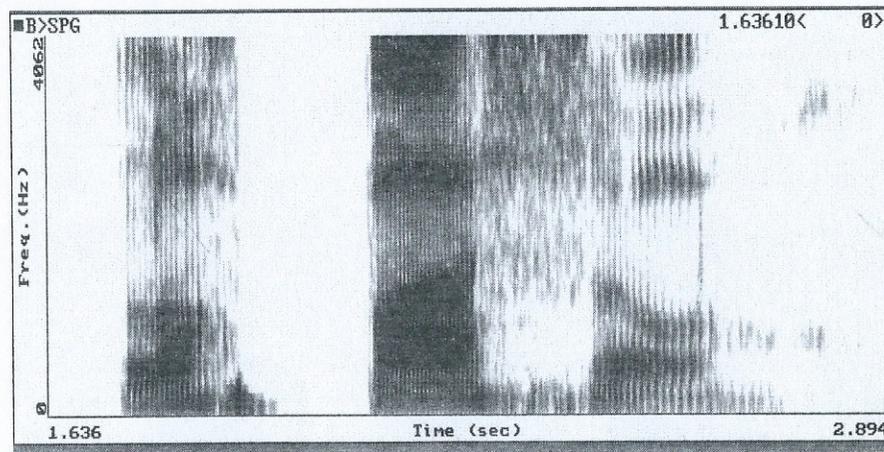


4. He aquí cuatro espectros de las cuatro fricativas sordas del castellano. Indicar qué determina su punto de articulación; obsérvese a qué frecuencia se da el primer pico, a qué frecuencia se da el pico más alto y también cuáles son sus amplitudes. (En [x] se ha marcado el primer pico porque es el más característico, aunque en este caso no sea el más alto.) La dirección de la intensidad también puede ser un índice relevante.



↘ Dirección de los picos de intensidad  
 ▲ Picos de mayor amplitud

5. Indicar si el plural del artículo de "los pasos" se ha realizado con [z] o [ʝ]. ¿Por qué? ¿La última [o] es abierta o cerrada?



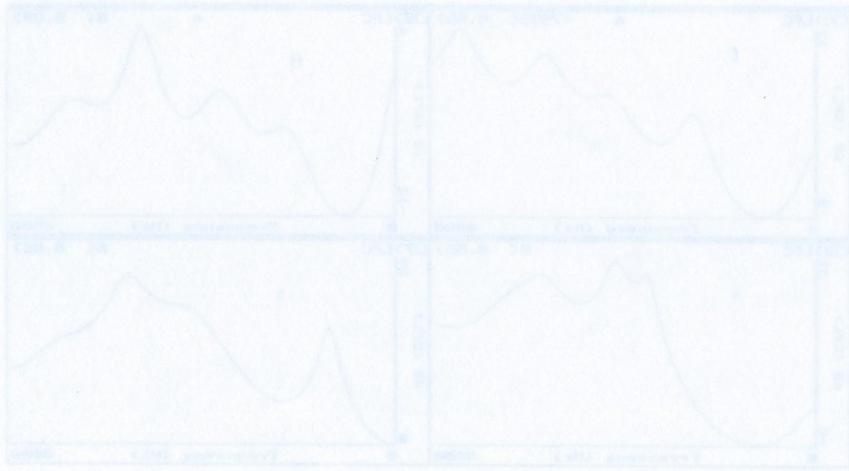


Figura 2.3. Espectrogramas de los sílabas de 'caja'.

¿Indicar el punto del sonido de "los pasos", se ha realizado con [s] o [ʃ]? ¿Por qué? ¿La última [s] es alveolar o dental?

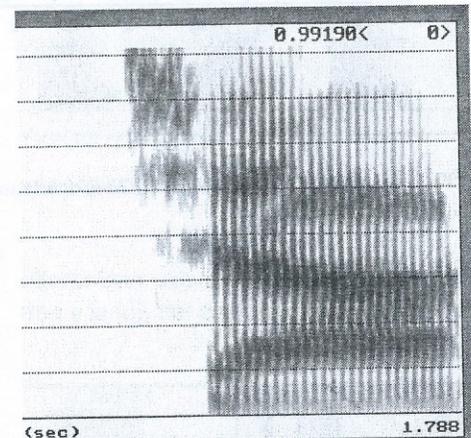


En el espectrograma y en el espectrograma siguientes hay una sibilante y una fricativa, analizar sus diferencias. Comentar también la aproximación con la fricativa sorda.



## CAPÍTULO 6 AFRICADAS

Los cuatro espectros de los cuatro fricativas sordas del castellano no indican que difieren en punto de articulación; observamos a qué frecuencia se da el primer pico, a qué frecuencia se da el pico más alto y también cuáles son sus amplitudes. En [x] se ha marcado el primer pico porque es el más característico, aunque en este caso no sea el más alto. La dirección de la intensidad también puede ser un índice relevante.



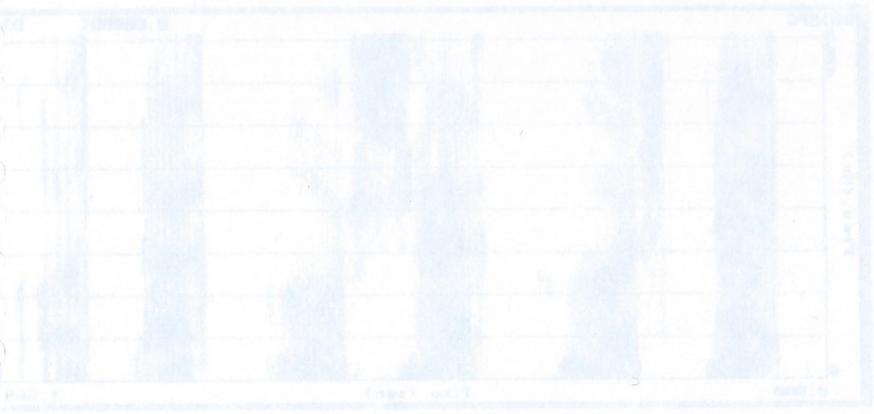
tj a:

FIG. 6.1. Espectrograma de la africada sorda.

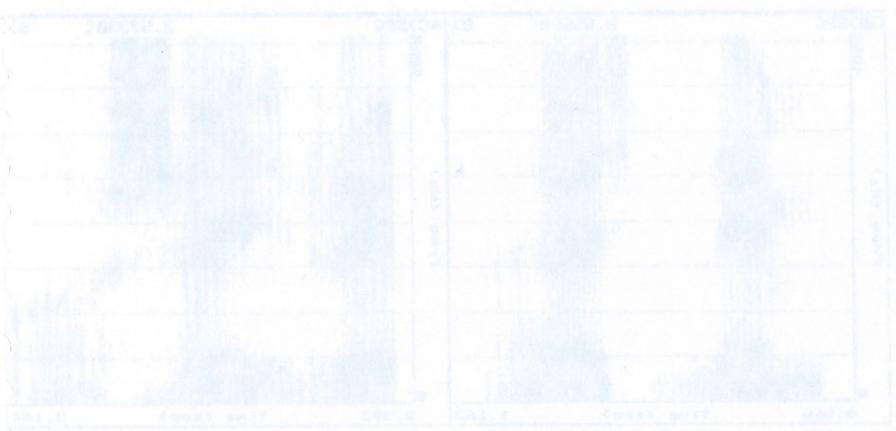
En la africada sorda existe un ruido tan largo o más que la zona de silencio correspondiente a la oclusión. En la sonora, además de las estrías de baja frecuencia debidas a la vibración de las cuerdas vocales, el silencio y el ruido (o las estrías de la aproximante) se reducen considerablemente.

Fricativas

1. Medir la duración de la oclusión y de la fricción de la sibilada por separado. Compararlas con las de la fricativa y la oclusiva de las otras palabras. Palabras: gachas, gachas-cato.



2. Describir las diferencias existentes entre la sibilada sonora [ʃ] del castellano en gallego y la [ʃ] del castellano en el resto.



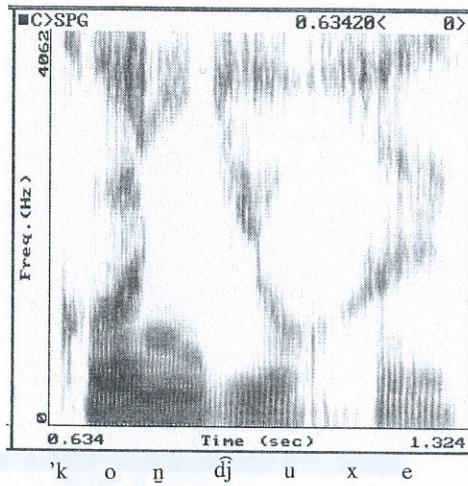
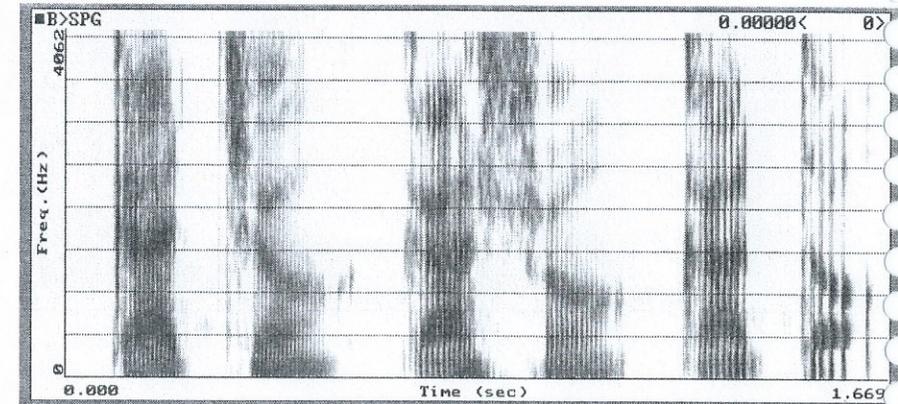


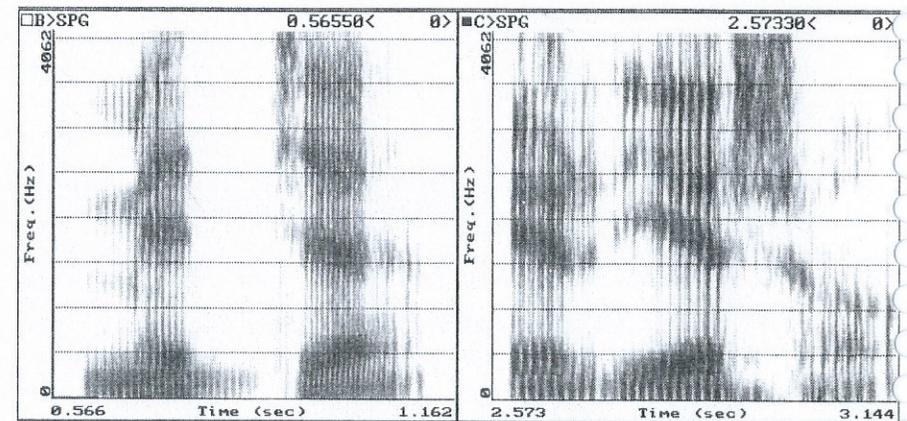
FIG. 6.2. Espectrograma de la africada sonora.

### Ejercicios

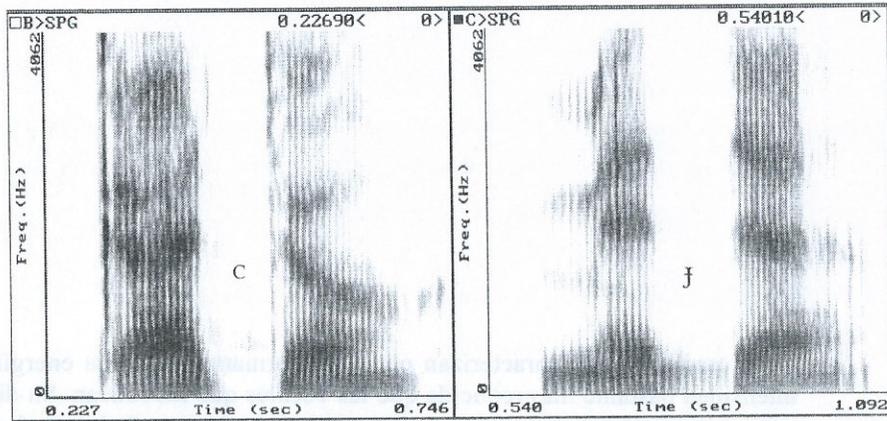
1. Medir la duración de la oclusión y de la fricación de la africada por separado. Compararlas con las de la fricativa y la oclusiva de las otras palabras. Palabras grabadas: *cacho-caso-cato*.



2. Describir las diferencias existentes entre la africada sonora [d̪] del catalán en *metge* y la [d̪] del castellano en *el yeso*.



3. Comparar las realizaciones africadas anteriores con las oclusivas palatales.



CAPÍTULO 7  
NASALES

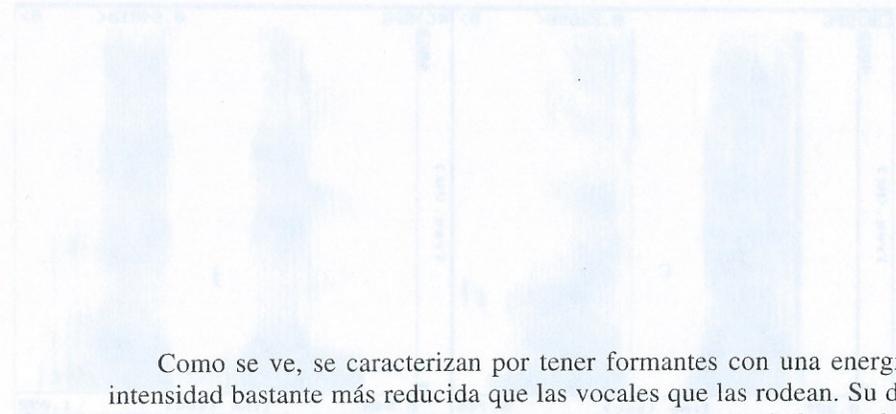
Formante	Frecuencia (Hz)	Amplitud (dB)
F1	100	20
F2	1000	18
F3	1500	18

Los datos siguientes corresponden a los cálculos realizados sobre este espectrograma:

(A) Frecuencias de las transiciones (Hz/ms):  
 -m: 1.285 - 1.000/44.45 = 0.41  
 -n: 1.356 - 1.292/67.2 = 1.12  
 -ŋ: 1.607 - 1.000/67.25 = 4.1

(B) F1 = 370 para las tres nasales

(C) F2 (m) = 1000 Hz  
 F2 (n) = 1300 Hz  
 F2 (ŋ) = 1500 Hz



Como se ve, se caracterizan por tener formantes con una energía o intensidad bastante más reducida que las vocales que las rodean. Su diferencia principal con las aproximantes es el paso abrupto de la vocal a la consonante y viceversa en las nasales; eso hace que se vean como bloques. Las aproximantes son transiciones suaves entre las vocales. Obsérvese cómo las transiciones nasales determinan su punto de articulación, además del segundo formante.

	<i>Intensidad vocal precedente</i>	<i>Intensidad conso- nante nasal</i>	<i>Diferencia de intensidad</i>
[m]:	27	21	6
[n]	30	18	12
[ɲ]:	23	16	7

He aquí algunos cálculos realizados sobre este espectrograma:

- A) Pendientes de las transiciones (Hz/ms):  
 -m:  $1.285 - 1.000/44.45 = 6.41$   
 -n:  $1.356 - 1.295/63.5 = 1.12$   
 -ɲ:  $1.607 - 2.000/95.25 = -4.13$
- B) F1 = 250 para las tres nasales
- C) F2 (m) = 1.000 Hz  
 F2 (n) = 1.400 Hz  
 F2 (ɲ) = 1.900 Hz

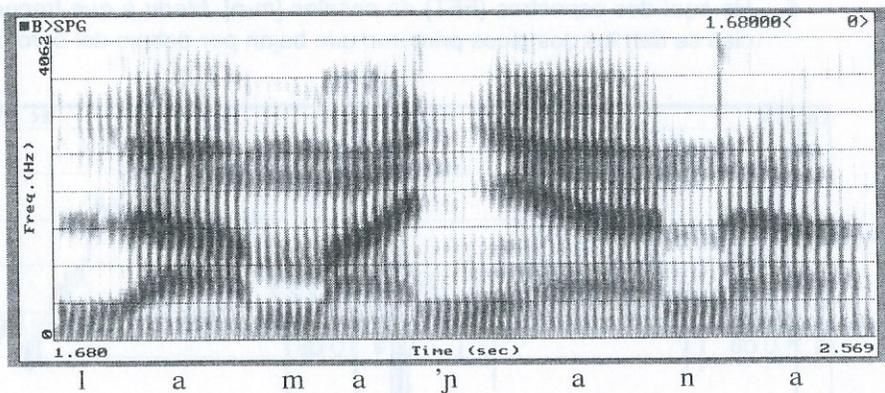


FIG. 7.1. Espectrograma de las nasales.

Las nasales se caracterizan también por poseer ceros en el espectro o antirresonancias. Se trata de picos negativos en el espectro nasal como los que mostramos en la figura 7.2:

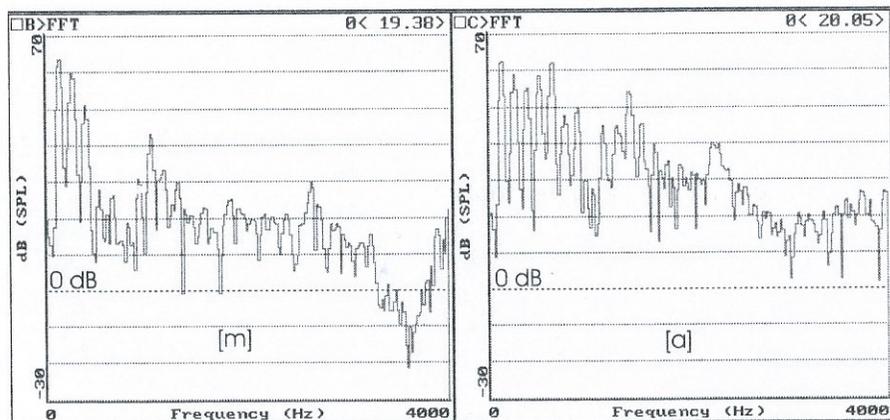
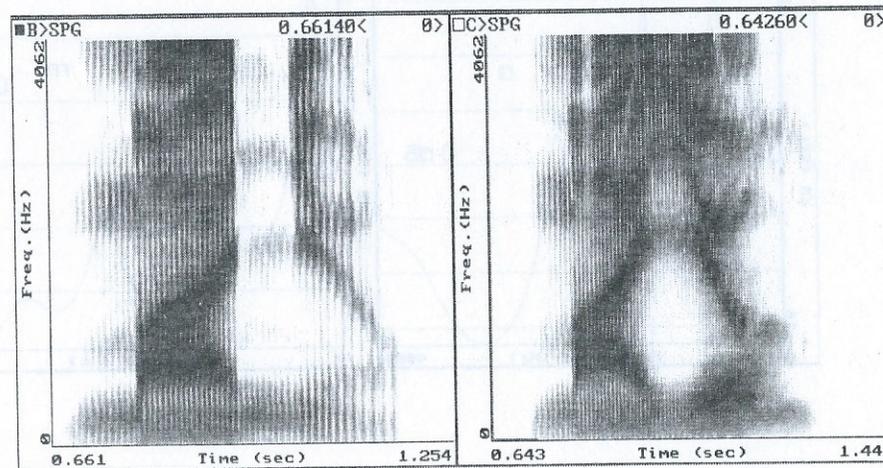


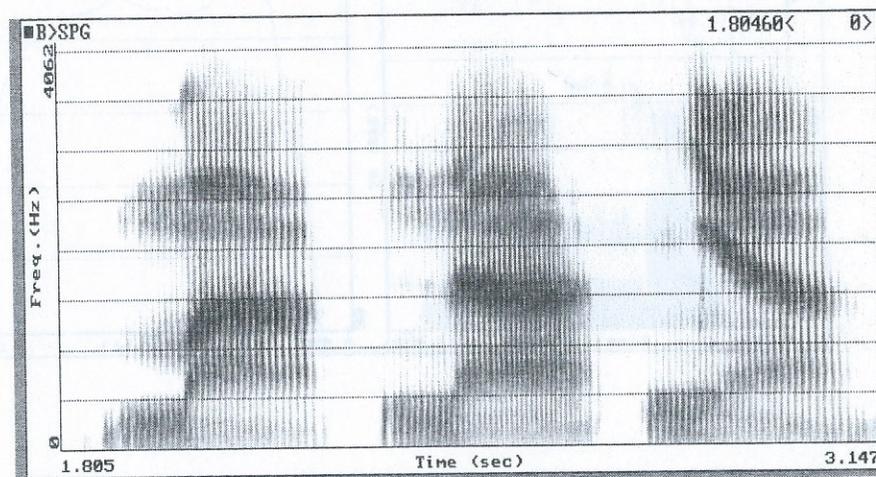
FIG. 7.2. Espectros de una nasal que muestra picos negativos, que son los ceros o antirresonancias. Obsérvese cómo en el espectro vocálico no hay ningún pico por debajo de la línea de puntos que es la línea cero dB.

### Ejercicios

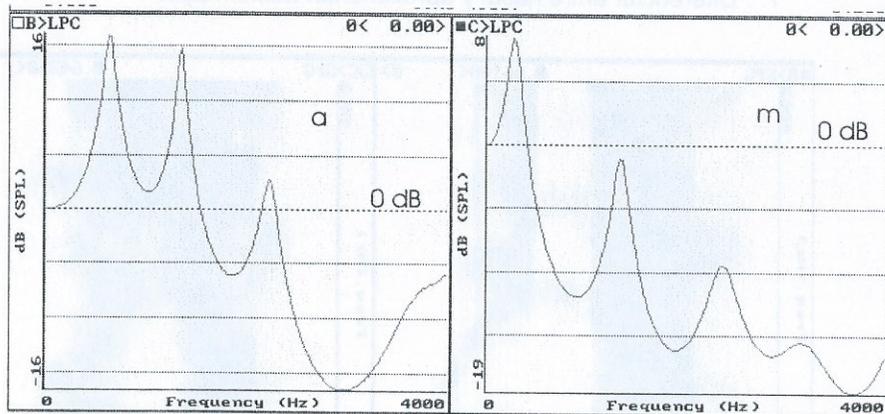
1. Diferenciar entre nasal y aproximante: *maño/mayo*.



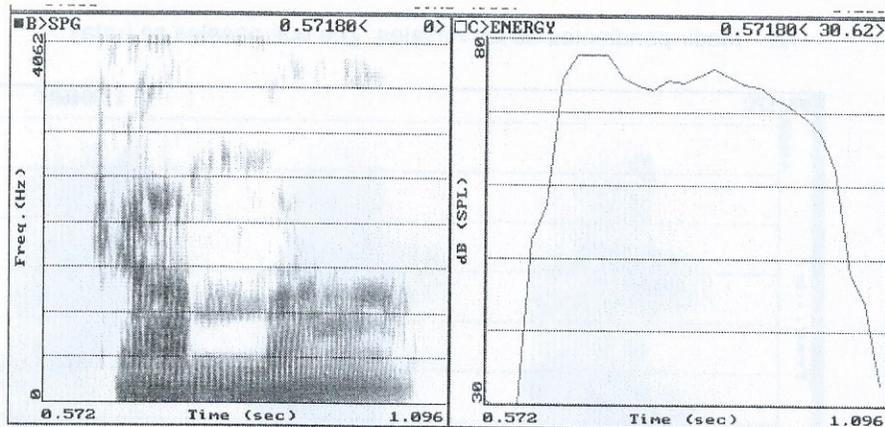
2. Medir pendientes de las nasales; ¿de qué nasales se trata?



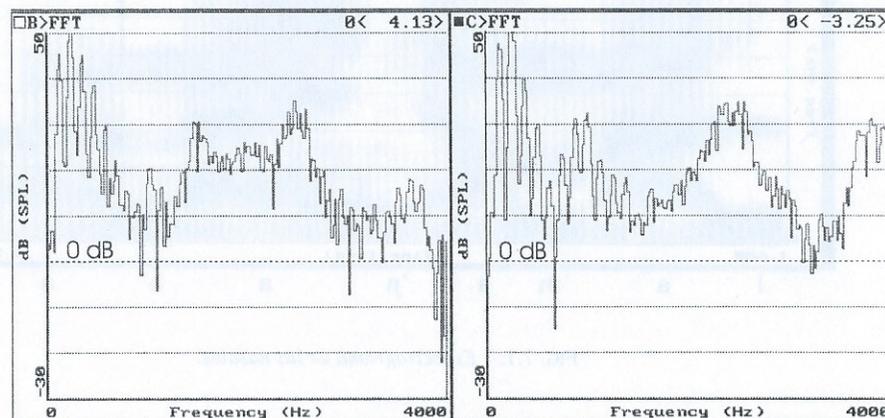
3. Dos espectros de la vocal precedente y de la nasal. Medir amplitudes de F1, F2 y F3. Comparar las diferencias.



4. Curva de intensidad global de la palabra "cama" para comparar las amplitudes de las vocales y de la nasal.



5. He aquí dos espectros (FFT) de nasales [m-n]. Medir a qué frecuencias se dan los dos picos primeros que bajan por debajo del cero dB.



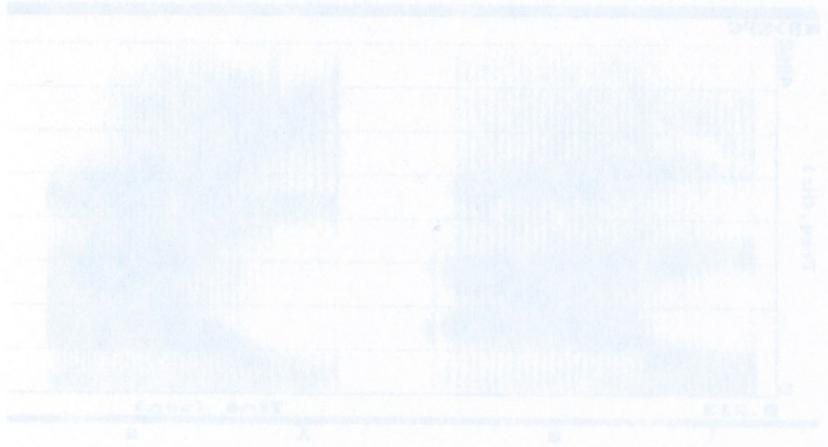
# CAPÍTULO 8 LÍQUIDAS

## 8.1. Lateralización

Se parecen a las nasales, pero poseen formantes como ellas y el paso  
 entre consonante y vocal es también abrupto. Se diferencian principal-  
 mente por la intensidad de los formantes altos, bastante mayor en las  
 laterales que en las nasales por regla general.

Formantes de las laterales

Formante	Nasales	Laterales
F1	~ 300 Hz	~ 300 Hz
F2	~ 1000 Hz	~ 1000 Hz
F3	~ 2500 Hz	~ 2500 Hz
F4	~ 4000 Hz	~ 4000 Hz



Formantes de las laterales

LIQUIDAS  
CAPITULO 8

### 8.1. Laterales

Se parecen a las nasales, pues poseen formantes como ellas y el paso entre consonante y vocal es también abrupto. Se diferencian principalmente por la intensidad de los formantes altos, bastante mayor en las laterales que en las nasales, por regla general.

	Promedio de intensidad (dB)		
	F2	F3	F4
vocales	24	17	23
nasales	9	6	7
laterales	20	18	16

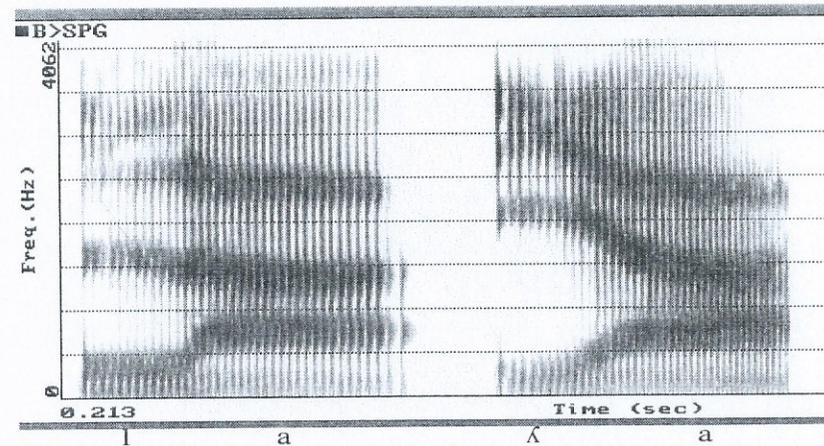


FIG. 8.1. Espectrograma de las laterales.

8.2. Vibrantes

Se caracterizan por las breves interrupciones de la energía que forman pequeñas oclusivas que llevan hasta barra de explosión. Característicamente, la vibrante simple sólo posee una interrupción, mientras que la múltiple tiene dos o más. En los espectrogramas que presentamos las oclusiones duran 25 ms aproximadamente. Existe una variante ocasional de la vibrante que es aproximante; por tanto, en vez de oclusión o silencio en el espectrograma, se observan las estrías características de las aproximantes.

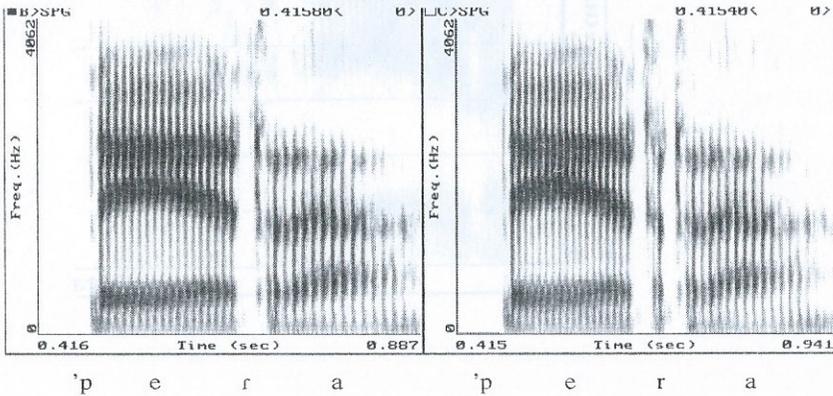


FIG. 8.2. Espectrograma de las vibrantes: simple y múltiple.

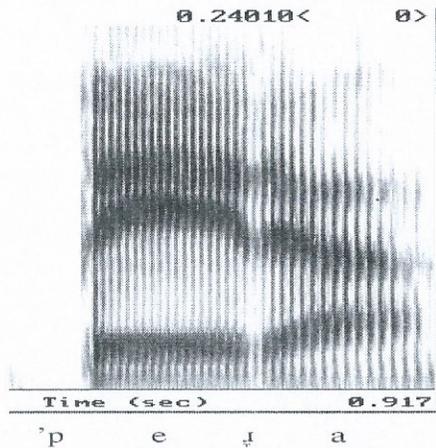
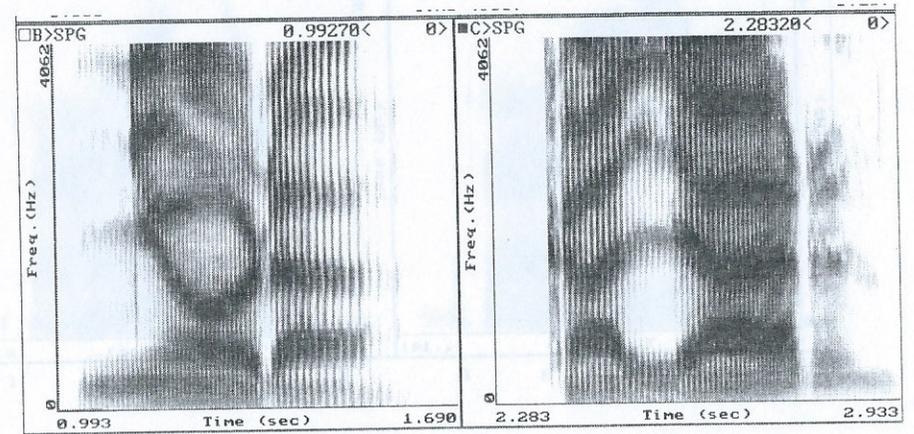


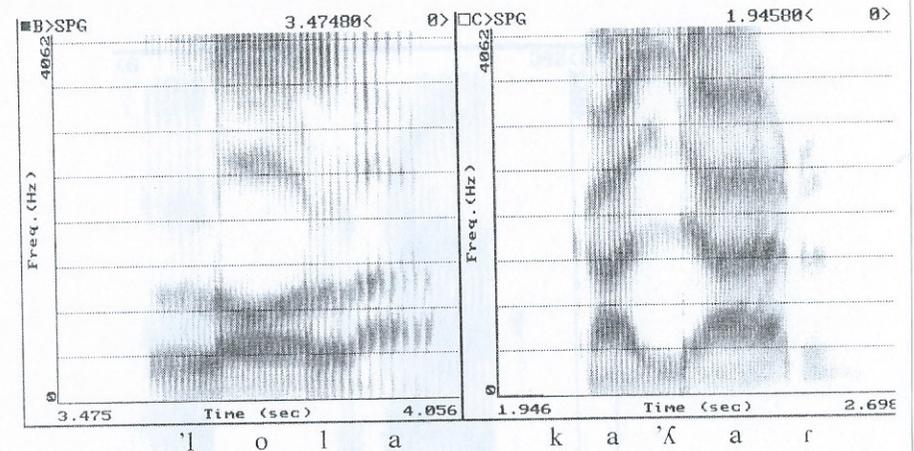
FIG. 8.3. Espectrograma de la aproximante.

Ejercicios

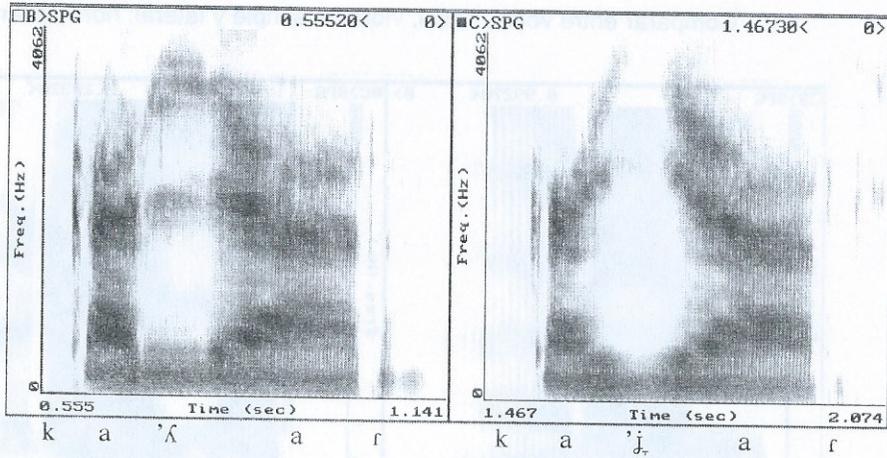
1. Comparar entre vocal, nasal, vibrante simple y lateral: ñora-callar.



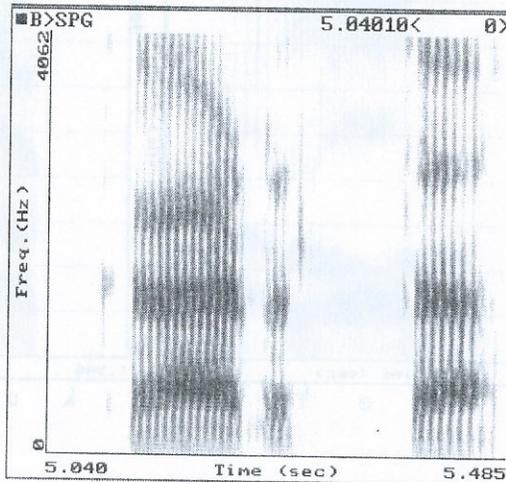
2. Medir pendiente de l/λ.



3. Describir diferencias entre la  $\lambda$  y su sustituta en el yeísmo.



4. Medir duración de la fase oclusiva y la vocálica en la  $r$  de la palabra "carta".



5. Medir la duración del elemento esvarabático en la palabra "tres". ¿La líquida es de tipo vibrante o aproximante?

