

# **REPRESENTACIÓN ESPECTRAL DE LA FONÉTICA ACÚSTICA ESPAÑOLA**

[1] Jesús Bernal, [2] Pedro Gómez y [1] Jesús Bobadilla

[1]

Departamento de Informática Aplicada  
Universidad Politécnica de Madrid  
Ctra. De Valencia Km. 7, 28031 Madrid  
Tfn: +34.913367860, Fax: +34.913367527  
e-mail: [jbernal@eui.upm.es](mailto:jbernal@eui.upm.es), [jbobi@eui.upm.es](mailto:jbobi@eui.upm.es)

[2]

Departamento de Arquitectura y Tecnología de Sistemas Informáticos  
Universidad Politécnica de Madrid  
Campus de Montegancedo, s/n, Boadilla del Monte, 28660 Madrid  
Tfn: +34.913367384, Fax: +34.913367412  
e-mail: [pedro@pino.datsi.fi.upm.es](mailto:pedro@pino.datsi.fi.upm.es)

## RESUMEN

Con este artículo queremos mostrar un conjunto de espectros que abarquen la mayoría de los alófonos que existen en la lengua española.

En un primer apartado se pretende dar una visión global para ver las evoluciones de los formantes y las diferencias espectrales existentes. Todas las grabaciones han sido realizadas por un mismo hablante. Los distintos fonemas que se describen se organizan por el modo de articulación.

En un segundo apartado se presentan los espectros de las cinco vocales realizadas por diferentes locutores.

El método utilizado para el cálculo de los espectrogramas ha sido basado en la Transformada de Fourier, aplicando técnicas originales para la eliminación de ruido.

Se presenta también el resultado de la extracción automática de los formantes mediante algoritmos originales; esto nos permite comprobar la fidelidad de la extracción y su utilización para un procesamiento posterior.

## ABSTRACT

This article pretends to show a set of spectra covering the most important Spanish language allophones.

The first shows an overview of the formant evolutions and the spectra differences in this topic.

The second section presents the five Spanish vowels spectra, pronounced by different speakers.

The spectra graphics have been computed using Fourier Transform analysis.

## 1. INTRODUCCIÓN

La forma tradicional para obtener el espectro de una señal ha sido mediante la utilización del espectrógrafo [Koe46]; más tarde, con la aparición de los ordenadores se utilizó la Transformada Rápida de Fourier (FFT) [Bri88]. En la actualidad existen herramientas que aportan prestaciones avanzadas: captura de formantes, captura del fundamental, programación de los parámetros de la FFT,...

La frecuencia de vibración de las cuerdas vocales es la frecuencia fundamental o *pitch*; las cavidades del conducto vocal actúan de resonadores que potencian o atenúan frecuencias específicas. El resultado a este proceso es la generación de una señal acústica en donde la energía se concentra alrededor de las frecuencias de resonancia, llamadas formantes. Por ello, los formantes y sus evoluciones establecen las características más importantes de la mayoría de los fonemas. También existe otro conjunto de fonemas cuyo sonido no está basado en las cuerdas vocales, sino en fricciones del aire.

**Es pues importante una visualización clara de los formantes del espectro y demás características acústicas, para así facilitar su estudio [Sch95].**

Se comenzó el trabajo aplicando filtros espaciales, pero con unos resultados poco satisfactorios. Nos parecieron interesantes las ideas de K. KONDERA [Kod78] y V. R. CHARI [Cha95], pero tampoco se alcanzaron la metas esperadas.

El método propuesto es absolutamente novedoso y se podría considerar como una etapa previa a los existentes; se podrían aplicar filtros posteriores para mejorar más el aspecto visual del espectro.

En este artículo pretendemos mostrar un conjunto de espectros que han sido procesados para tener una visión global de las representaciones espectrales de la lengua castellana, realizado en un sólo hablante.

Los espectros se han clasificado según el modo de articulación. En

la tablas siguientes tenemos un resumen del conjunto de representaciones.

	Bilabial		Labiodental		Dental		Interdental	
	Sor.	Son.	Sor.	Son.	Sor.	Son.	Sor.	Son.
Oclusivas	[p]	[b]			[t]	[d]		
Fricativas		[β]	[f]				[θ]	[ð]
Africadas								
Nasales		[m]				[n]		
Laterales						[l]		[ɭ]

	Alveolar		Palatal		Velar	
	Sor.	Son.	Sor.	Son.	Sor.	Son.
Oclusivas					[k]	[g]
Fricativas	[s]			[j]	[x]	[ɣ]
Africadas			[tʃ]			
Nasales		[n]		[ɲ]		[ŋ]
Laterales		[l]		[ʎ][l̄]		
Vibrante simple		[r]				
Vibrante doble		[r̄]				

Aquí presentamos la grafía de los alófonos referenciados y algunos ejemplos de ellos.

Alófono	Grafía	Ejemplos
b	b, v	<b>B</b> ote
β	b, v	C <b>v</b> a
d	d	<b>D</b> ada
ð	d	D <b>ð</b> a
g	g, gu	<b>g</b> ama, <b>g</b> uisa, <b>h</b> ongo
ɣ	g, gu	<b>p</b> aga, <b>seg</b> uido, <b>am</b> igo
p	p	<b>P</b> apá
t	t	<b>T</b> apa
k	c, qu, k	<b>c</b> asa, <b>q</b> uita, <b>c</b> osa
m	m	<b>M</b> amá
n	n	<b>n</b> ana, <b>a</b> nc
ŋ		<b>D</b> onde, <b>l</b> ento

ŋ		T <u>engo</u>
ɲ	ñ	Le <u>ña</u>
f	f	F <u>ama</u>
θ	c, z	C <u>ena</u>
s	s	S <u>oy</u>
x	g, j	P <u>aja</u>
ʝ	y, hi	M <u>ayo</u>
tʃ	ch	Pe <u>cho</u>
l	l	A <u>la</u>
l̄	l	Al <u>za</u>
l̄	l	to <u>ldo</u>
l,	l	col <u>cha</u>
ʎ	ll	ll <u>ave</u>
r	r	pa <u>ra</u>
rr̄	r, rr	per <u>ro</u>

## 2. FONEMAS REALIZADOS POR UN MISMO LOCUTOR

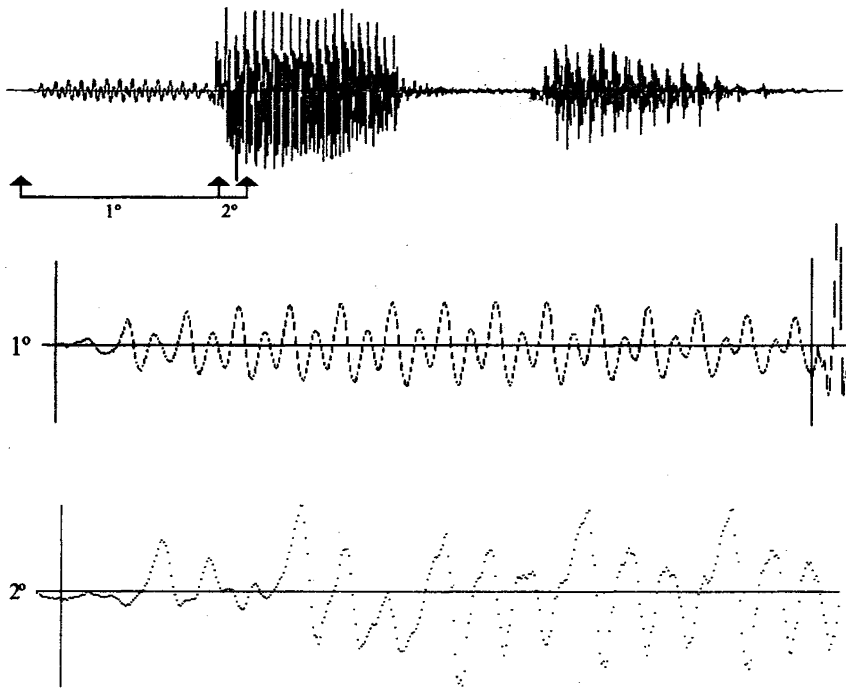
En este apartado se realizará el estudio utilizando un mismo locutor, concretamente el autor de este artículo.

### 1.1. Oclusivas

#### 1.1.1. Fonema /b/

##### Alófono [b]

Bilabial, oclusivo y sonoro. Se produce después de pausa o consonante nasal. Ejemplos: *bote, vaso, cambio*.

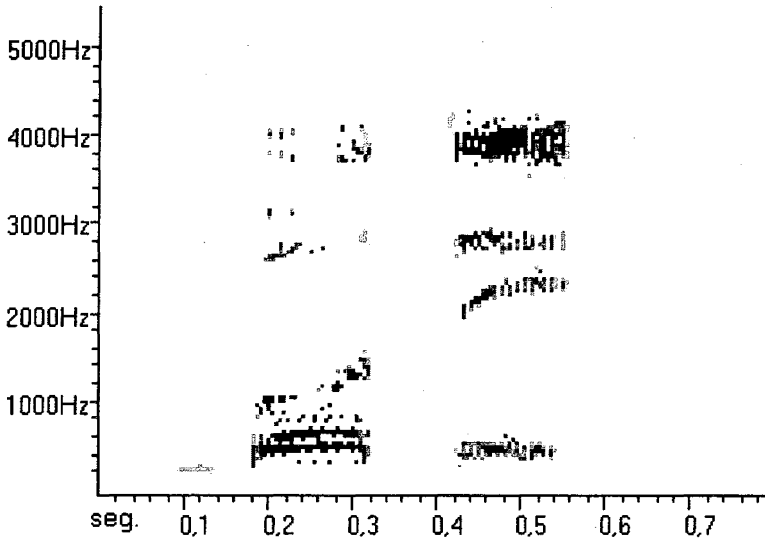


*Figura 1. Realización de 'bote'. En el tramo 1° se inicia la vibración de las cuerdas vocales para la pronunciación de la 'b'. En el tramo 2° se produce la abertura de la boca.*

La Figura 1 representa la señal temporal producida por la pronunciación de la palabra 'bote'. En la gráfica superior se presenta la totalidad de la misma. Las dos gráficas inferiores son ampliaciones de la región marcada en la superior. En el tramo 1° aparecen reflejadas las vibraciones de las cuerdas vocales correspondientes al alófono [b], estando la boca cerrada; comienzan en el instante 48 ms y tienen una duración de 135 ms; además presentan una energía baja.

En el instante 183 ms se inicia la apertura de la boca produciéndose la salida del aire acumulado con brusquedad; de forma paulatina se va formando la vocal 'o'. Las líneas verticales dibujadas en las gráficas marcan los instantes referidos.

En el espectro correspondiente se detecta el inicio de la vibración de las cuerdas vocales en el instante 68 ms con una frecuencia media de 215 Hz; al estar la boca cerrada tiene una energía muy débil. En el instante 186 ms se detecta el comienzo de los formantes de la 'o' con unas frecuencias iniciales de 431 Hz y 904 Hz para los dos primeros formantes. El tercero aparece con 2.584 Hz y el cuarto con 3.790 Hz.



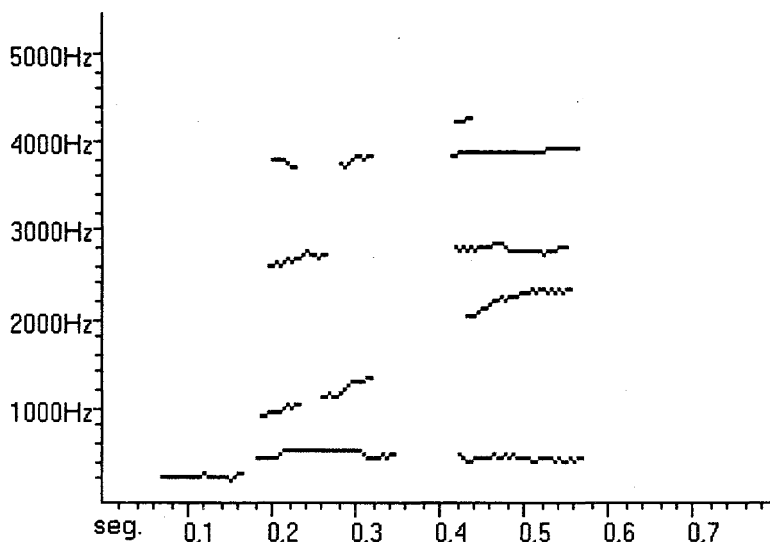


Figura 2. Espectro de la palabra 'bote'.

### Alófono [β]

Bilabial, fricativa y sonora. Ejemplos: *cava*, *pavo*, *abrir*.

Se han realizado dos grabaciones: la primera representa la palabra 'cava' pronunciada con normalidad; en la segunda se ha forzado voluntariamente el fonema /v/ para que sea más fricativo.

En la Figura 3, en la parte superior tenemos el fonema /β/; su representación ha sido amplificada ya que posee poca energía. Comienza en el instante 162 ms hasta 222 ms. En el espectro de la Figura 5, en la palabra 'cava' primero se detecta el tramo fricativo con una energía muy débil. La evolución de los formantes contrasta claramente con el fonema /k/, ya que este último tiene el *locus* en una frecuencia más alta.



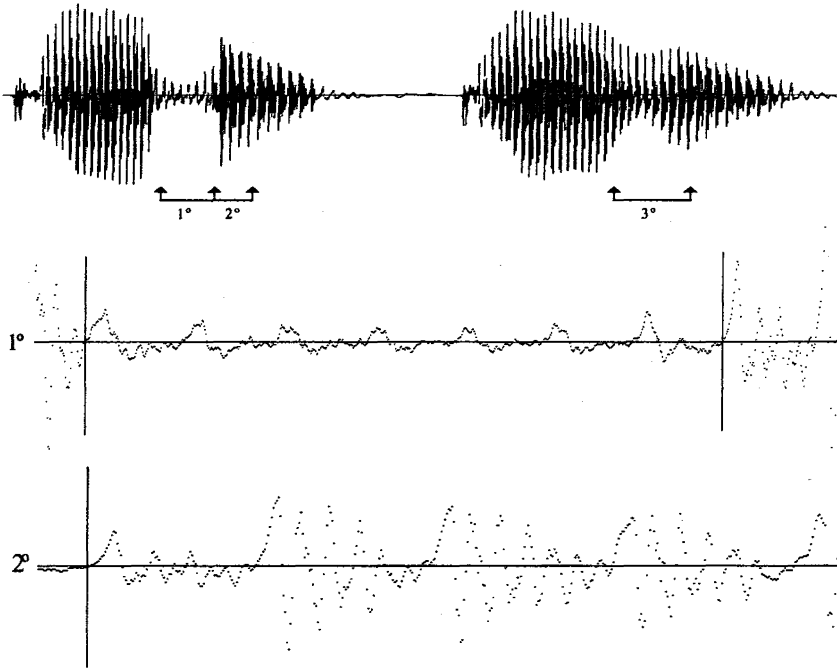


Figura 3. Realización de 'cava'. El tramo 1° corresponde a la 'v'. En el tramo 2° se inicia la apertura de la boca para la pronunciación de la 'a'.

En la segunda grabación de la palabra 'cava' se marca, de forma aproximada, el centro en el instante 674 ms. Se aprecia que no existe ninguna oclusión, los formantes primero y segundo sufren una caída en frecuencias y en energía. Si escuchamos el sonido a partir del punto indicado se percibe el fonema [b] con una explosión muy pobre.

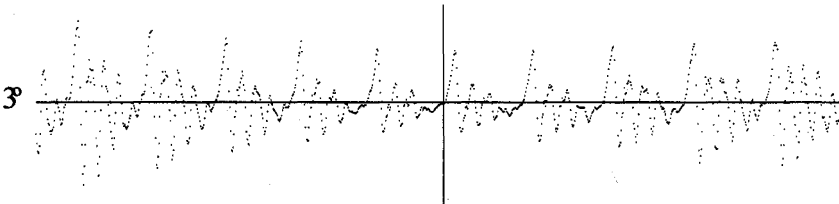


Figura 4. En el tramo 3° se marca el centro de la consonante 'v' de la palabra 'cava'. Se ha forzado para que sea claramente fricativa.

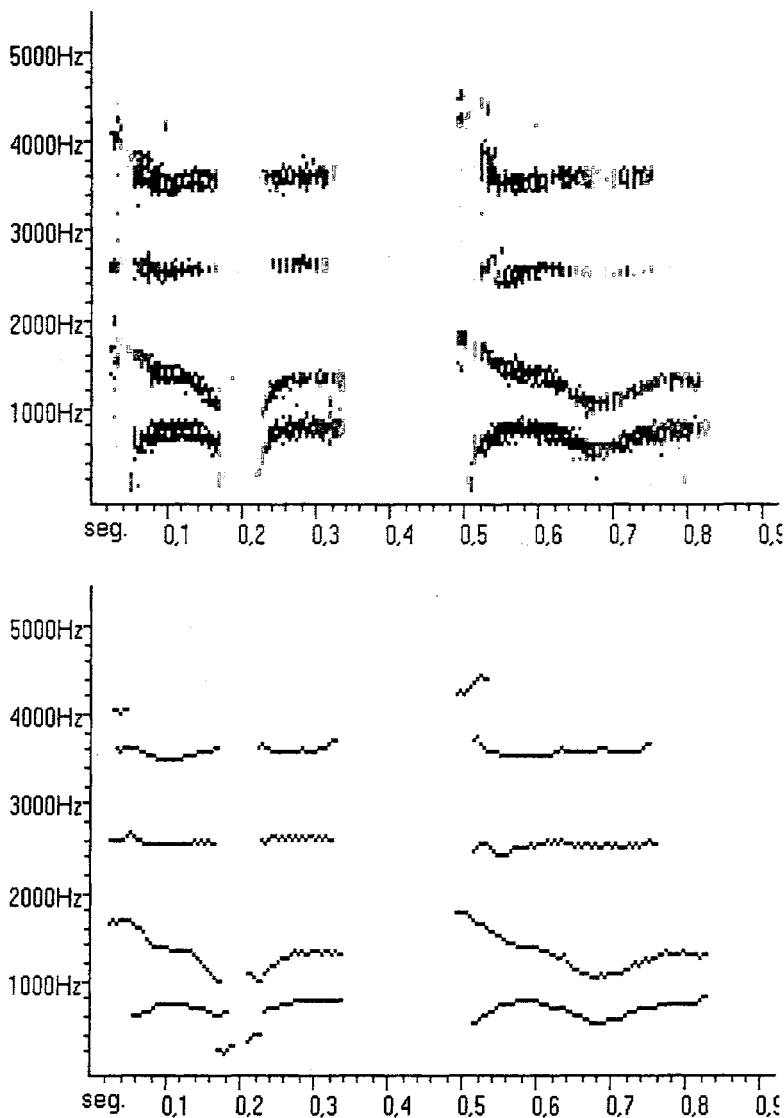


Figura 5. Espectro de la palabra 'cava'. Se ha grabado dos veces; en la parte derecha se ha forzado para que la [β] sea más fricativa.



### 1.1.2. Fonema /d/

#### Alófono [d]

Linguodental, oclusivo y sonoro. Se produce cuando va precedido por una pausa después de consonante nasal y lateral. Ejemplos: *dada*, *donde*, *falda*.

Igual que ocurre con el alófono [b], las cuerdas vocales vibran antes de iniciar la explosión correspondiente al alófono; por ello, en la Figura 6 aparece dicha vibración marcada en el primer tramo. Comienza en el instante 91 ms y dura 80 ms; a partir del instante 171 ms se produce la explosión y se va formando la onda correspondiente a la vocal 'a'.

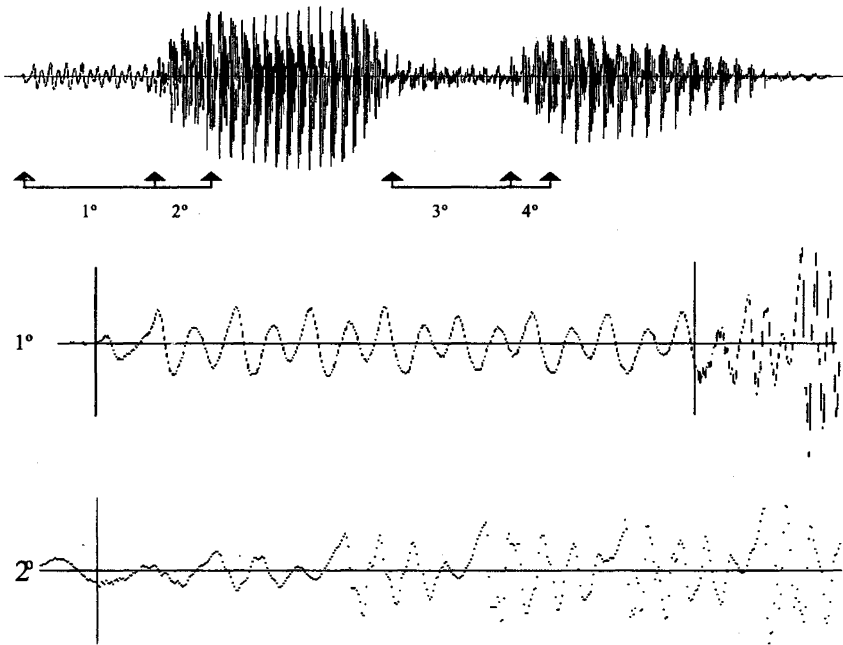


Figura 6. Realización de 'dada'. Los tramos 1° y 2° corresponden al alófono [d].

En el espectro se aprecian los elementos mencionados. La vibración

inicial de las cuerdas se detecta en el instante 95 ms y su frecuencia oscila entre 172 Hz y 215 Hz. Los dos primeros formantes del fonema /a/ tienen frecuencias 732 Hz y 1.421 Hz en mitad de la pronunciación.

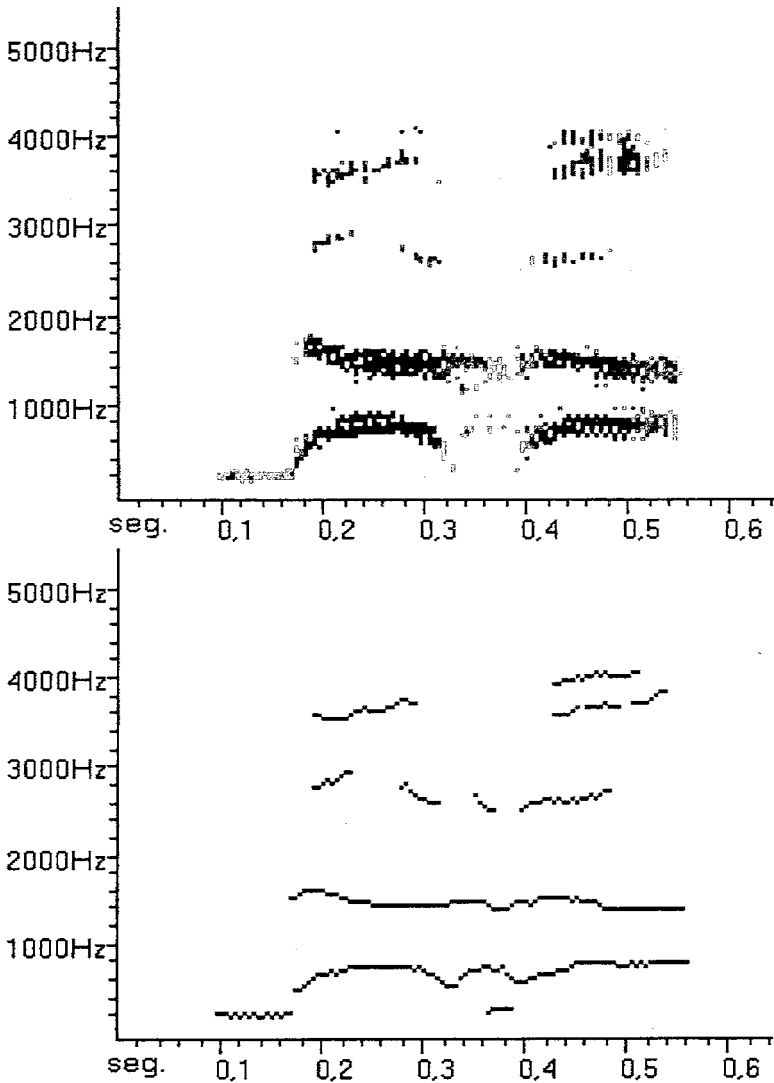


Figura 7. Espectro de la palabra 'dada'.

Alófono [ð]

Linguodental, fricativo y sonoro. Se produce si no va precedido de pausa, ni de nasal, ni de lateral. Ejemplos: *dada*, *codo*, *pardo*.

En la Figura 8, en la gráfica superior, se ha capturado el tramo temporal que corresponde con el alófono [ð]. Es difícil determinar los momentos exactos, pero se podría decir que comienza en el instante 321 ms y dura hasta el instante 397 ms. En el espectro aparece con una energía muy débil.

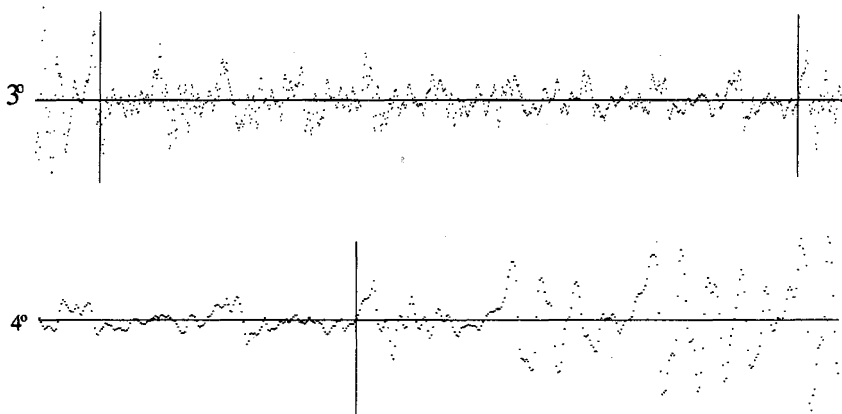


Figura 8. Tramo 3º y 4º correspondiente al alófono [ð] de la palabra 'dada' (ver Figura 6).

Son los dos alófonos típicos de las oclusivas sonoras [b], [d] y [g]. En mitad de palabra, o al principio sin pausa previa, son fricativas por comodidad y fluidez de la articulación. Igual que la [b] se puede forzar para que sea más fricativa, con unos resultados muy similares.

El espectrograma es el mismo que se presentó con el alófono anterior, ver Figura 7.

## 1.1.3. Fonema /g/

Alófono [g]

Linguovelar, oclusivo y sonoro. Precedido de pausa o la nasal /n/.  
Ejemplos: *gama*, *guisa*, *hongo*.

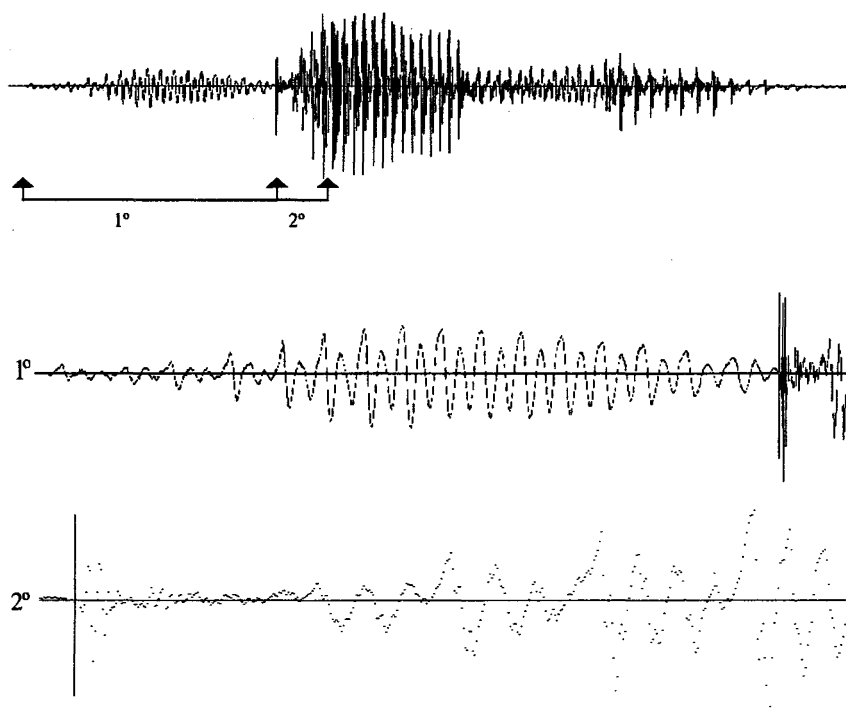


Figura 9. Realización de 'gama'. El tramo 1° representa la sonoridad de las cuerdas vocales. El tramo 2° representa la explosión del alófono [g].

Como todos los fonemas sonoros, aparece la vibración de las cuerdas vocales antes de iniciada la explosión; empieza en el instante 153 ms y dura hasta el 286 ms, momento en que se inicia la oclusión. La vibración de las cuerdas tiene una frecuencia media de 215 Hz.

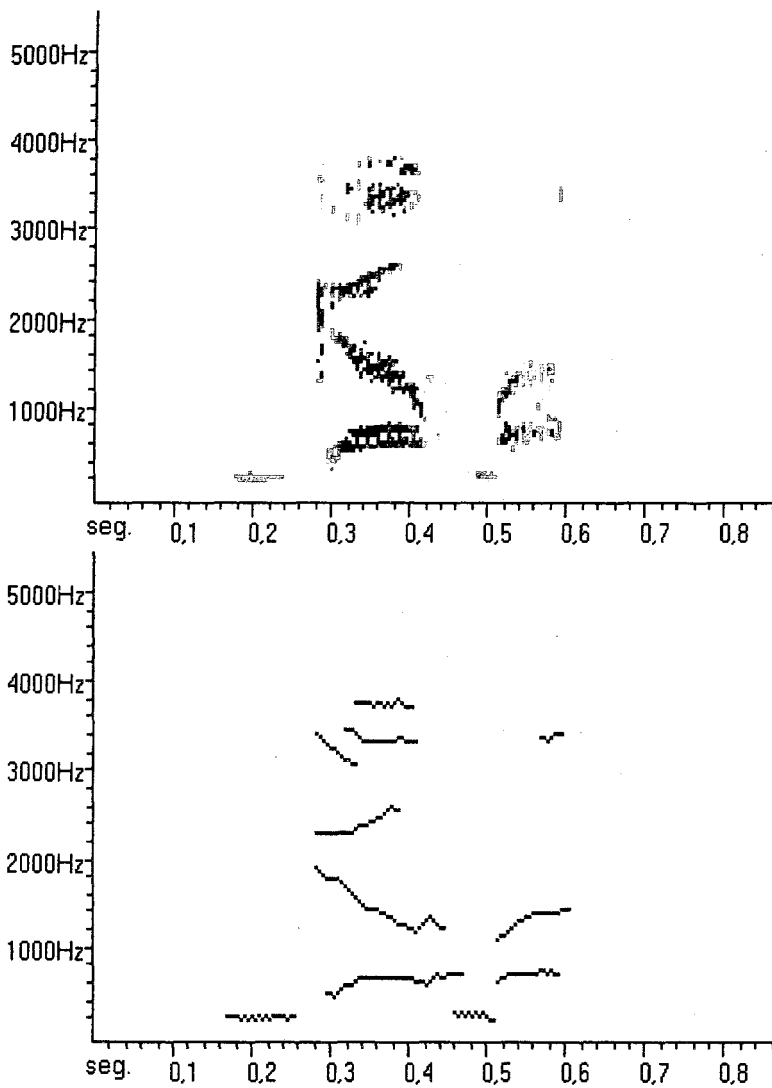


Figura 10. Espectro de la palabra 'gama'.

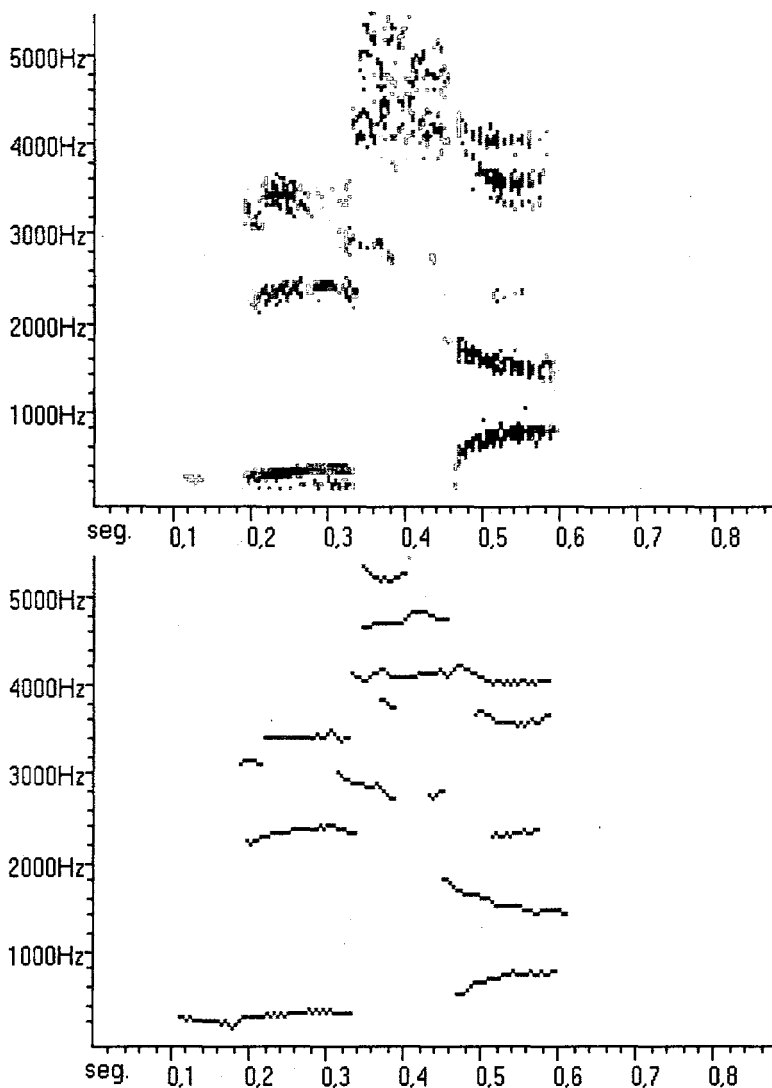


Figura 11. Espectro de la palabra 'guisa'.

Las características espectrales del alófono [g] en la palabra 'guisa' son muy similares a la palabra 'gama'. La diferencia más notable es que la distribución de energías en la barra de oclusión es distinta. La



vibración de la cuerdas vocales se inician en el instante 92 ms y la explosión en el instante 187 ms.

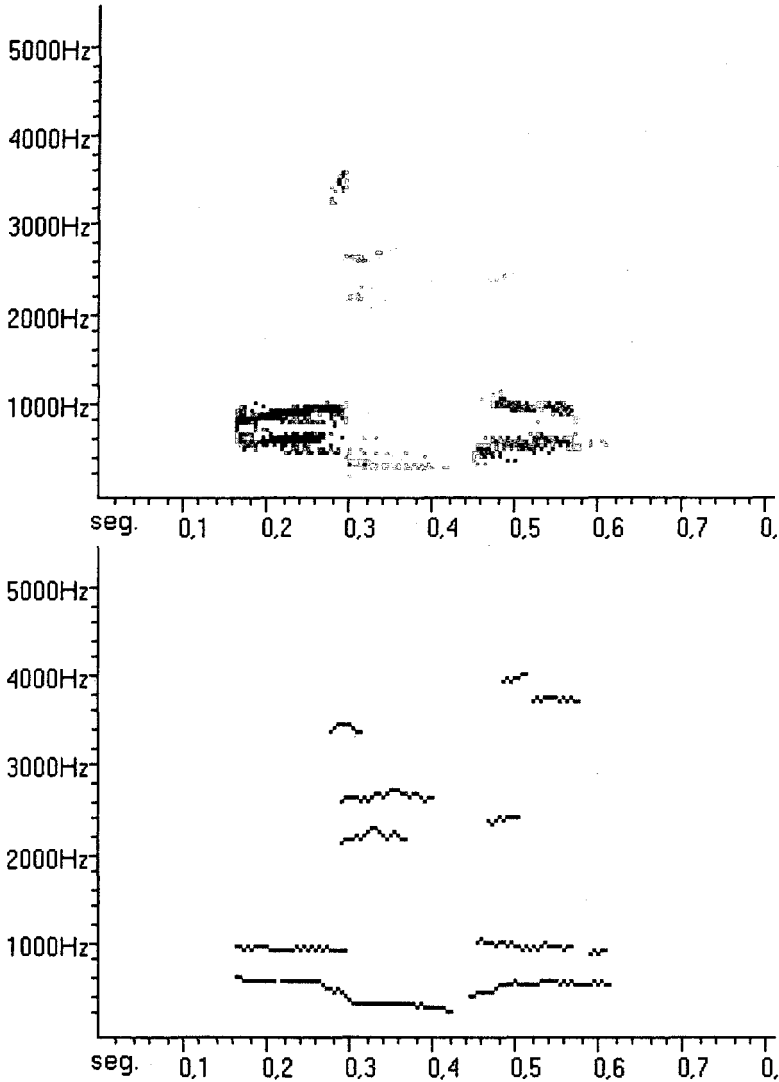


Figura 12. Espectro de la palabra 'hongo'.

Se podría decir que en el instante 443 ms comienza 'go'. No se

aprecia ninguna barra de explosión.

Se han escogido los tres espectros por semejanza a los tres alófonos del fonema /n/; se da una característica muy similar. La energía en la barra de explosión presenta la misma distribución.

### Alófono [ɲ]

Linguovelar, fricativo y sonoro. Si no va precedido de pausa, ni del fonema /n/ y va acompañado de vocal central. Ejemplos: *paga*, *seguido*, *amigo*.

La primera marca está en 316 ms y la segunda en 384 ms; son difíciles de precisar.

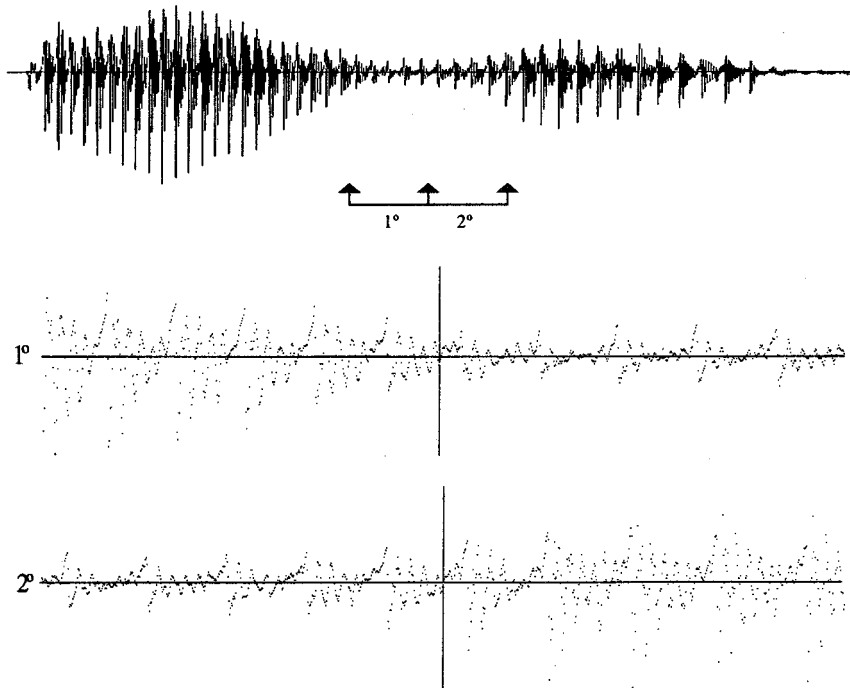


Figura 13. Realización de 'paga'.

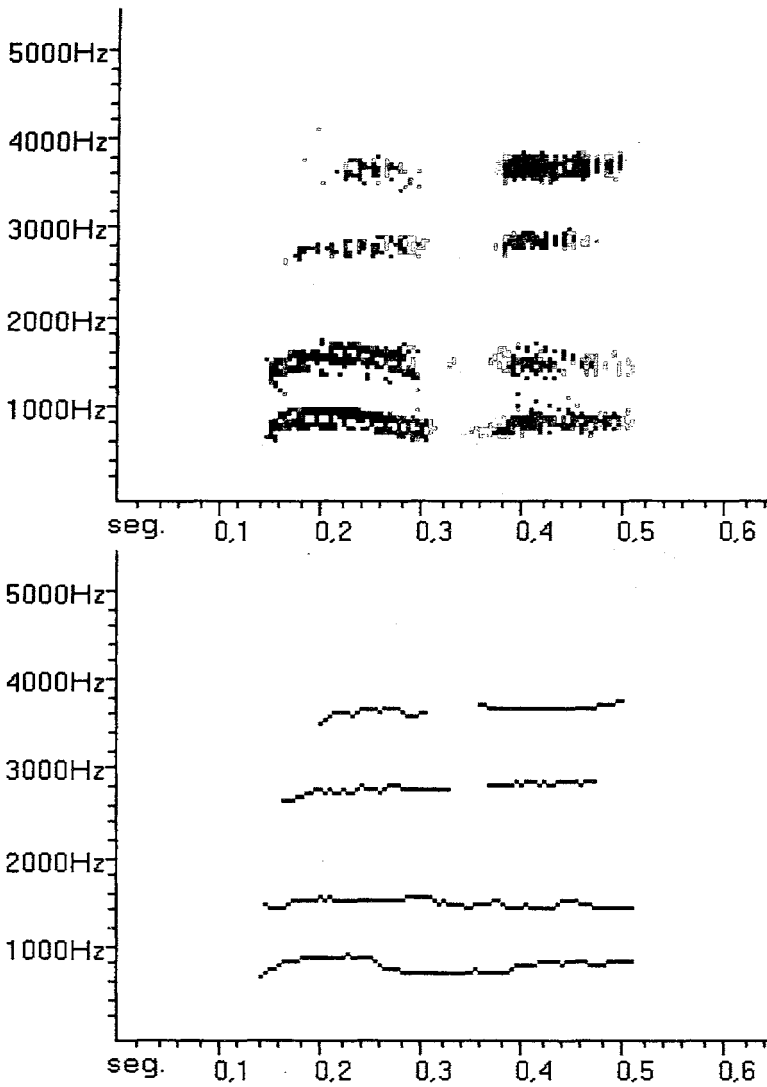


Figura 14. Espectro de la palabra 'paga'.

En el alófono [ɣ] de la Figura 14 se aprecia que la oclusión a desaparecido completamente. A diferencia del alófono [β] los formantes primero y segundo ofrecen una bajada de frecuencia muy tenue.

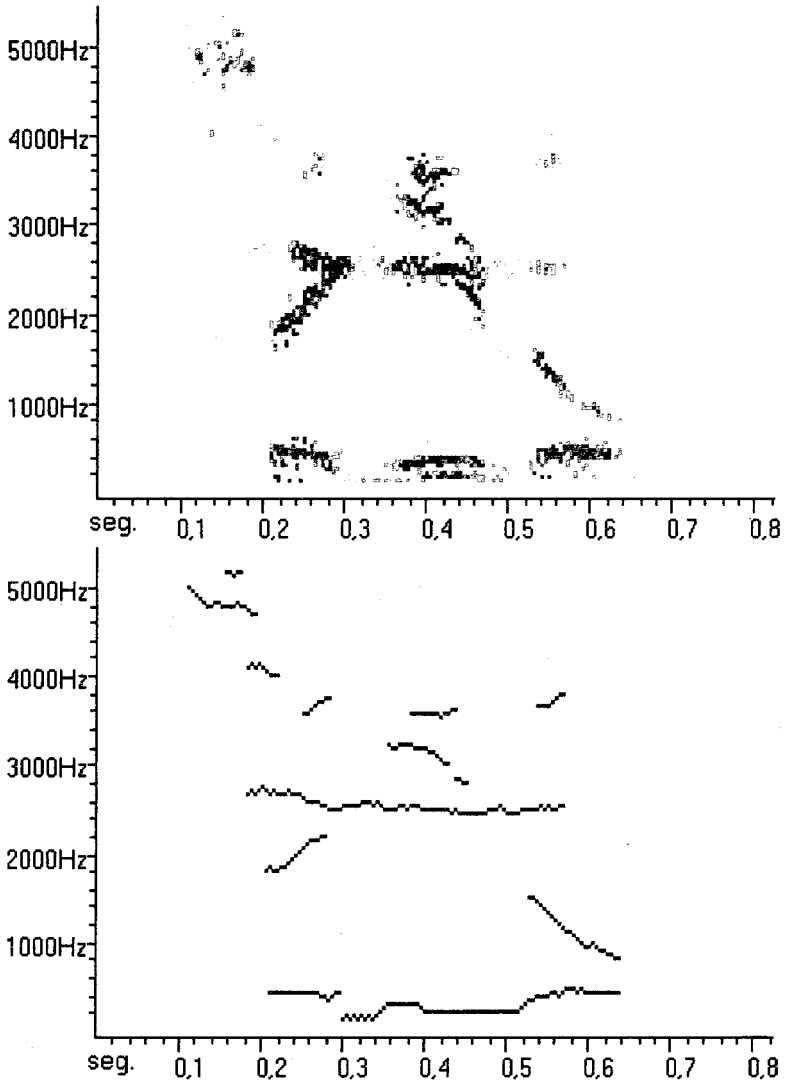


Figura 15. Espectro de la palabra 'seguido'.

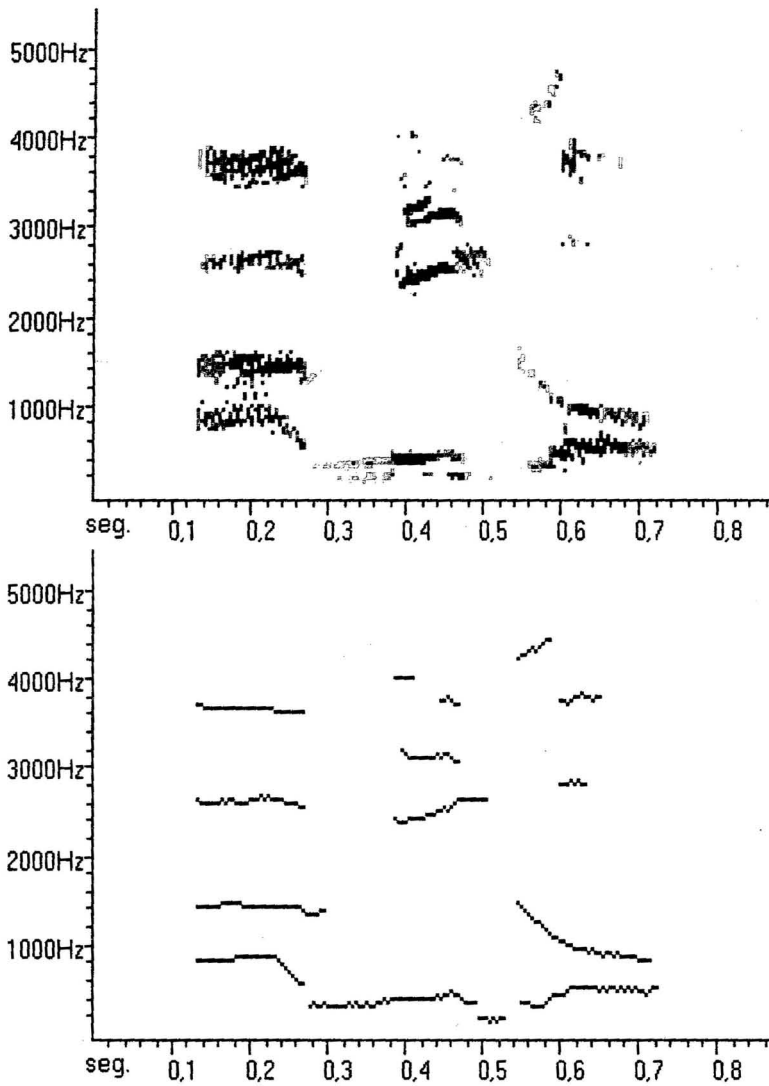


Figura 16. Espectro de la palabra 'amigo'.

#### 1.1.4. Fonema /p/

Bilabial, oclusivo y sordo. Ejemplos: *pedra, capa, opción*.

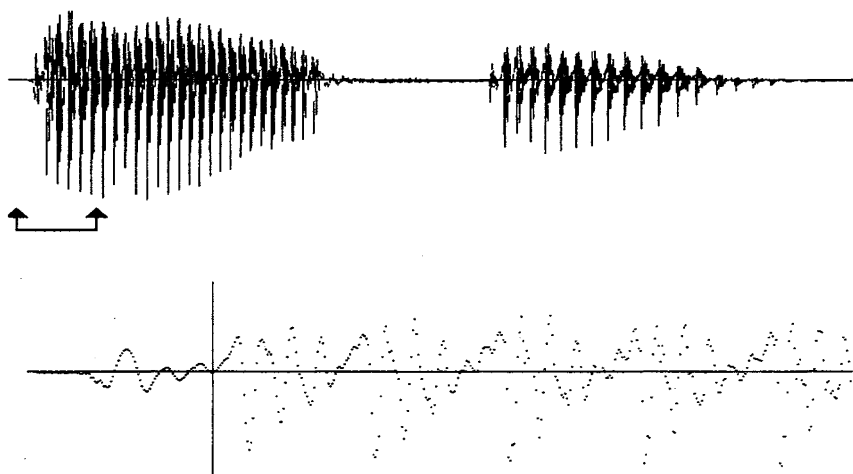


Figura 17. Realización de 'papa'. El tramo representa al alófono [p].

La explosión se inicia en el instante 85 ms y dura hasta 92 ms. Al inicio de los formantes los dos primeros tienen una frecuencia de 689 Hz y 1.206 Hz, en mitad de los mismos son de 818 Hz y 1.464 Hz y al final de 689 Hz y 1.249 Hz.

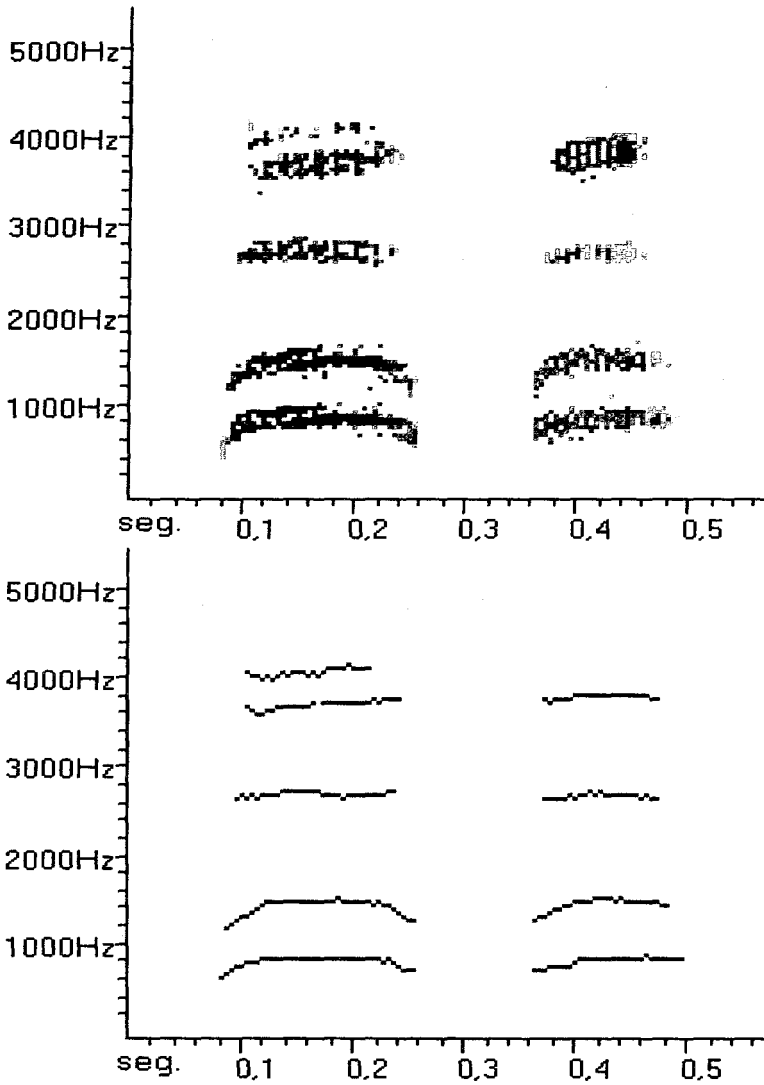


Figura 18. Espectro de la palabra 'papa'.

Se aprecia en la Figura 18 que no aparece la barra de explosión. Viendo el tramo temporal en la Figura 17 la explosión dura 7 ms (recordar que una ventana abarca a 9 ms); ello hace que la barra de

explosión abarque un solo punto gráfico en el eje temporal, y teniendo en cuenta que las ventanas no siempre caen en las posiciones ideales, es difícil que se capture de forma adecuada dicha explosión.

1.1.5. Fonema /t/

Dental, oclusivo y sordo. Se produce en todos los casos excepto cuando le precede un fonema interdental. Ejemplos: *tapa*, *tres*, *atar*.

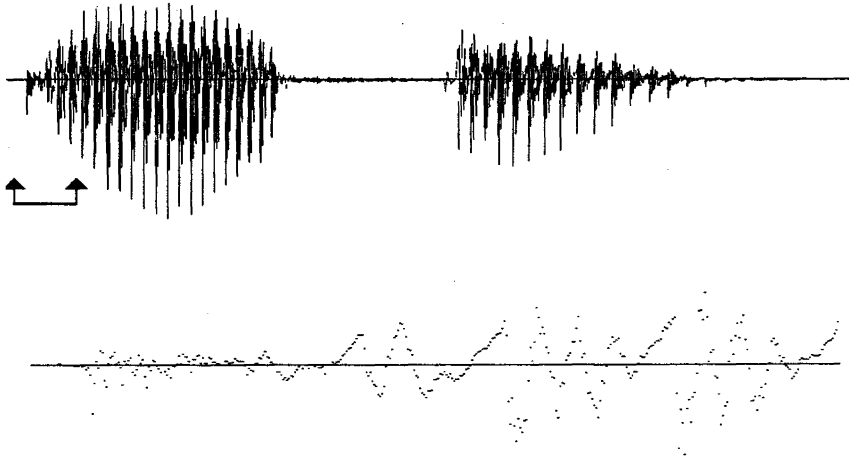


Figura 19. Realización de 'tapa'. El tramo representa el alófono [t].



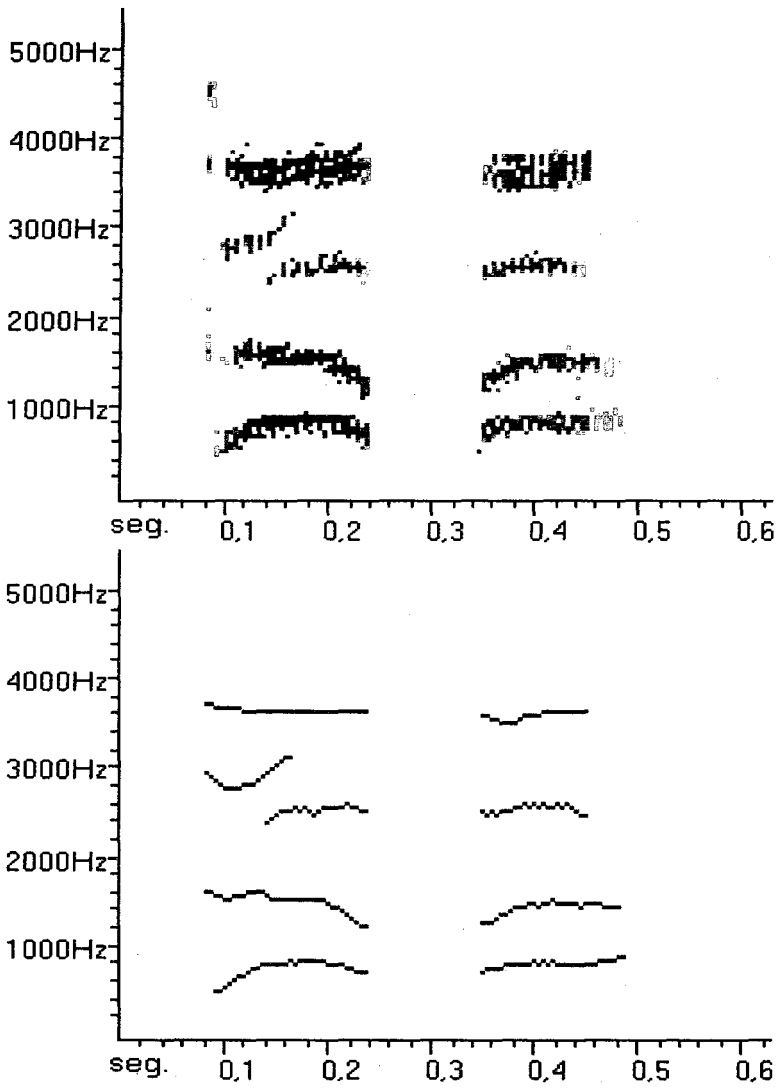


Figura 20. Espectro de la palabra 'tapa'.

A diferencia del fonema /p/, la barra de explosión dura lo suficiente como para poderse representar en el espectro. En la Figura 20 aparece con una duración de varios puntos gráficos del eje temporal.

## 1.1.6. Fonema /k/

Alófono [k]

Linguovelar, oclusivo y sordo. Se da si le sigue la vocal /a/.  
Ejemplos: *casa*, *arca*.

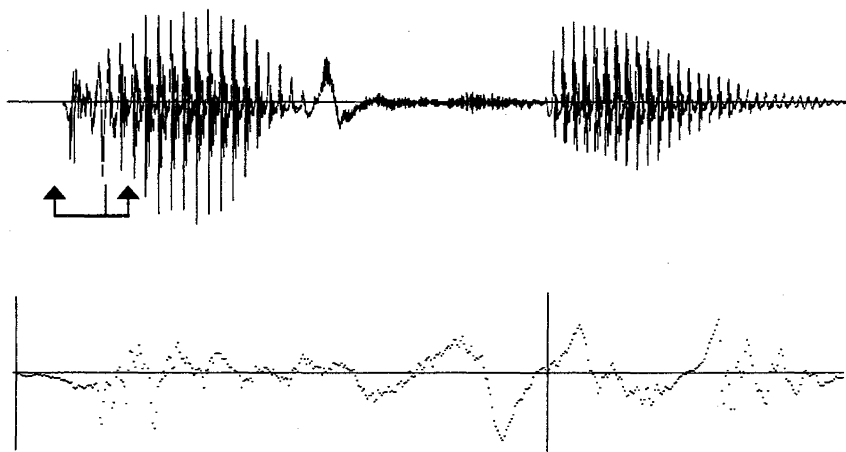


Figura 21. Realización de 'casa'. El tramo corresponde al alófono [k].

En la Figura 21 hemos marcado el tramo temporal referente al alófono [k]; corresponde a la barra de explosión. Comienza en el instante 41 ms y dura aproximadamente 26 ms.

En el espectro de la Figura 22 se aprecia la barra de explosión correspondiente al alófono [k]. En la detección de los formantes se unen los propios de la vocal 'a' con la barra de explosión; ello se debe a que están muy próximos, pero con un estudio de la distribución de energías se podría detectar la presencia de la barra de explosión. Una vez estabilizados, el primer y segundo formante tienen una frecuencia de 689 Hz y 1.680 Hz.

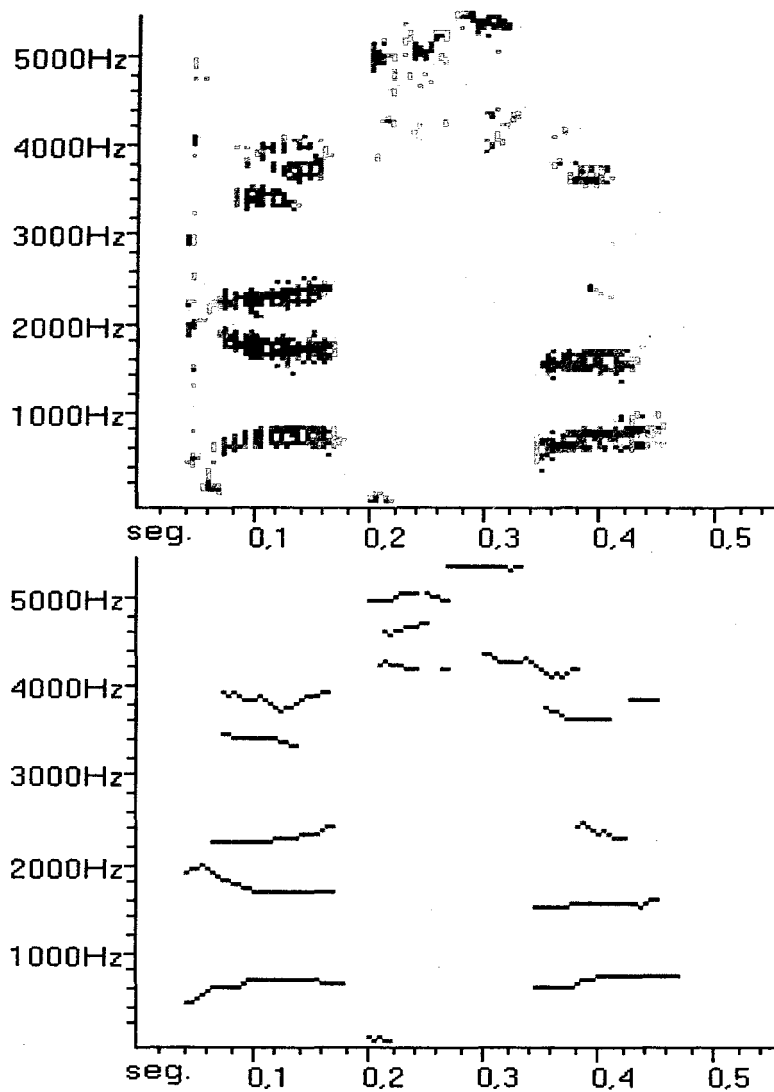


Figura 22. Espectro de la palabra 'casa'.

Alófono [k+]

Linguopostpalatal, oclusivo y sordo. Se da con las vocales /e, i/.

Ejemplos: *quita, kilo, queso.*

En la Figura 23 tenemos la evolución temporal del alófono [k+]; comienza en el instante 78 ms y termina, aproximadamente, en el 111 ms.

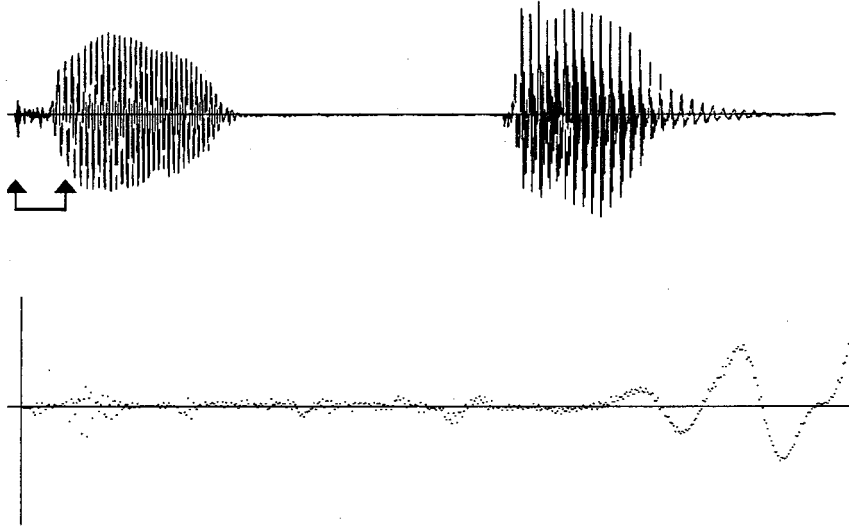


Figura 23. Realización de 'quita'. El tramo representa el alófono [k+].

En este caso existe una mayor concentración de energía en las frecuencias altas. Al estar el formante de la vocal lo suficientemente alejado se detecta de forma independiente.

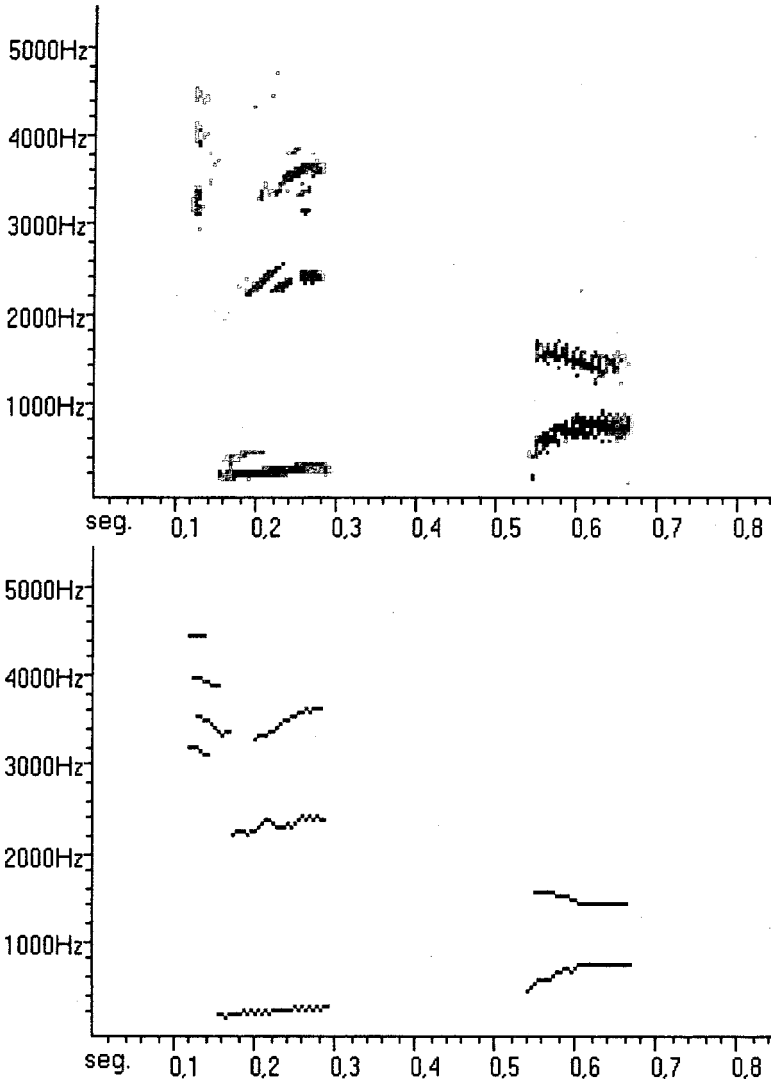


Figura 24. Espectro de la palabra 'quita'.

Alófono [k-]

Linguopostvelar, oclusivo y sordo. Se da con las vocales /o, u/.

Ejemplos: *cosa*, *cupo*, *acuña*r.

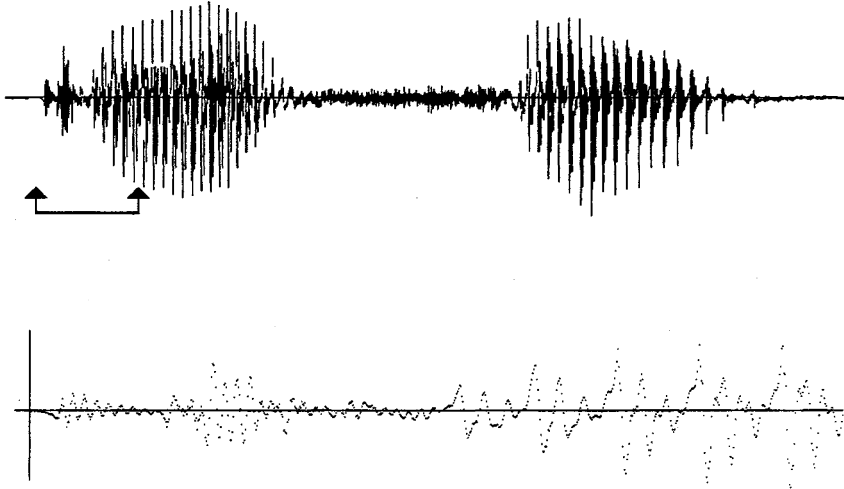


Figura 25. Representación de 'cosa'. El tramo representa el alófono [k-].

En esta grabación comienza en el instante 94 ms y tiene una duración de 36 ms.

Aunque tiene energía en las frecuencias altas, la máxima concentración está alrededor del segundo formante de la vocal 'o'. Los dos primeros formantes se encuentran en las frecuencias 517 Hz y 947 Hz.

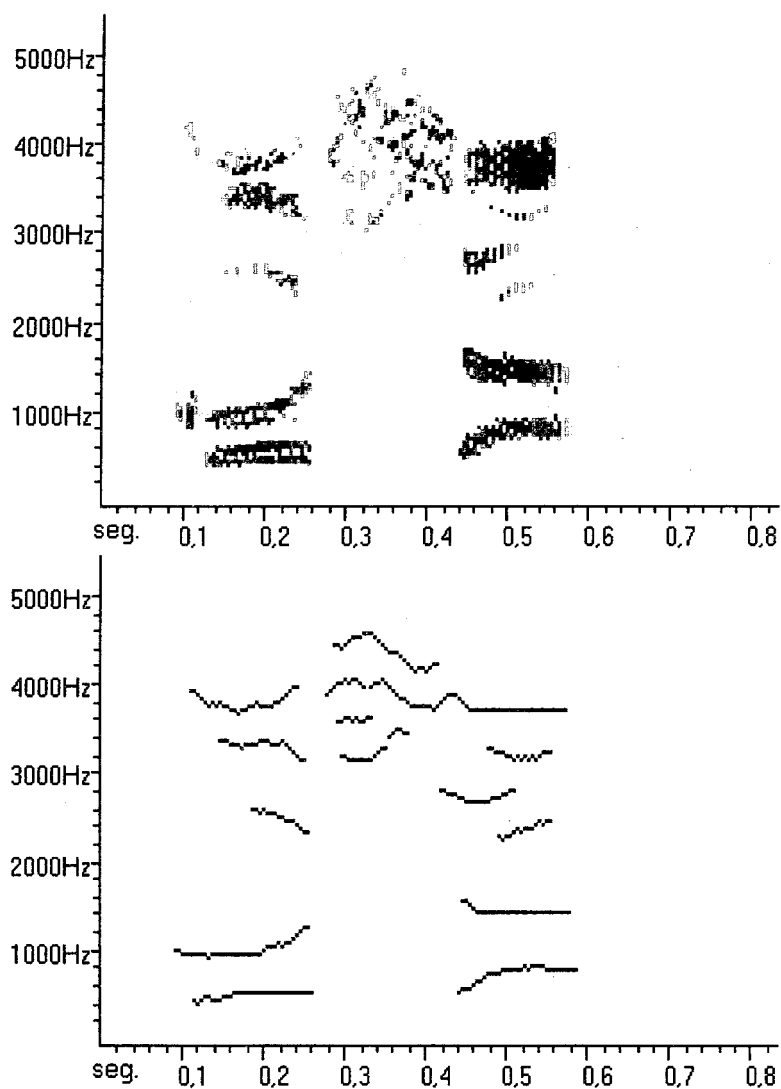


Figura 26. Espectro de la palabra 'cosa'.

La diferencia notable entre los tres alófonos es la distribución de energías dentro de la barra de explosión. Para [k] tiende a repartirse a lo largo de todas las frecuencias, pero con cierta concentración en las

frecuencias superiores, centrales e inferiores. Para la [k+] tiende a concentrarse en las frecuencias altas; y para [k-] se concentra en las frecuencias bajas.

## 1.2. Nasaes

### 1.1.7. Fonema /m/

Bilabial, nasal y sonoro. Ejemplos: *mamá*, *moda*, *bomba*.

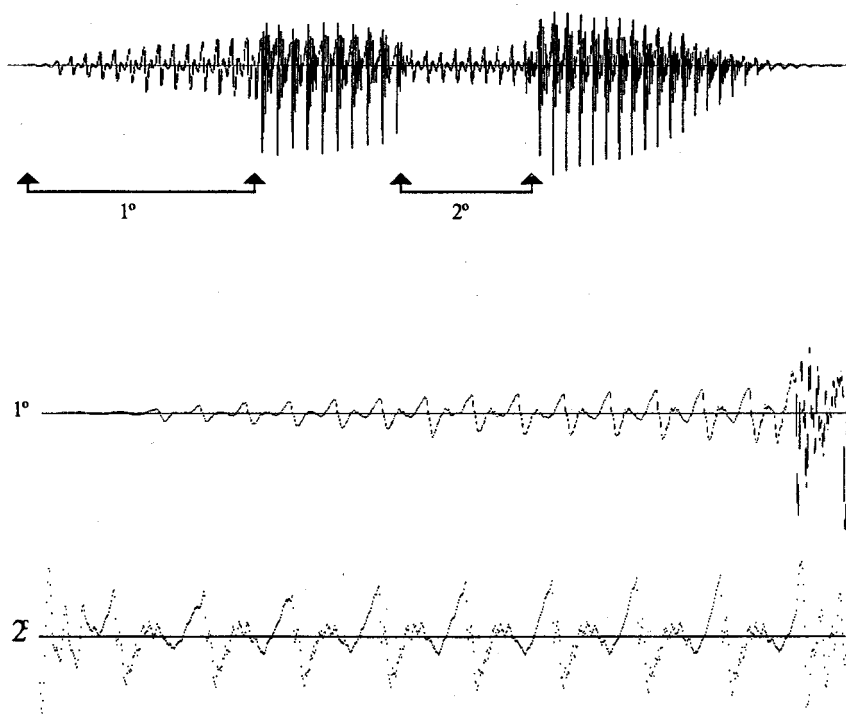


Figura 27. Realización de 'mamá'. El tramo 1° representa la primera 'm' de 'mama' y el tramo 2° representa la 2° 'm'.

En la Figura 22 vemos el tramo temporal. El fonema /m/ comienza en el instante 57 ms y a partir del instante 208 ms se inician los



formantes de la vocal.

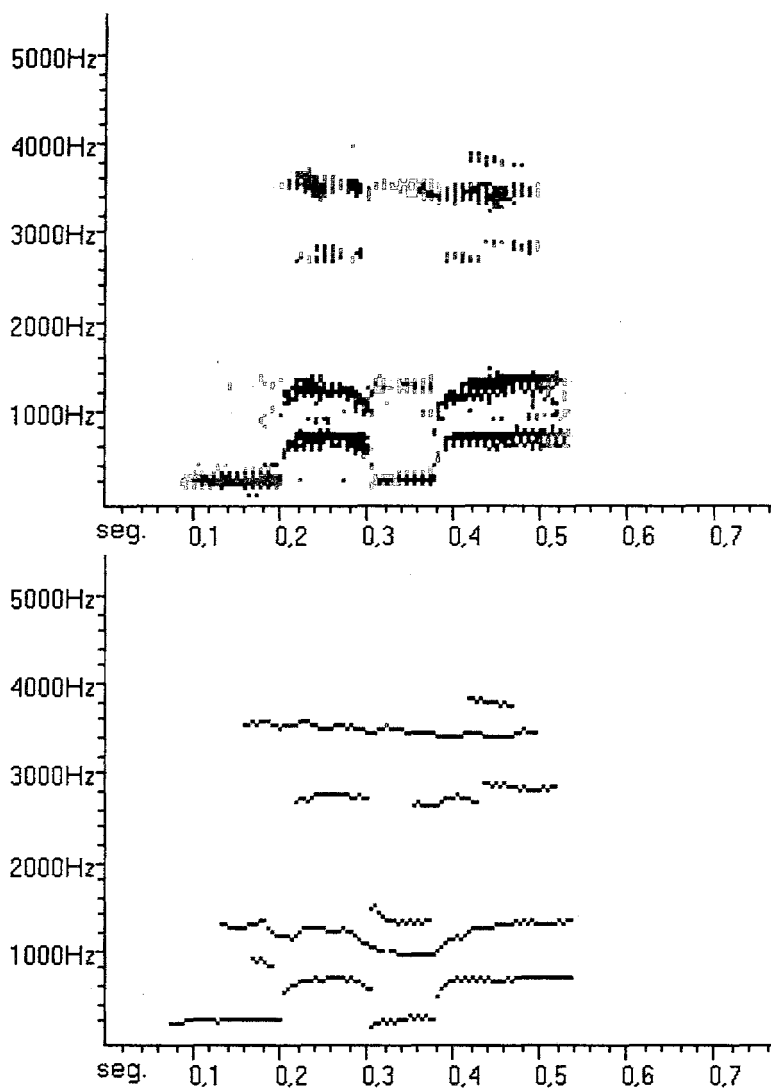


Figura 28. Espectro de la palabra 'mamá'.

Principalmente se detectan tres formantes, con las frecuencias de 215 Hz, 1.249 Hz y 3.531 Hz, aunque es el primer formante el que tiene

más energía.

En la segunda /m/ aparecen también tres formantes con las frecuencias de 215 Hz, 1.292 Hz y 3.488 Hz, muy similares a los de la primera /m/.

#### 1.1.8. Fonema /n/

##### Alófono [n]

Linguoalveolar, nasal y sonoro. Ejemplos: *nana*, *anca*.

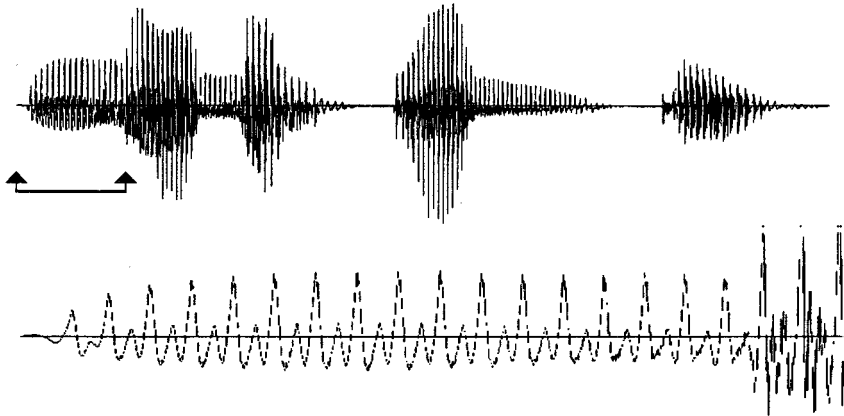


Figura 29. Realización de 'nana anca'. El tramo representa el alófono [n] de la 'nana'.

En la Figura 29 podemos ver el tramo temporal correspondiente al fonema /n/; comienza en el instante 63 ms y a partir del instante 217 ms cambia la señal por el inicio de la 'a'.

Las frecuencias medias de los tres formantes más importantes son: 172 Hz, 1.378 Hz y 2.584 Hz.

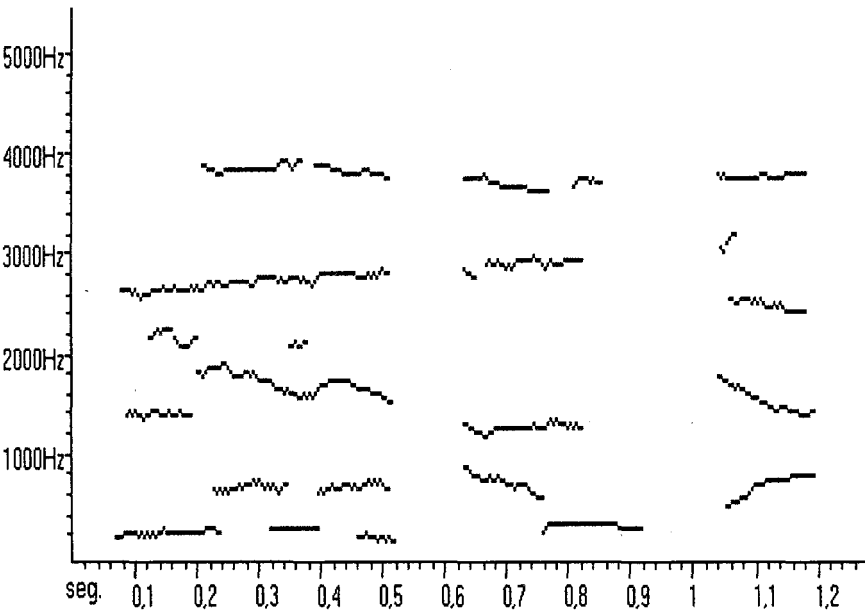
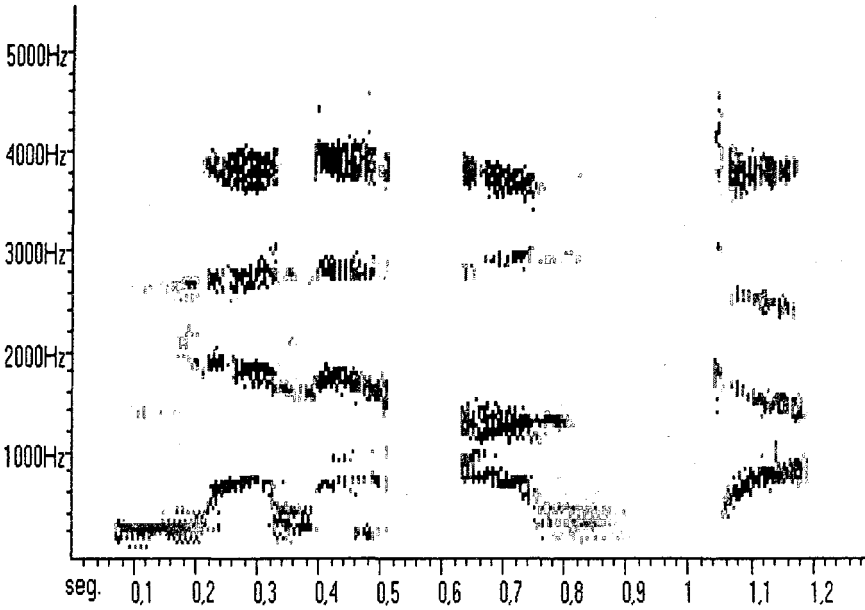


Figura 30. Espectro de las palabras 'nana' y 'anca'.

Se puede comprobar que la energía de los formantes es claramente inferior a los producidos por vocales. Además, tiene un primer formante con una frecuencia media muy baja (172 Hz).

### Alófono [ŋ]

Linguodental, nasal y sonoro. Ejemplos: *donde, lento*.

En la Figura 31 podemos ver el tramo temporal producido por la palabra 'donde'. Se puede considerar que al alófono [ŋ] empieza en el instante 382 ms y tiene una duración de 165 ms.

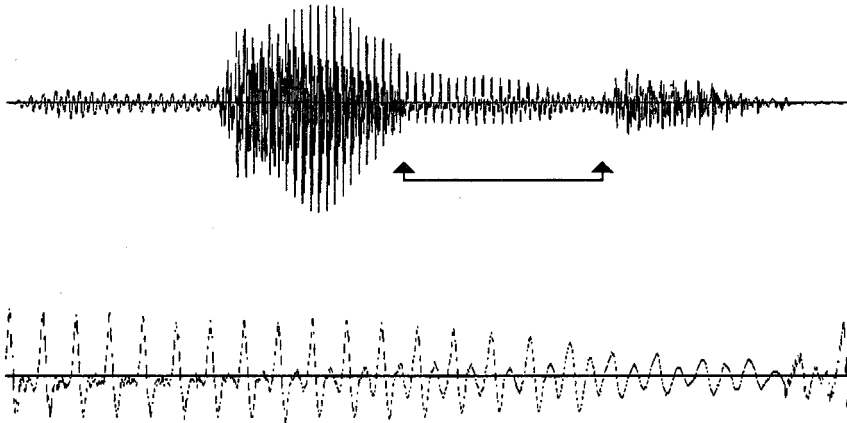


Figura 31. Realización de 'donde'. El tramo corresponde con el alófono [ŋ].

En el espectro se aprecia que el primer formante tiene mucha más energía que los restantes, aparece con una frecuencia de 258 Hz. El tercer formante tiene una frecuencia final de 2.799 Hz. El segundo y cuarto formante son tan tenues que apenas se detectan.

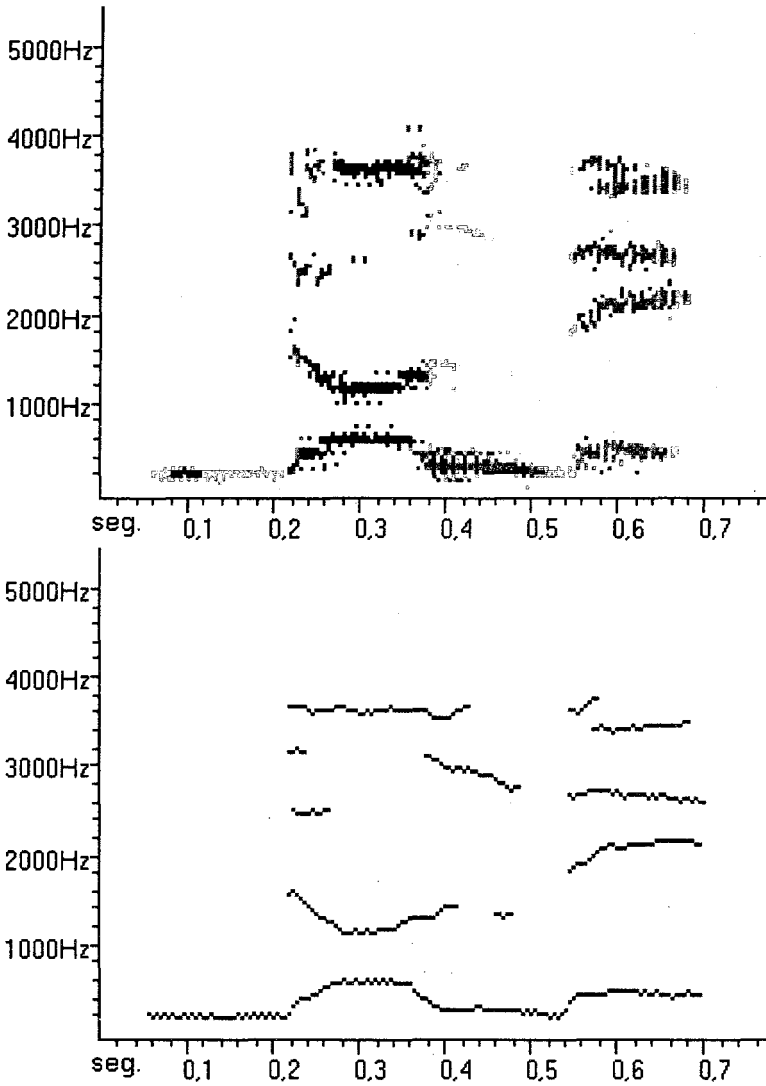


Figura 32. Espectro de las palabra 'donde'.

### Alófono [ŋ]

Lingüovelar, nasal y sonoro. Ejemplos: *tengo*.

La Figura 33 representa el tramo temporal de la palabra 'tengo'. El alófono [ŋ] se inicia en el instante 161 ms y tiene una duración de 166 ms.

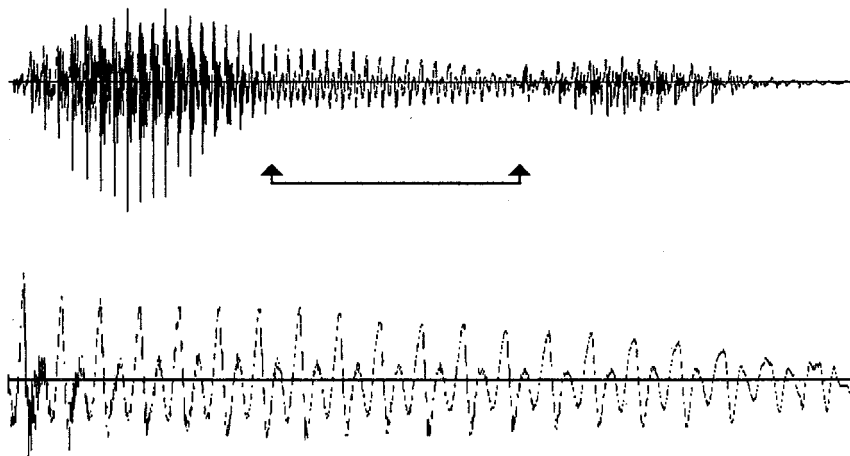


Figura 33. Realización de 'tengo'. El tramo corresponde con el alófono [ŋ].

En el espectro se aprecia que los formantes segundo y tercero tienen más energía que en el alófono [ŋ]. Los tres primeros formantes tienen una frecuencia de 215 Hz, 2.239 Hz y 2.627 Hz respectivamente.

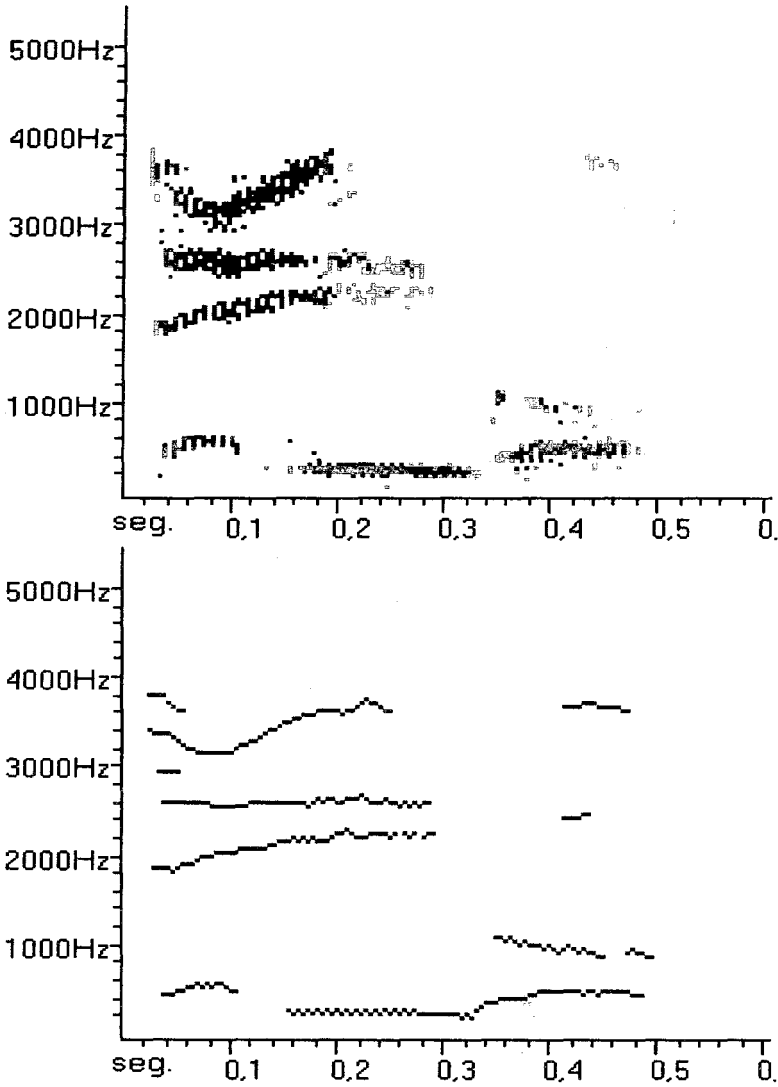


Figura 34. Espectro de las palabra 'tengo'.

1.1.9. Fonema /ɲ/

Linguopalatal, nasal y sordo. Ejemplos: *caña*, *leña*, *añoranza*.

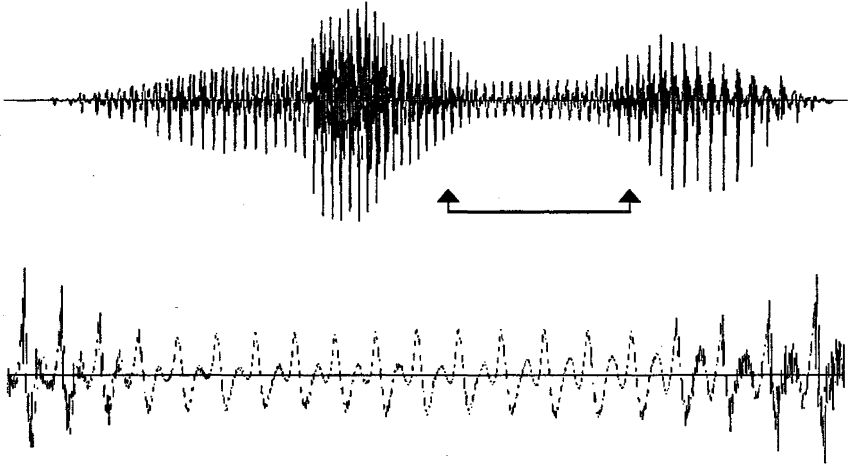


Figura 35. Realización de 'leña'. El tramo corresponde con el alófono /ɲ/.

Al tener los formantes cierta continuidad, la señal temporal no cambia con la brusquedad suficiente como para detectar en qué instante se produce, pero aproximadamente va desde el instante 394 ms hasta el 487 ms.



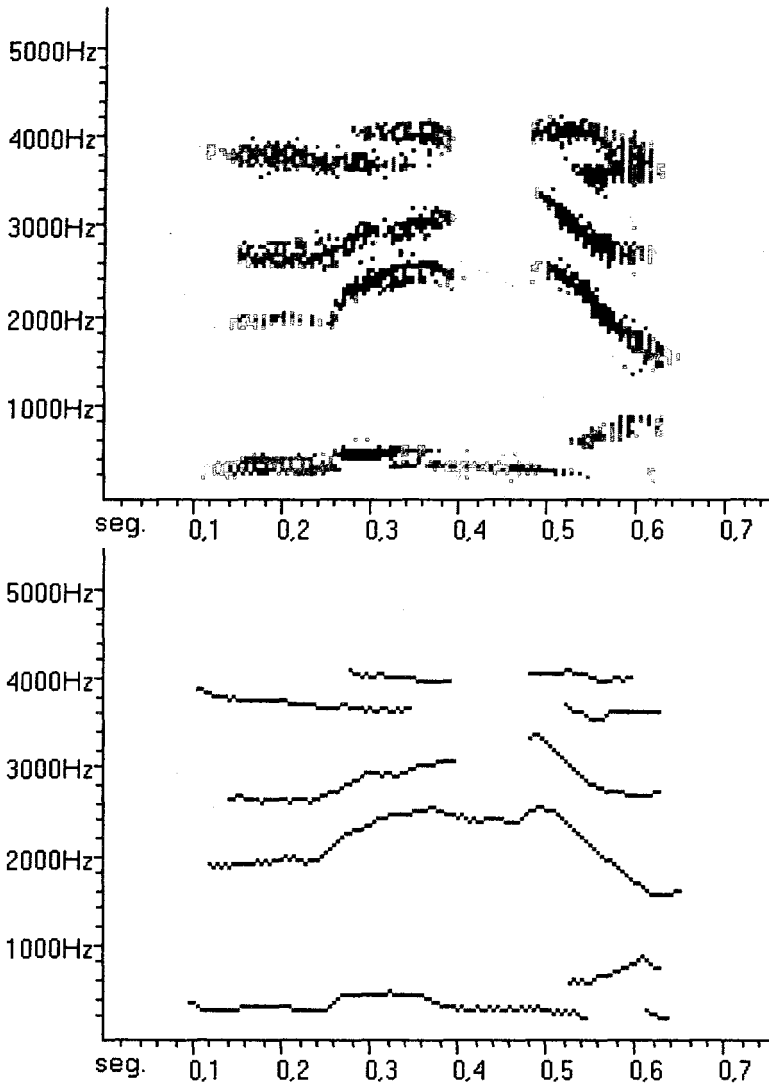


Figura 36. Espectro de la palabra 'leña'.

Tiene una frecuencia media de formantes de 258 Hz y 2.412 Hz. La energía de los formantes es inferior a la energía de los formantes vocálicos y similar a los otros fonemas nasales. Se aprecia en el espectro

la fuerte caída en frecuencias de los formantes segundo y tercero de la vocal posterior.

### 1.3. Fricativas

#### 1.1.10. Fonema /f/

Labiodental, fricativo y sordo. Ejemplos: *fama*, *café*, *feliz*.

Al ser un fonema fricativo, la primera parte corresponde al sonido del aire al paso por alguna región estrecha. En la Figura 6-33 está marcada la fricción del aire, comienza en el instante 106 ms y tiene una duración de 66 ms.

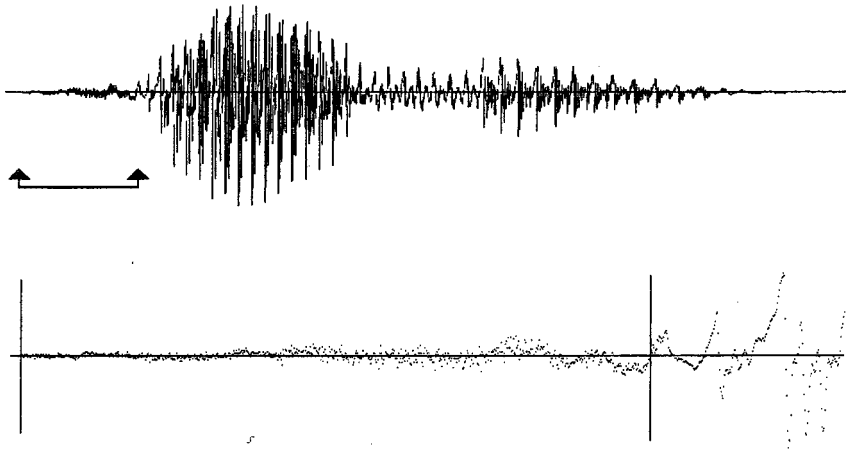


Figura 37. Realización de 'fama'. El tramo correspondiente con el alófono [f].

En el espectro se aprecia cómo la fricción corresponde a frecuencias altas. En general dispone de poca energía. La vocal siguiente, la 'a', tiene sus dos primeros formantes a 689 Hz y 1.249 Hz en el instante 231 ms, en mitad de la pronunciación.

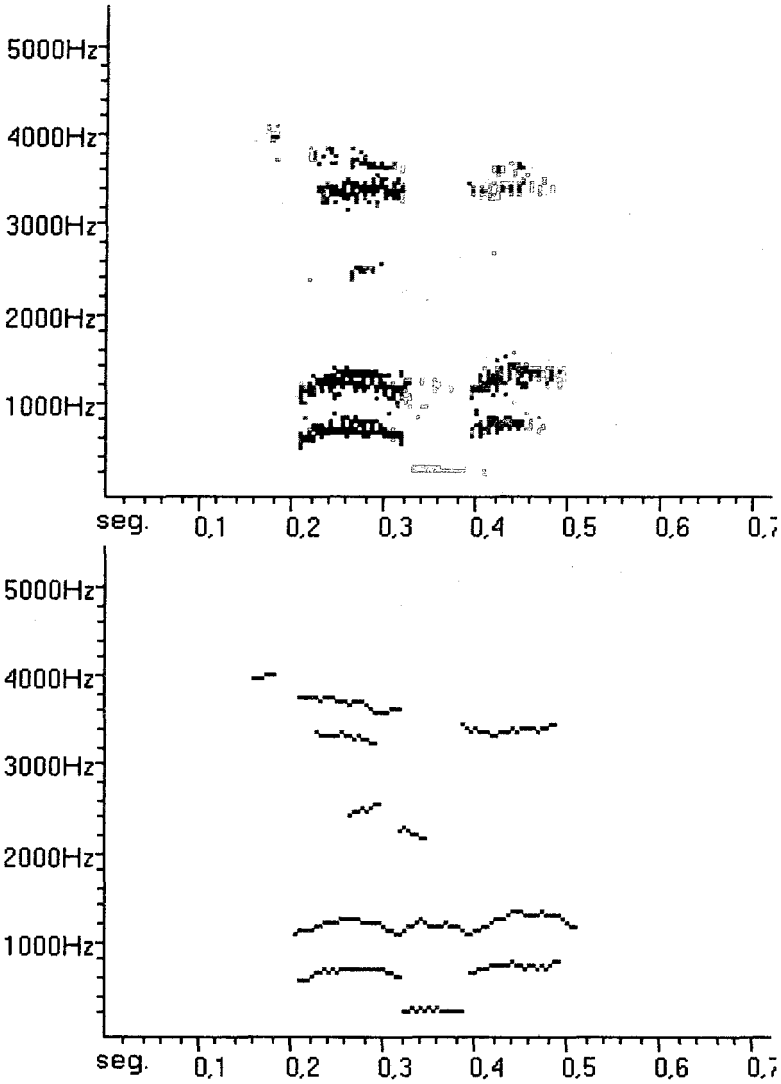


Figura 38. Espectro de la palabra 'fama'.

Entre los fonemas fricativos, éste es el que tiene menor energía. Es un ruido muy similar al producido por el propio micrófono; de hecho, se eliminó la parte de la fricción de la /f/ y se substituyó por ruido del

micrófono amplificado y no se aprecia diferencia.

### 1.1.11. Fonema /θ/

Interdental, fricativo y sordo. Ejemplos: *cena*, *caza*, *cocer*.

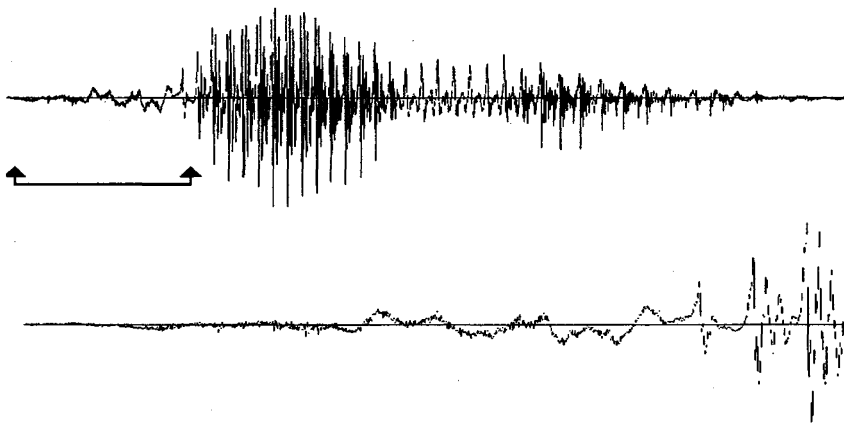


Figura 39. Realización de 'cena'. El tramo corresponde con el alófono [θ].

En la Figura 39 se muestra el espectro del fonema /θ/ dentro de la palabra 'cena'. Tiene una energía muy débil, similar al fonema /f/.

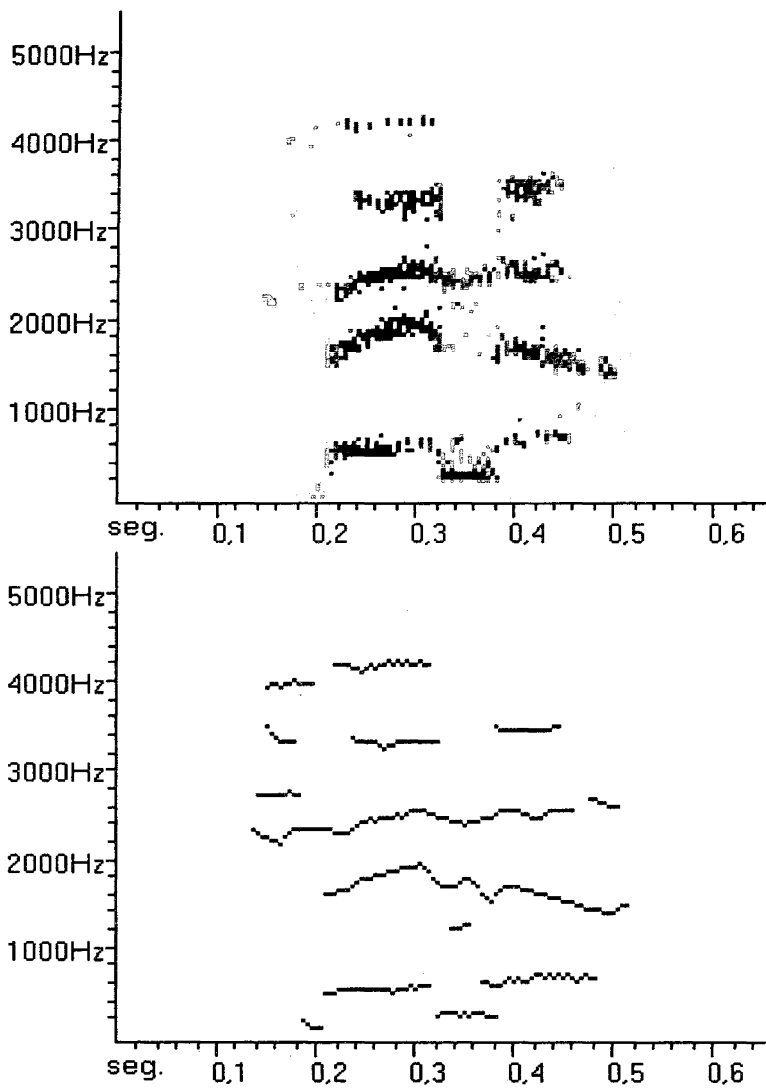
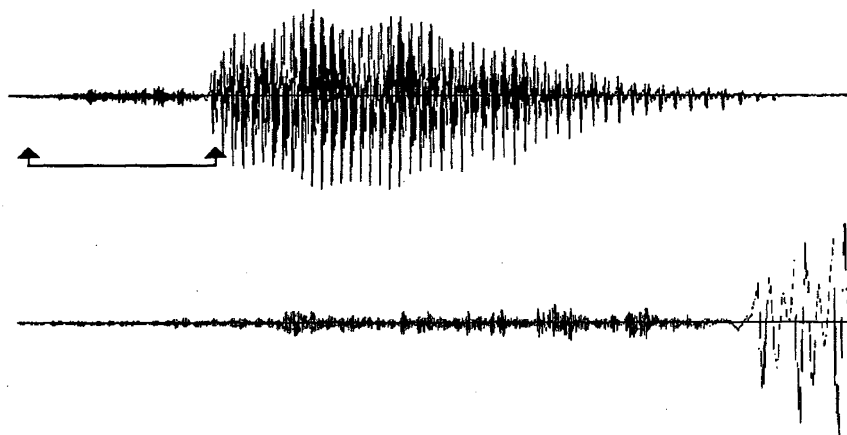


Figura 40. Espectro de la palabra 'cena'.

### 1.1.12. Fonema /s/

Linguoalveolar, fricativo y sordo. Se da cuando no precede a

ninguna consonante sonora. Ejemplos: *soy, dos, casa*.



*Figura 41. Tramo temporal inicial de la palabra 'soy'.*

En la Figura 41 se presenta el tramo temporal del fonema /s/, de la palabra 'soy'. Comienza en el instante 113 ms y dura hasta el 223 ms.

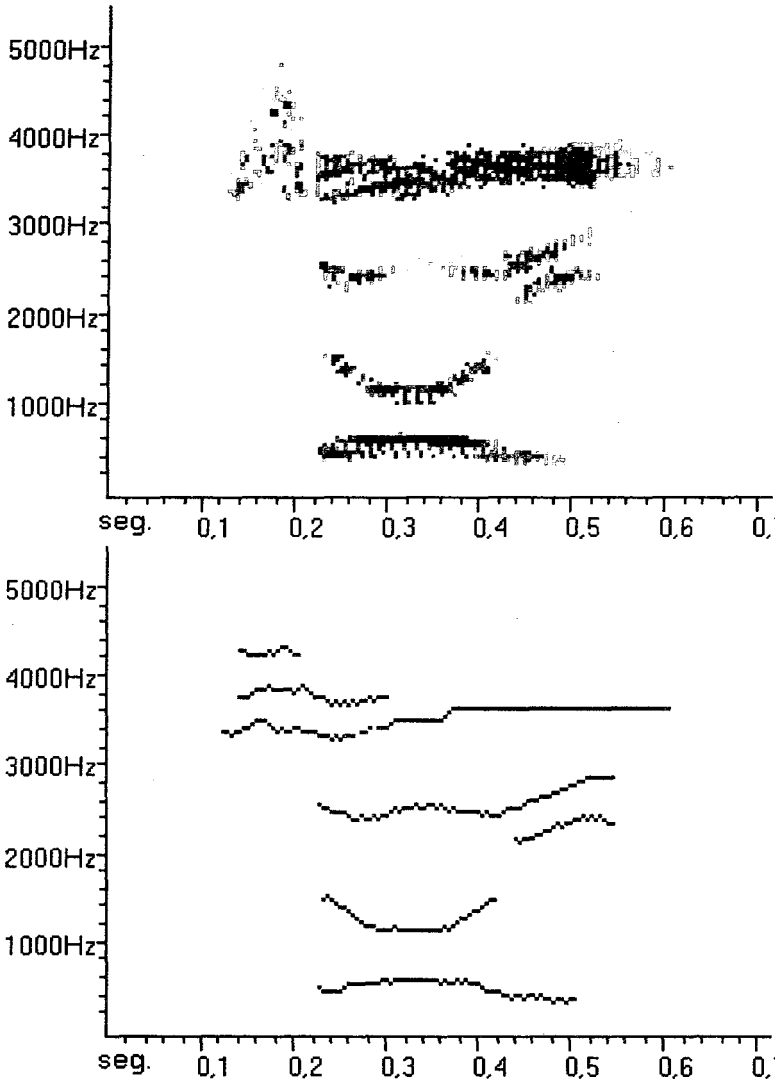


Figura 42. Espectro de la palabra 'soy'.

### 1.1.13. Fonema /x/

Linguovelar, fricativo y sordo. Ejemplos: *paja*, *gitano*.

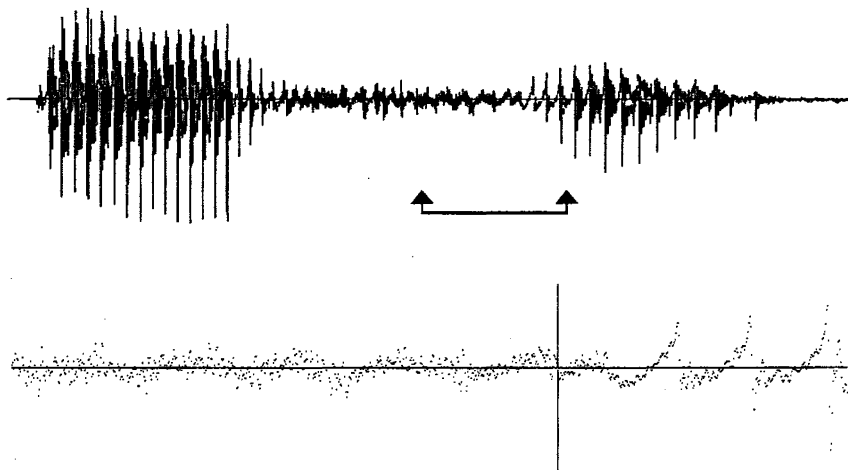


Figura 43. Realización de 'paja'. El tramo corresponde con el alófono [x].

En la Figura 43 presentamos la señal temporal. Aunque no se puede precisar con exactitud, se marca el instante 369 ms como el inicio donde se empieza a formar la onda de la vocal.



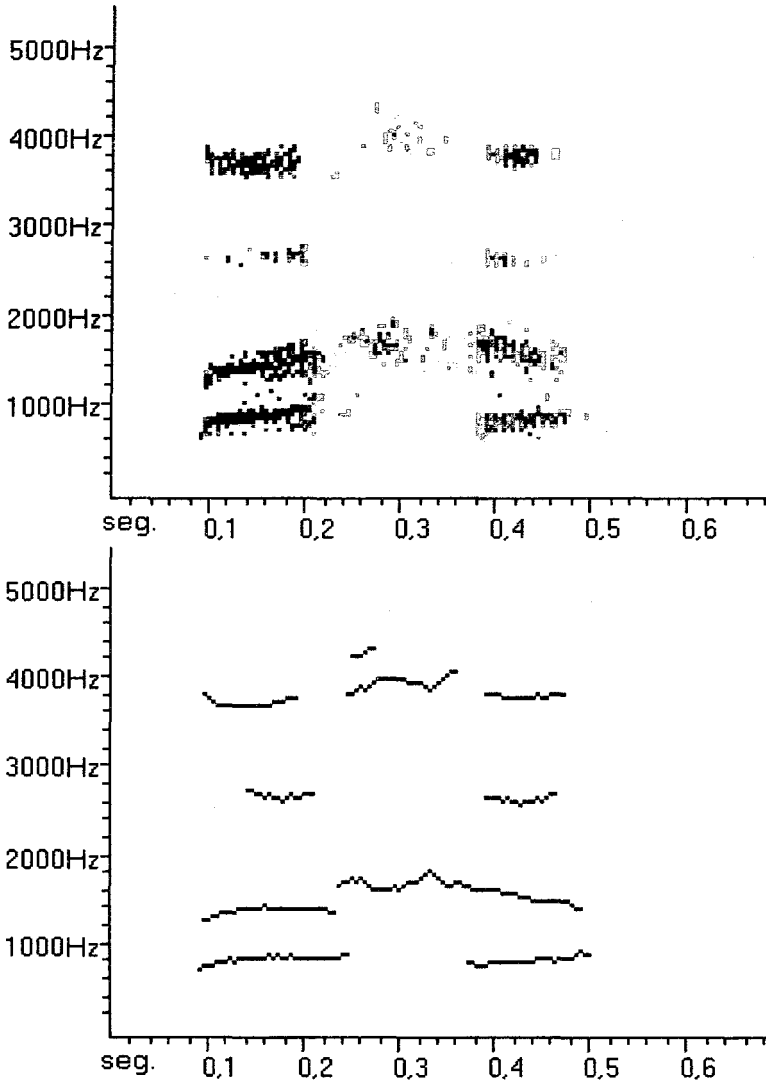


Figura 44. Espectro de la palabra 'paja'.

Una característica fundamental que la distingue de otras fricativas es su energía en frecuencias relativamente bajas (en la zona de 1.800 Hz); también posee energía en las zonas altas.

## 1.1.14. Fonema /j/

Palatal, fricativo y sonoro. Se da cuando no se encuentra precedido ni por pausa, ni por nasal o lateral. Ejemplos: *mayo*, la *hierba*.

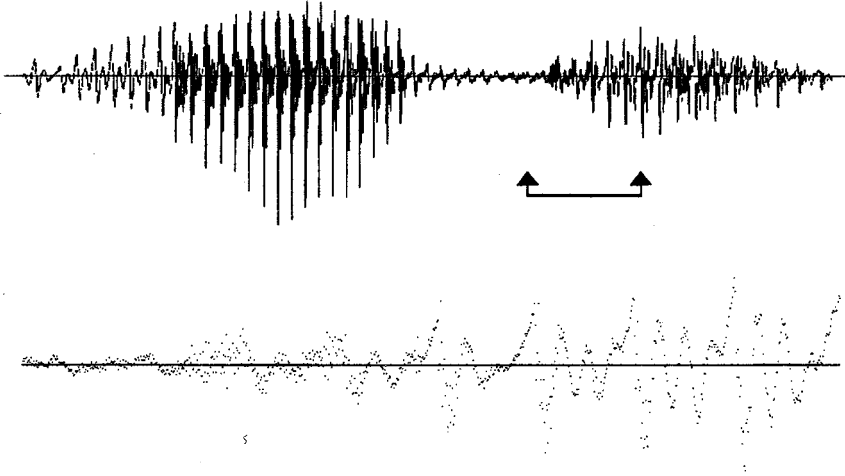


Figura 45. Realización de 'mayo'. El tramo corresponde con el alófono [j].

En la presentación espectral del fonema /j/ se distinguen dos partes; en primer lugar el ruido correspondiente a la fricción (instante 367 ms), y en segundo lugar una variación brusca del segundo formante similar a la secuencia 'io'.

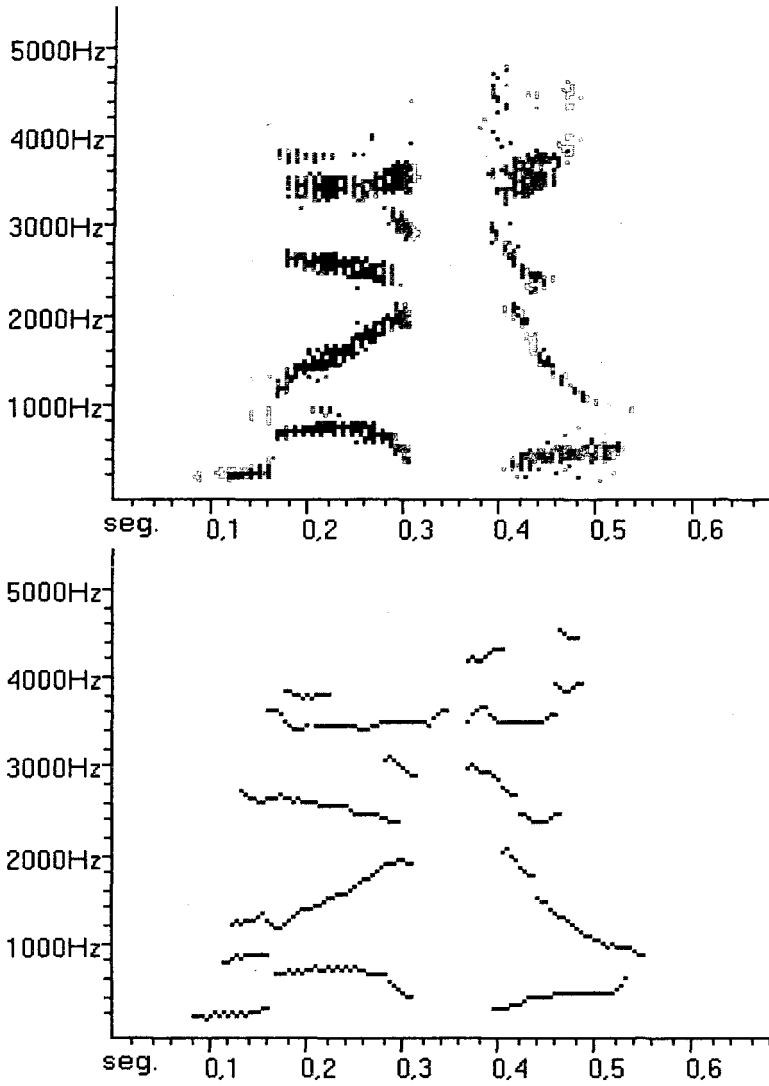


Figura 46. Espectro de la palabra 'mayo'.

## 1.4. Africadas

### 1.1.15. Fonema /tʃ/

Linguopalatal, africado y sordo. Ejemplos: *pecho*, *chico*, *coche*.

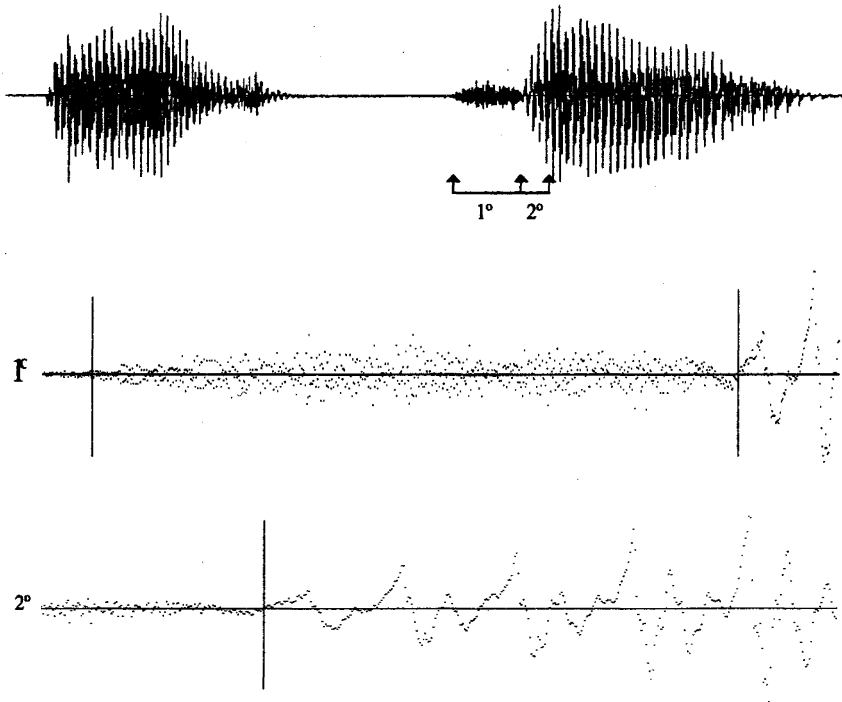


Figura 47. Realización de 'pecho'. El tramo 1° representa la zona fricativa y el tramo 2° la explosión del alófono [tʃ].

En la Figura 47 se tiene marcado el conjunto de muestras temporales que corresponde a la parte fricativa, antes de que se produzca la explosión; comienza en el instante 522 ms hasta el 587 ms.

En el espectro de la Figura 48 se aprecia perfectamente el ruido característico del fonema /tʃ/. Los formantes de la 'o' se detectan en el instante 590 ms.

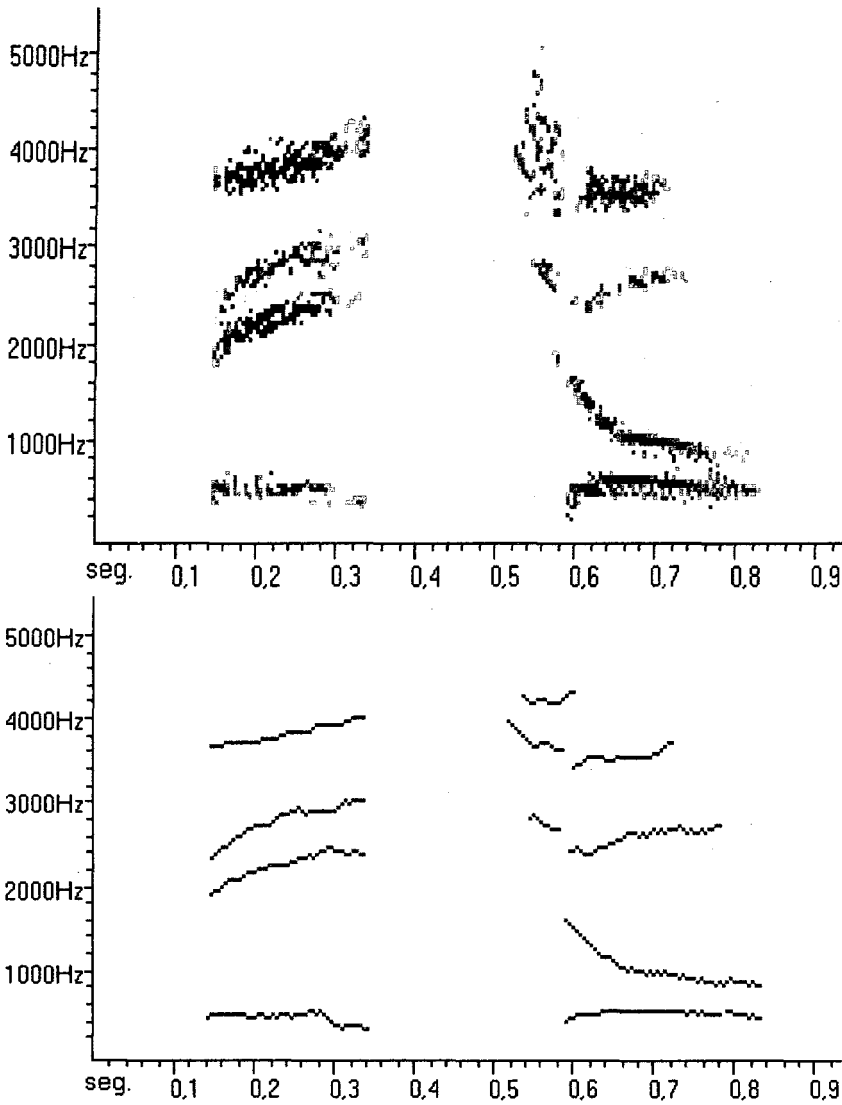


Figura 48. Espectro de la palabra 'pecho'.

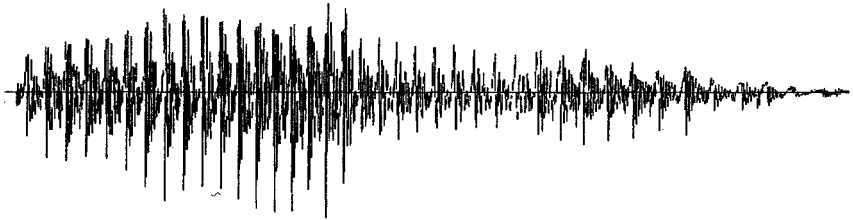
La característica más relevante que distingue a la /tʃ/ de la /s/ es su duración temporal; de hecho, si a una /s/ se le quita un tramo temporal suena como una /tʃ/.

## 1.5. Laterales

### 1.1.16. Fonema /l/

#### Alófono [l]

Linguoalveolar, lateral y sonoro. Se da cuando está en posición prenuclear o en postnuclear seguido de pausa, vocal o consonante distinta de [t, d, θ]. Ejemplos: *ala, mal, alférez*.



*Figura 49. Realización de 'ala'.*

Analizando la forma de onda de la señal temporal, se hace muy difícil distinguir el cambio entre la vocal 'a' y la consonante 'l'; tenemos un mejor indicio estudiando el cambio de volúmenes.

La mitad del alófono [l] se encuentra en 322 ms; aparece con cinco formantes con frecuencias 517 Hz, 1.723 Hz, 2.756 Hz, 3.747 Hz y 4.264 Hz, aunque el último posee muy poca energía. Una característica destacable es la continuidad de los formantes de la vocal con los del alófono [l].

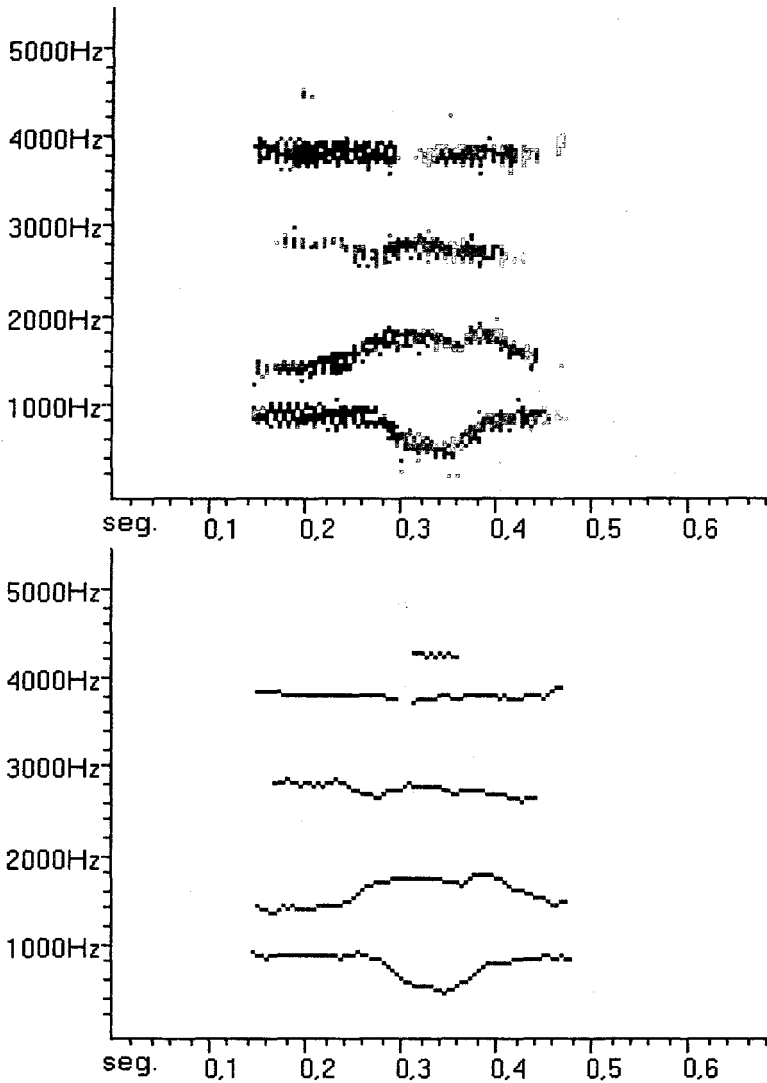


Figura 50. Espectro de la palabra 'ala'.

Alófono [l]

Linguointerdental, lateral y sonoro. En posición postnuclear seguido

del fonema [θ]. Ejemplos: *alza*, *dulce*.

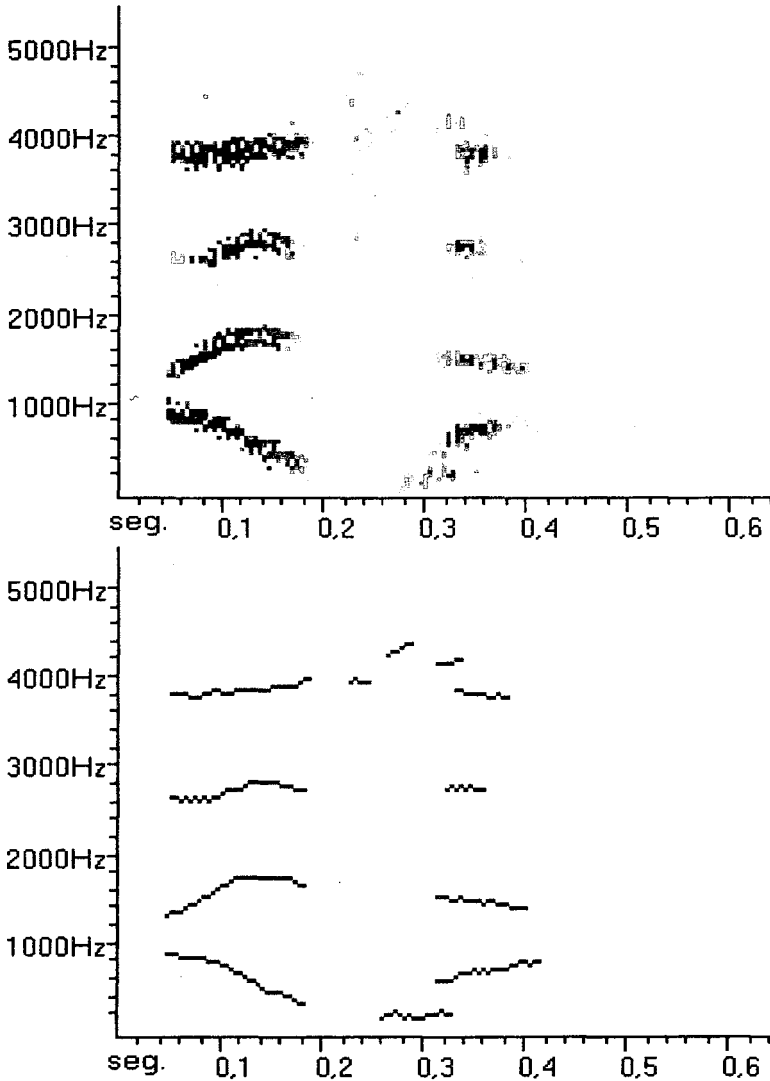


Figura 51. Espectro de la palabra 'alza'.

Al final del alófono (177 ms) los formantes tienen unas frecuencias de 301 Hz, 1.637 Hz, 2.713 Hz y 3.919 Hz. Destacan por el cambio de



energía que sufren en el último tramo y por la evolución de los formantes de la vocal anterior.

Alófono [l]

Linguodental, lateral y sonoro. En posición postnuclear seguido de los fonemas [t, d]. Ejemplos: *toldo, el toro*.

Al final del alófono (236ms) los formantes tienen una frecuencia de 301 Hz, 1.421 Hz, 2.584 Hz y 3.747 Hz. Los alófonos [l], [ l̥ ], y [l̥] tienen mucha similitud y es difícil distinguirlos.

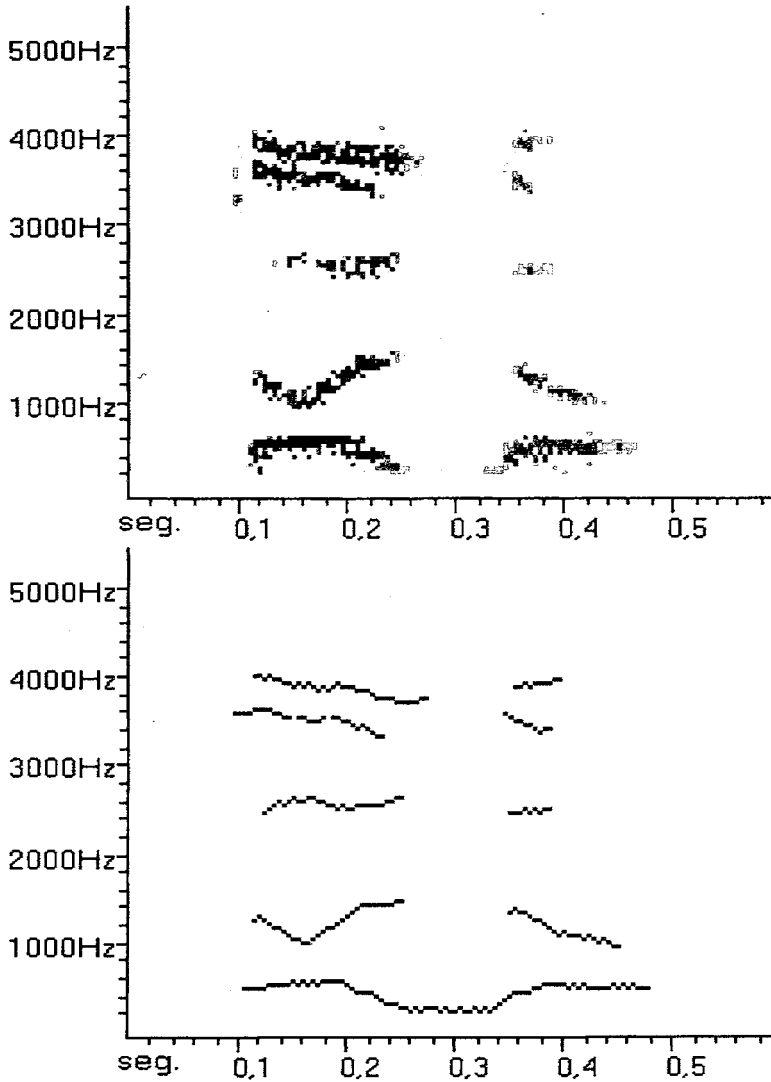


Figura 52. Espectro de la palabra 'toldo'.

Alófono [L]

Linguoprepalatal, lateral y sonoro. Se produce cuando precede a un

fonema palatal. Ejemplos: *colcha*, *colchón*.

Al final del alófono (218ms) los formantes tienen una frecuencia de 301 Hz, 1.723 Hz, 2.412 Hz y 3.575 Hz.

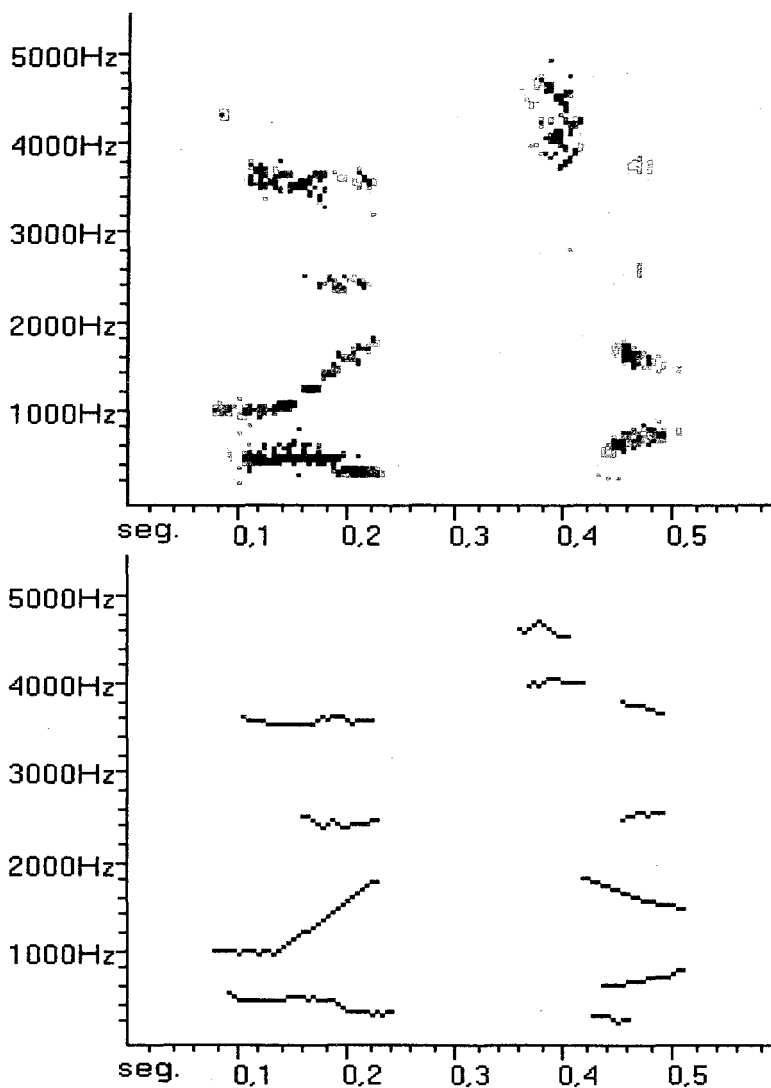


Figura 53. Espectro de la palabra 'colcha'.

## 1.1.17. Fonema /ʎ/

Linguomediopalatal, lateral y sonoro. Ejemplos: *llave, calle, cepillo*.

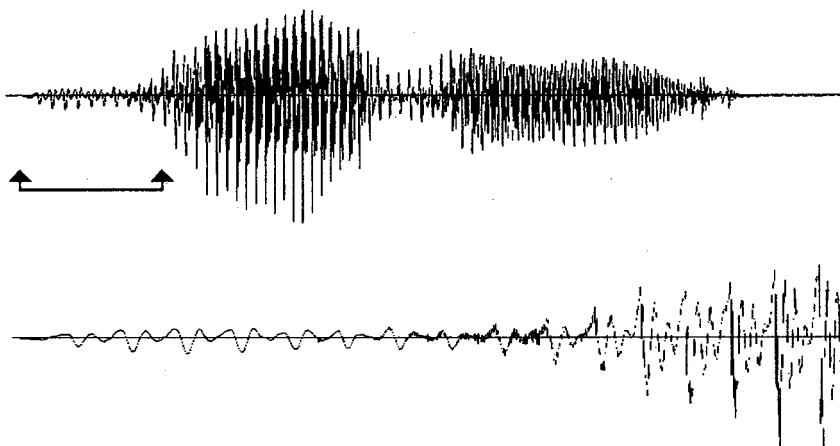


Figura 54. Realización de 'llave'. El tramo corresponde con el alófono [ʎ].

Al ser sonora, las cuerdas vocales comienzan en el instante 97 ms con una frecuencia de entre 172 Hz y 215 Hz. A partir del instante 149 ms aparece el ruido de altas frecuencias como consecuencia de la fricción. Hacia el momento 199ms se configuran los formantes de la vocal 'a'.

En el espectro de la Figura 55 se aprecian las etapas comentadas. Las cuerdas vocales se detectan en el instante 104 ms y el ruido de altas frecuencias en el instante 159 ms. Como en la captura se buscan formantes, se unen el ruido y el formante de la vocal posterior.

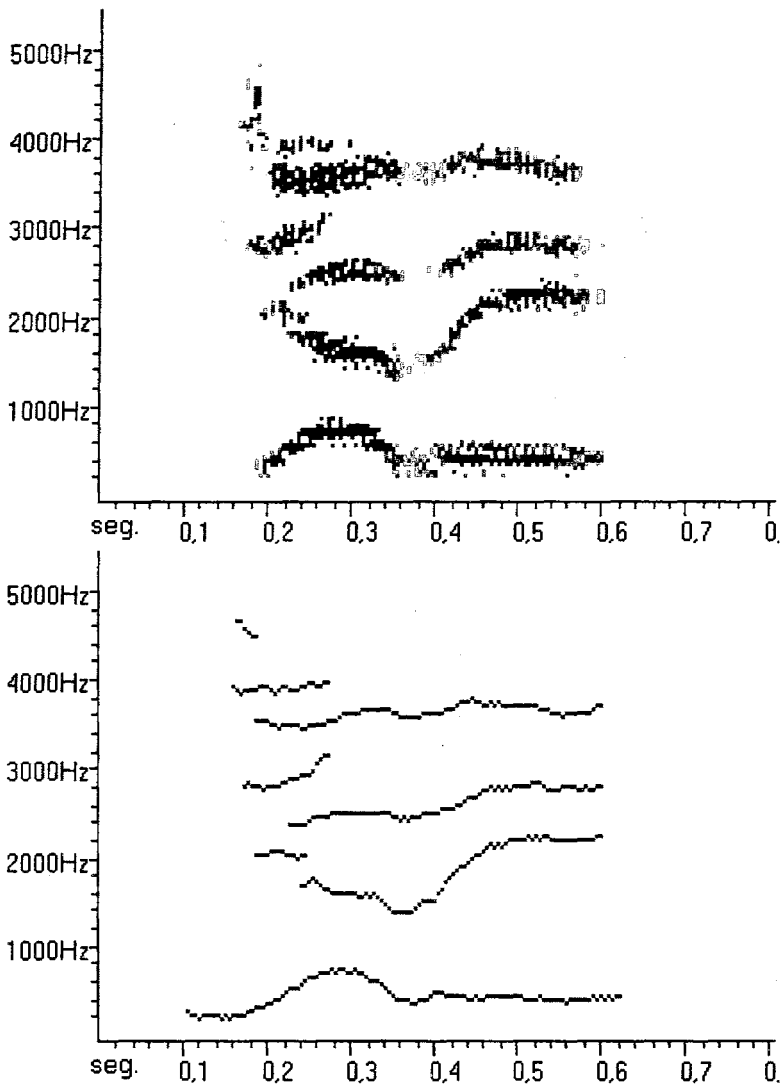


Figura 55. Espectro de la palabra 'llave'.

## 1.6. Vibrantes

### 1.1.18. Fonema /r/

Linguoalveolar, vibrante simple y sonoro. Se da cuando se encuentra en interior de palabra. Ejemplos: *para*, *norte*, *coro*.

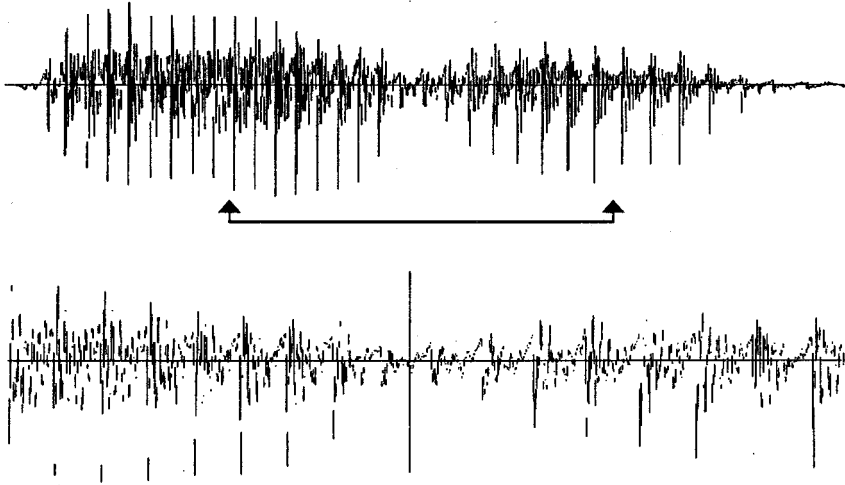


Figura 56. Realización de 'para'. El tramo corresponde con el alófono [r].

En la Figura 56 se representa la parte central de la 'r'. En este caso no se distingue fácilmente la oclusión, aunque se percibe disminución de la energía.

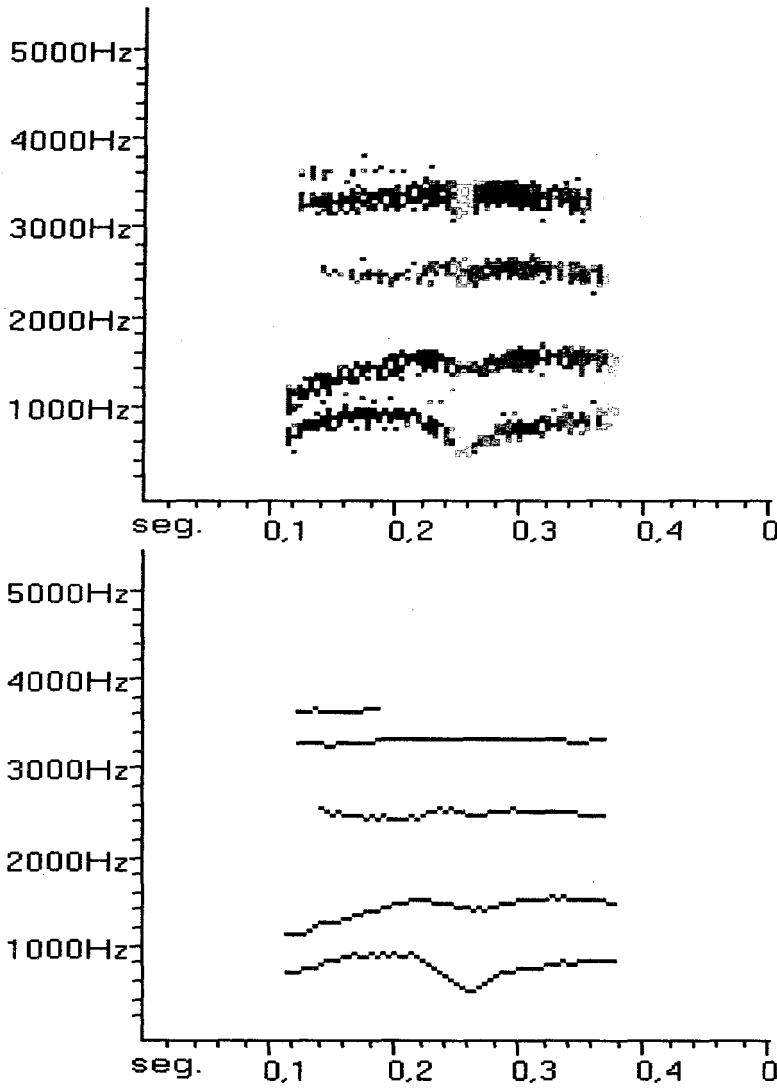


Figura 57. Espectro de la palabra 'para'.

En la Figura 57 vemos el espectro. Los cuatro formantes que corresponden con el fonema /r/ tienen las frecuencias 474 Hz, 1.378 Hz, 2.455 Hz y 3.316 Hz, en el instante 263 ms.

## 1.1.19. Fonema /r/

Linguoalveolar, vibrante múltiple y sonoro. Ejemplos: *perro*, *remo*, *enroscar*.

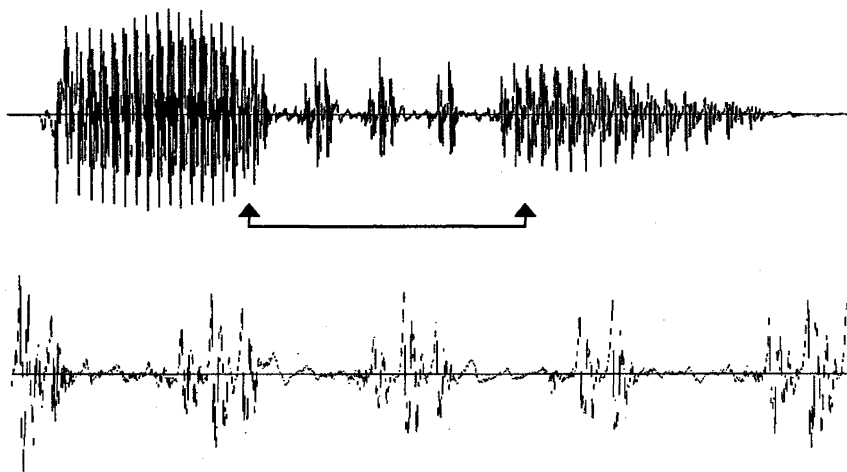


Figura 58. Realización de 'perro'. El tramo corresponde con el alófono [r].

Al contrario que para el fonema /r/, tanto en el tramo de la Figura 58 como en el espectro de la Figura 59 se aprecian claramente cuatro oclusiones.



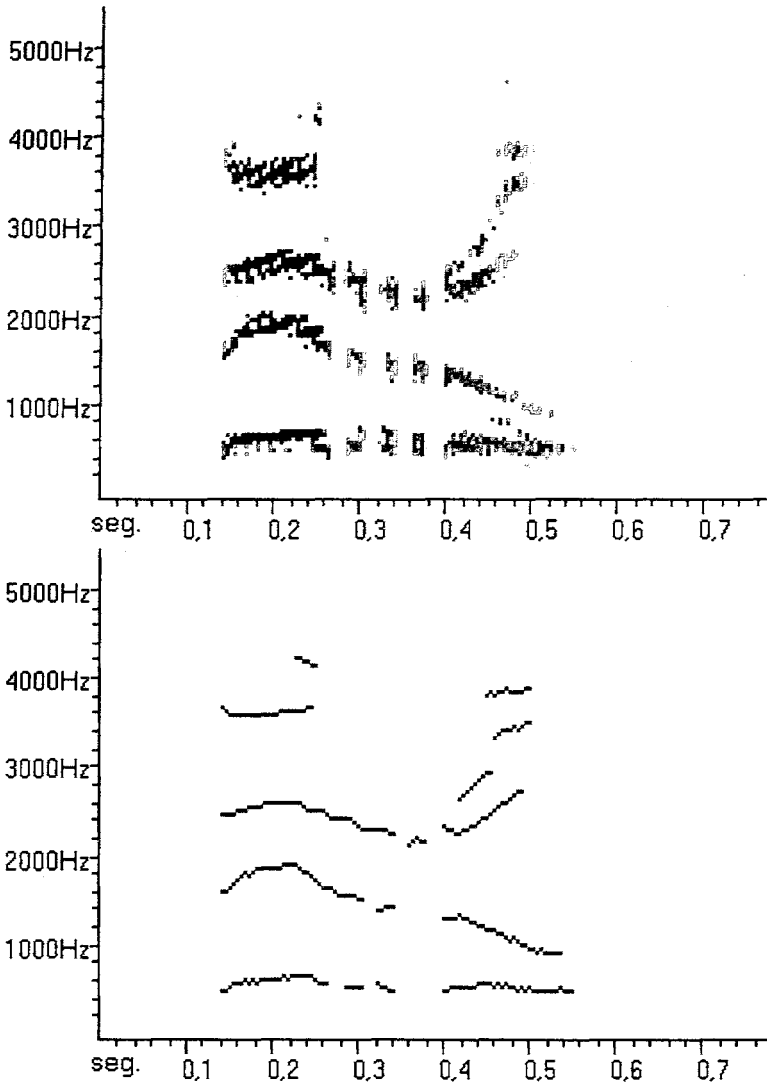


Figura 59. Espectro de la palabra 'perro'.

1.7. VOCALES

1.1.20. Fonema /a/

Central, abierta, sonora.

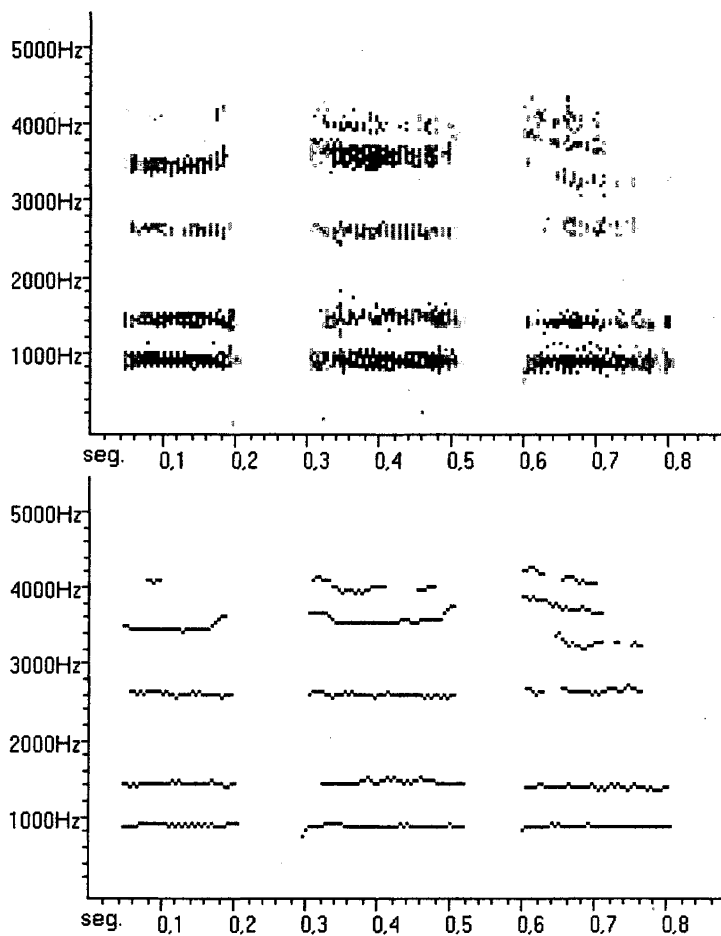


Figura 60. Espectros de la vocal 'a'.

La Figura 60 presenta el espectro del fonema /a/ grabado tres veces por el mismo hablante. En él destaca la regularidad de los tres primeros

formantes; el primero en una frecuencia de 904 Hz, 861 Hz y 861 Hz; el segundo de 1.421 Hz, 1.464 Hz y 1.378 Hz; y el tercero de 2.627 Hz, 2.584 Hz y 2.627 Hz. Los siguientes dependiendo de la grabación, salen con valores diferentes.

## 1.1.21. Fonema /e/

Anterior, media, sonora.

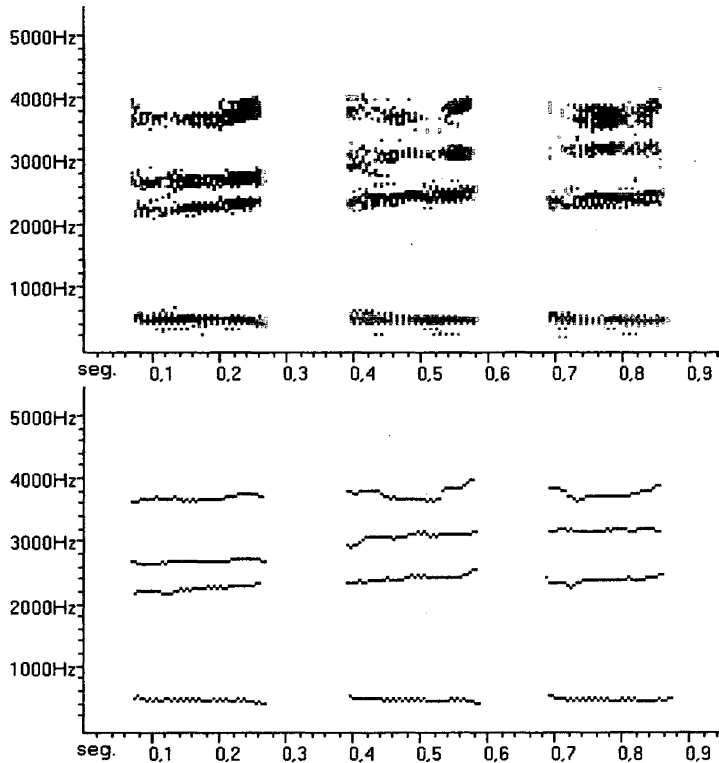


Figura 61. Espectros de la vocal 'e'

La Figura 61 presenta el espectro de la vocal 'e', grabado en tres ocasiones. En este caso, son los formantes primero, segundo y cuarto los que mantienen con regularidad la frecuencia; el tercer formante oscila entre el segundo y el cuarto formante. El primero tiene una frecuencia de 431 Hz, el segundo de 2.239 Hz, 2.412 Hz y 2.369 Hz, y el cuarto de 3.661 Hz, 3.661 Hz y 3.704 Hz.

## 1.1.22. Fonema /i/

Anterior, cerrada, sonora.

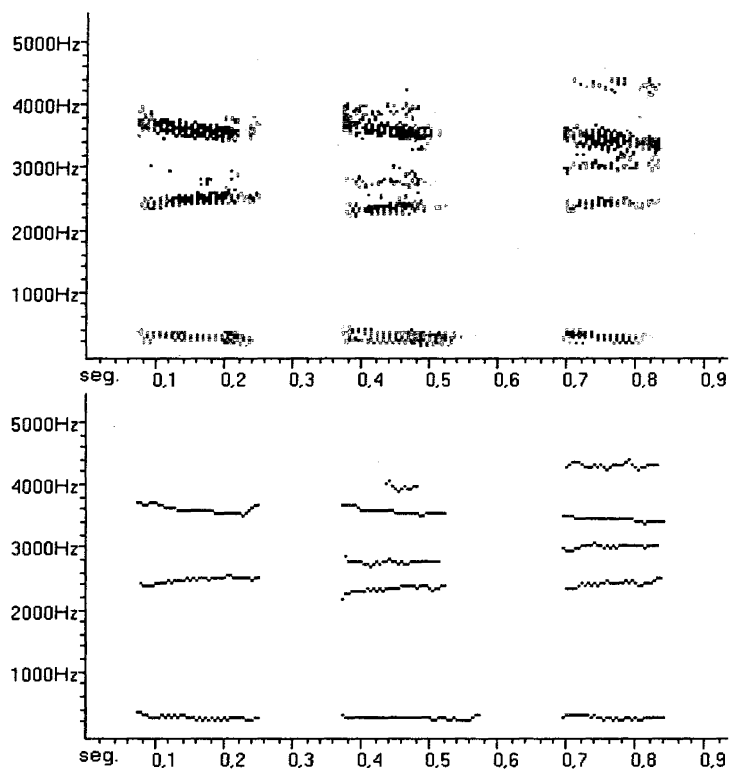


Figura 62. Espectros de la vocal 'i'.

La Figura 60 presenta el espectro de la grabación de la 'i'. Se produce una situación muy similar a la de la 'e': los tres formantes principales se encuentran en las frecuencias de 258 Hz, 2.412 Hz y 3.531 Hz. Pero surgen unos formantes con una energía más débil que dependiendo de la grabación, aparecen en frecuencias distintas.

## 1.1.23. Fonema /o/

Posterior, media, sonora.

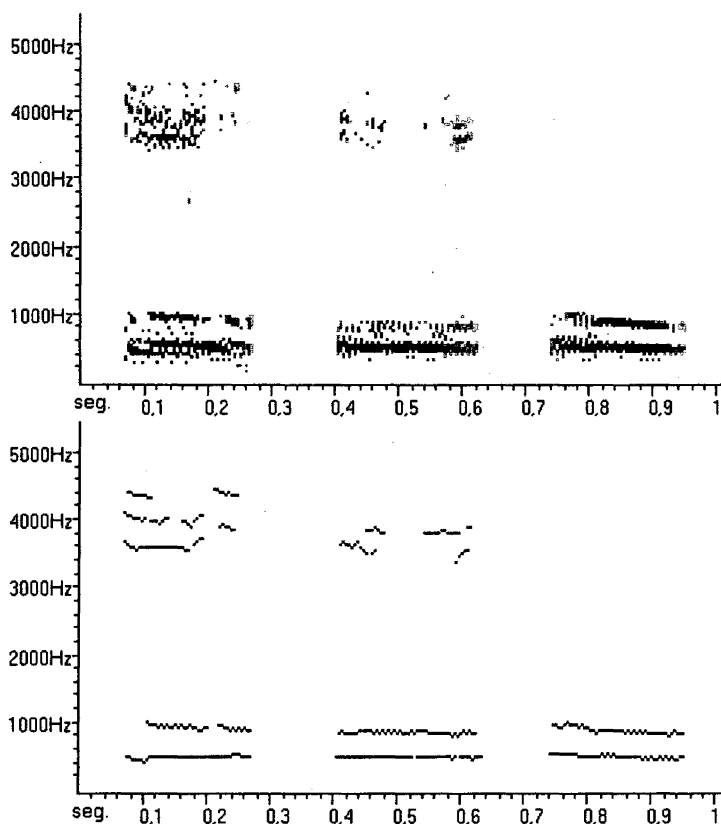


Figura 63. Espectros de la vocal 'o'.

La Figura 63 representa el espectro de la vocal 'o'. Los dos primeros formantes son muy estables y se dan en las frecuencias de 474 Hz y 861 Hz. Los formantes de frecuencias altas, según la grabación, aparecen alrededor de los 3.900 Hz, o no aparecen ya que disponen de muy poca energía.

## 1.1.24. Fonema /u/

Posterior, cerrada, sonora.

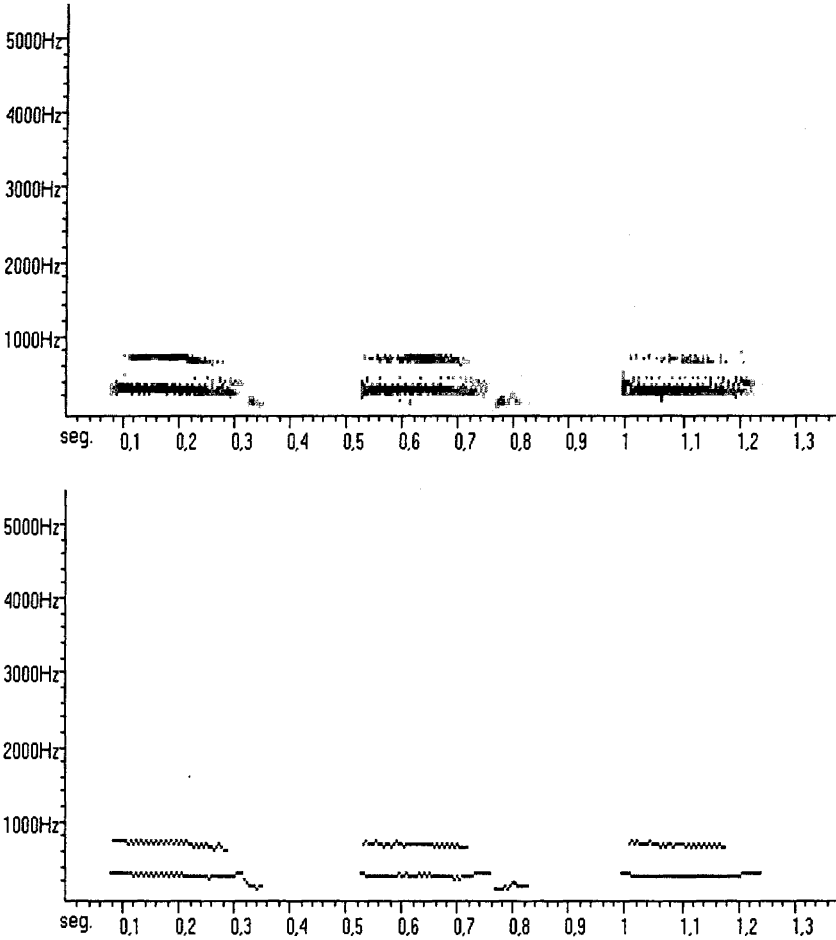


Figura 64. Espectros de la vocal 'u'.

La Figura 64 contiene el espectro de la vocal 'u'. En este caso sólo aparecen dos formantes bastante estables entre las distintas grabaciones. El primero aparece en la frecuencia de 258 Hz y el segundo en la de 689 Hz.

### **3. FONEMAS VOCÁLICOS REALIZADOS POR DIFERENTES LOCUTORES**

En este apartado se pretende probar el método propuesto en grabaciones realizadas por diferentes locutores.

Todas la grabaciones que se presentan a continuación corresponden a las cinco vocales pronunciadas de forma aislada y en orden alfabético. Se han establecido tres grupos de edades y dentro de cada una se han tomado muestras de dos hombres y dos mujeres.



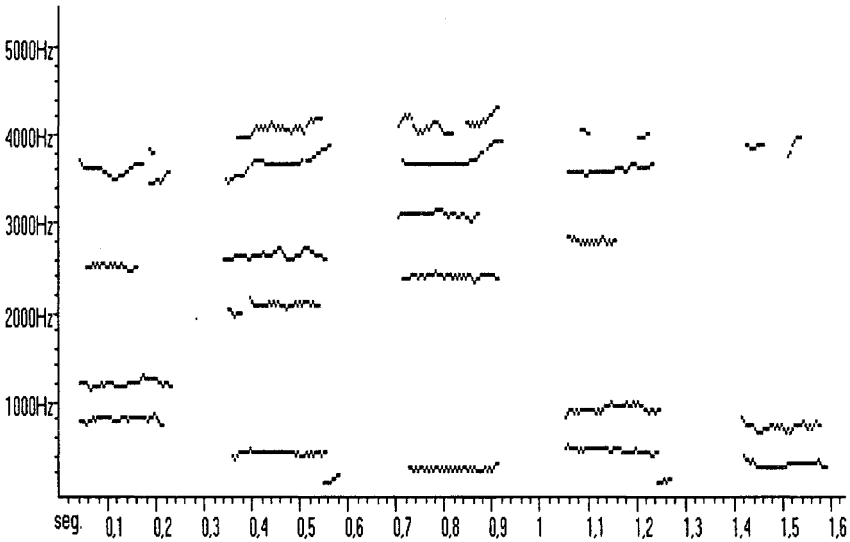
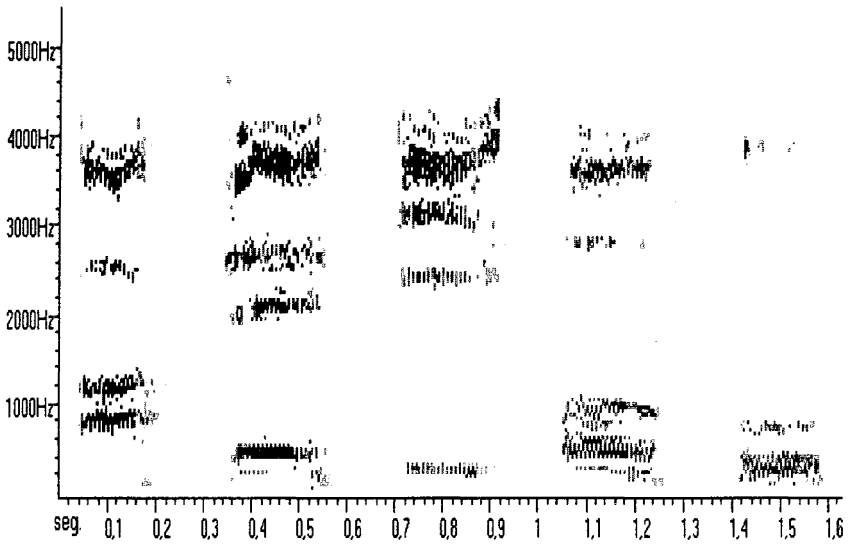


Figura 65. Espectro de la secuencia 'a e i o u' realizado por un hombre joven.

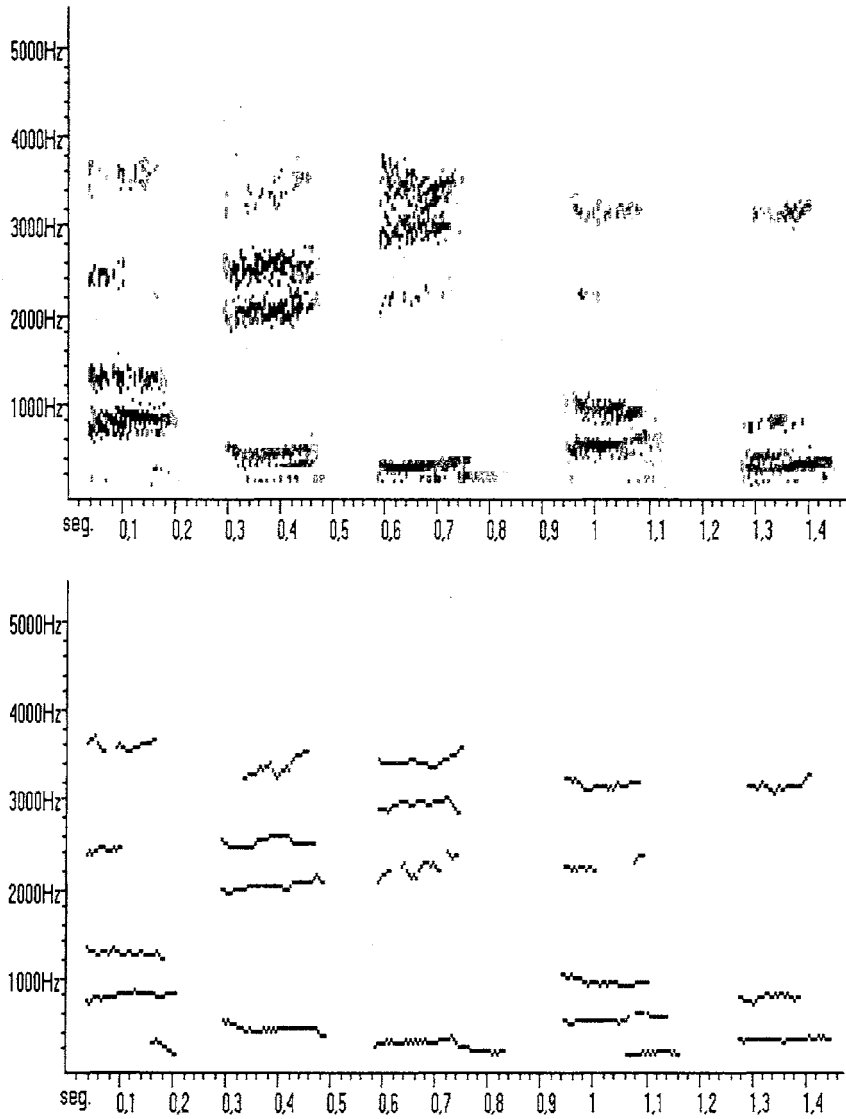


Figura 66. Espectro de la secuencia 'a e i o u' realizado por un hombre joven.

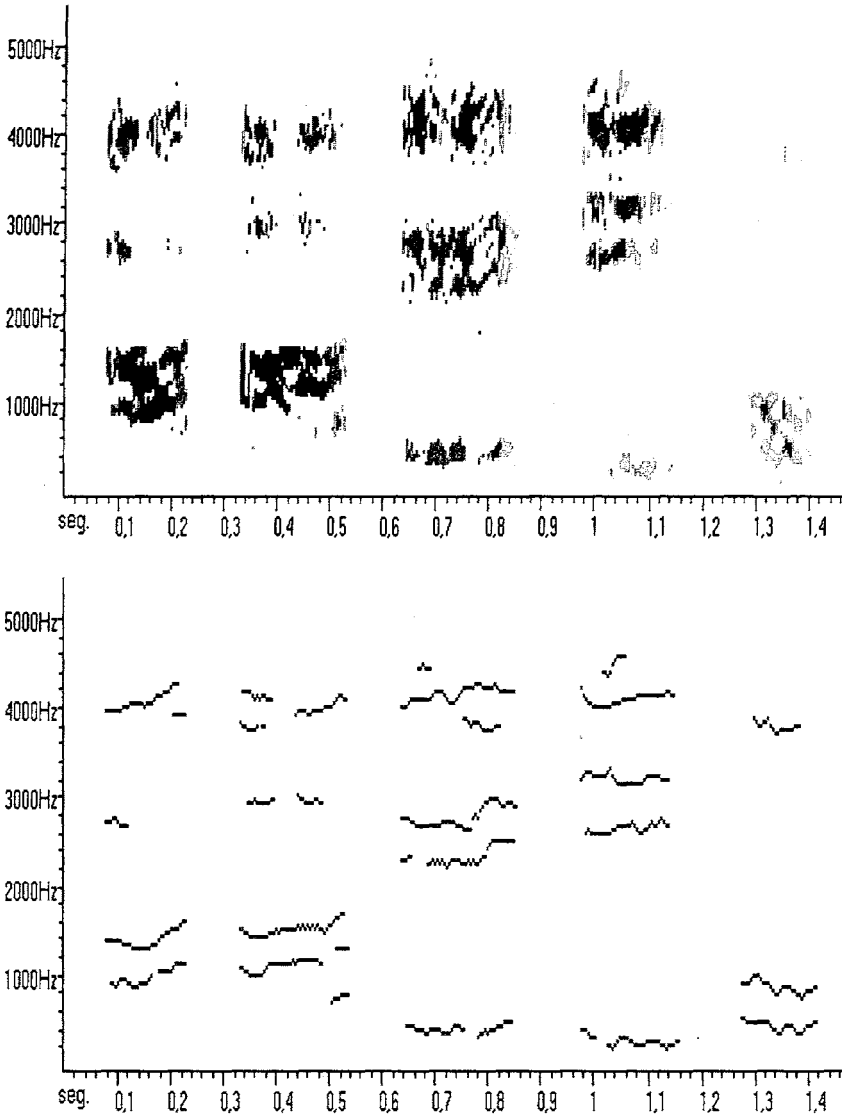


Figura 67. Espectro de la secuencia 'a e i o u' realizado por una mujer joven.

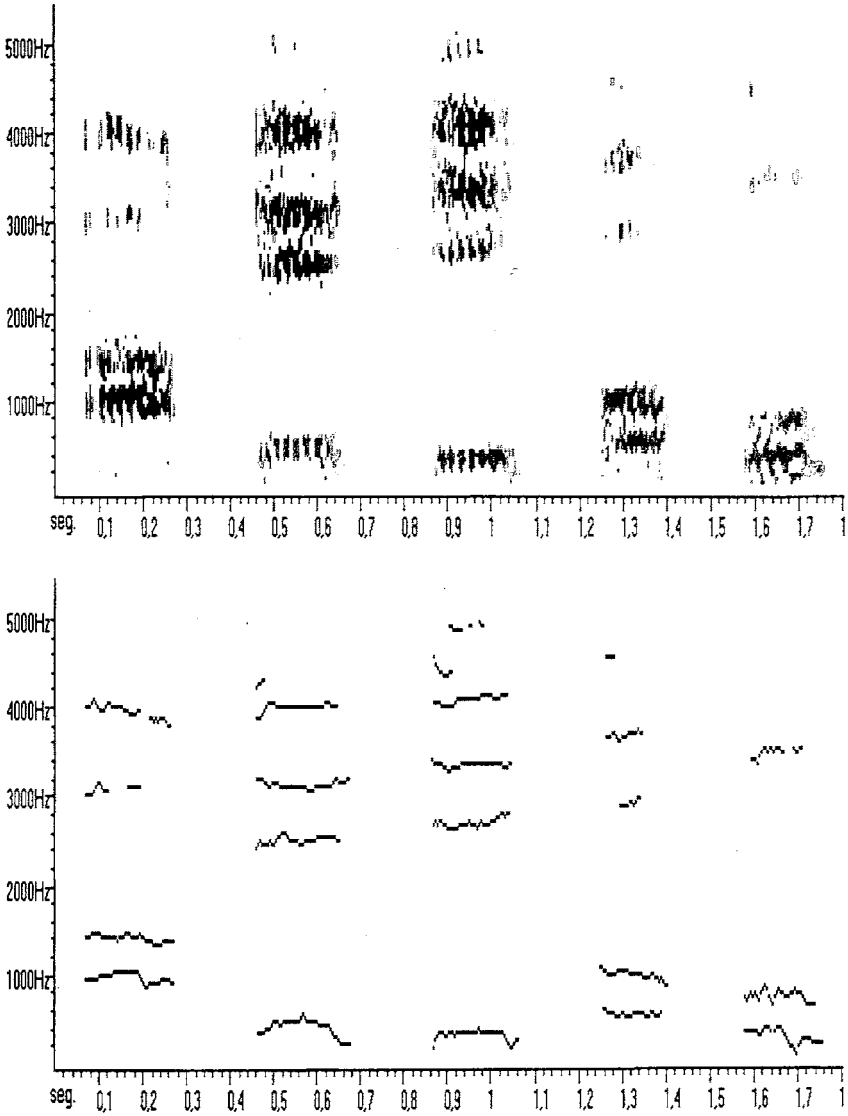


Figura 68. Espectro de la secuencia 'a e i o u' realizado por una mujer joven.

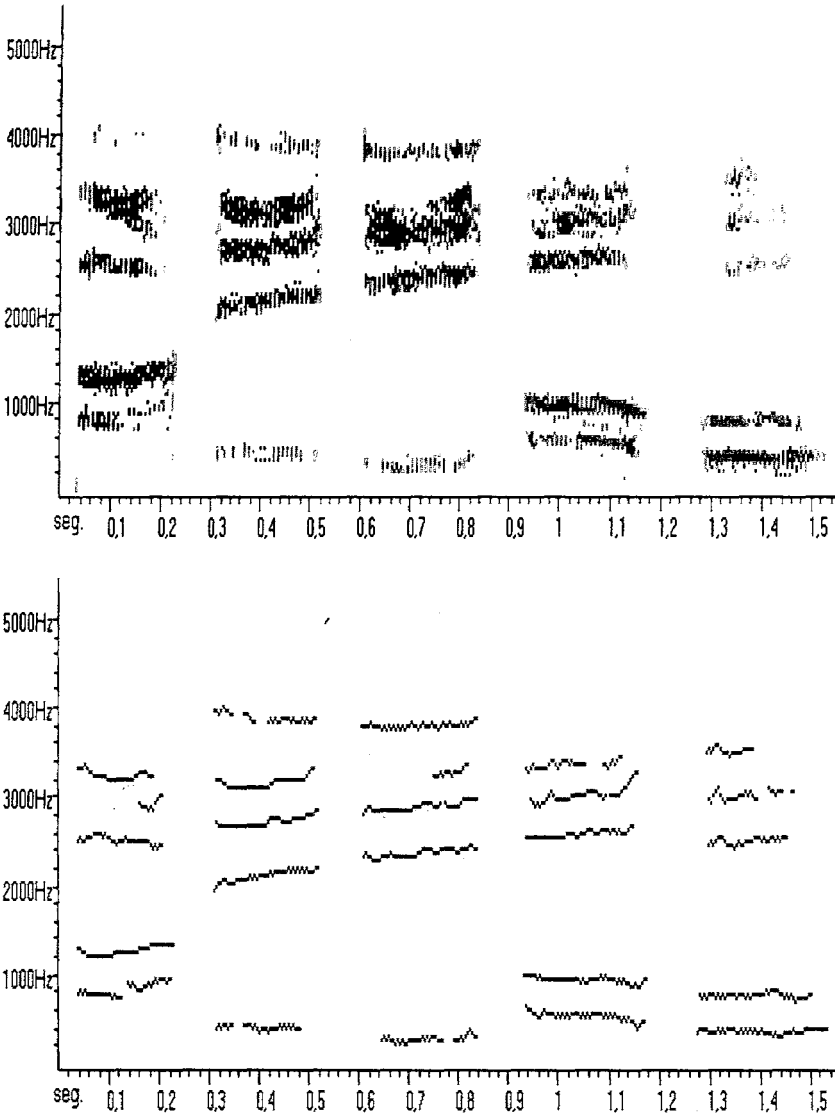


Figura 69. Espectro de la secuencia 'a e i o u' realizado por un hombre de media edad.

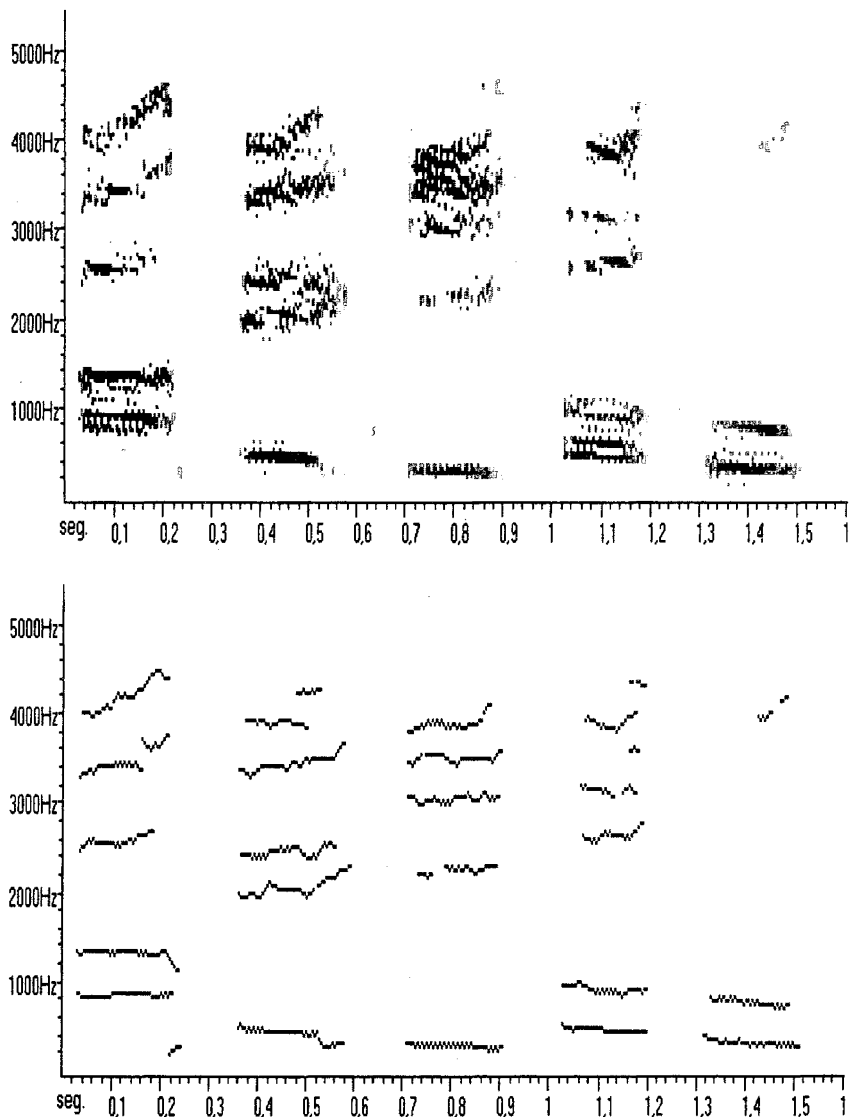


Figura 70. Espectro de la secuencia 'a e i o u' realizado por un hombre de media edad.

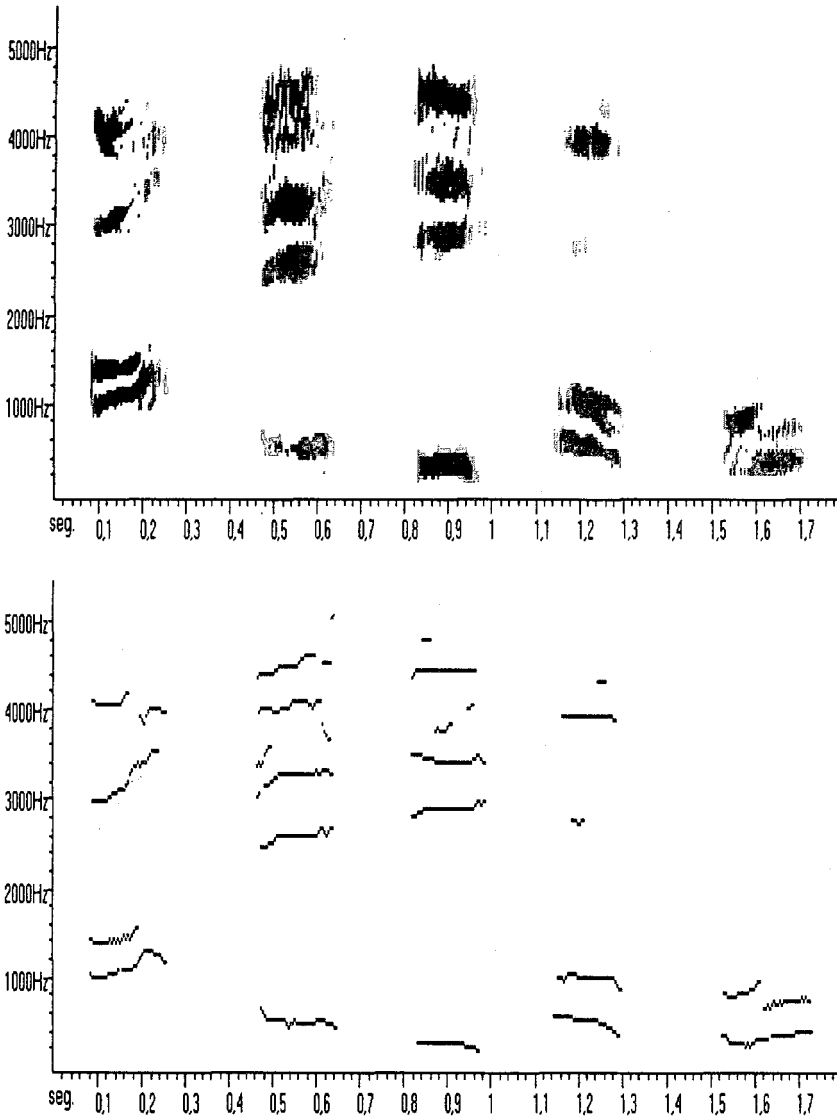


Figura 71. Espectro de la secuencia 'a e i o' realizado por una mujer de media edad.

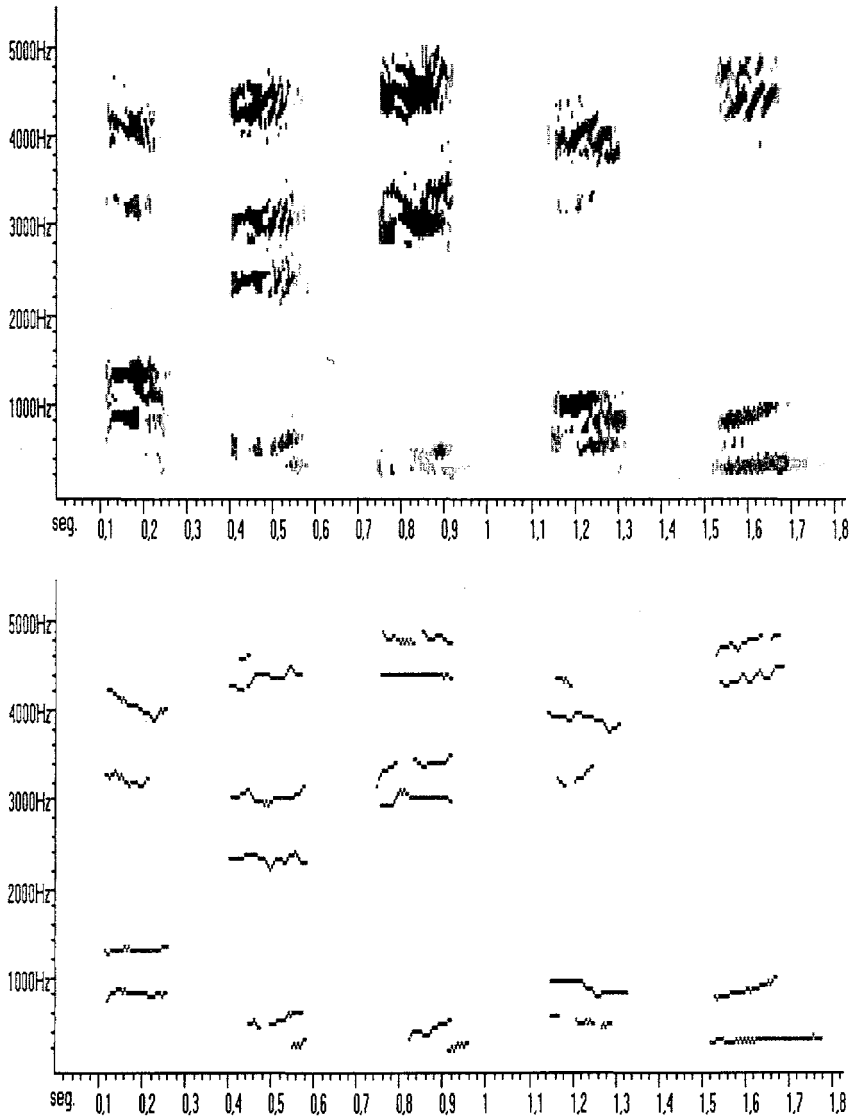


Figura 72. Espectro de la secuencia 'a e i o u' realizado por una mujer de media edad.



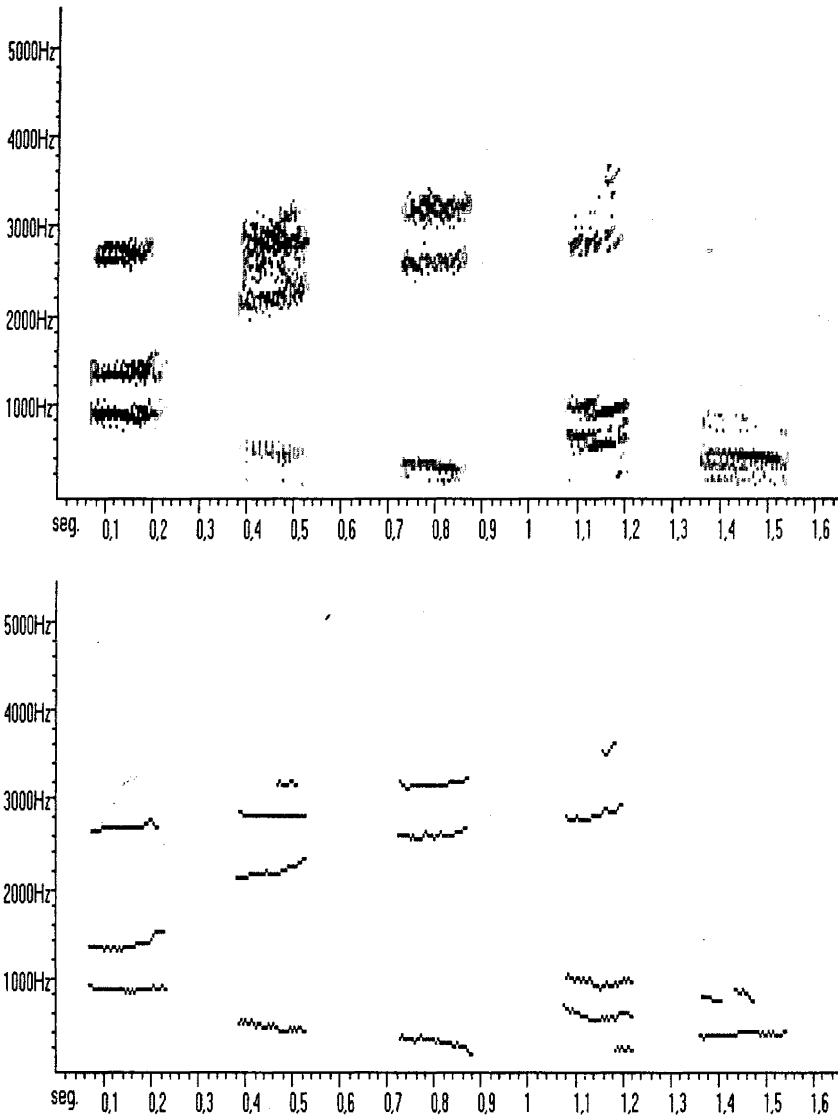


Figura 73. Espectro de la secuencia 'a e i o' realizado por un hombre mayor.

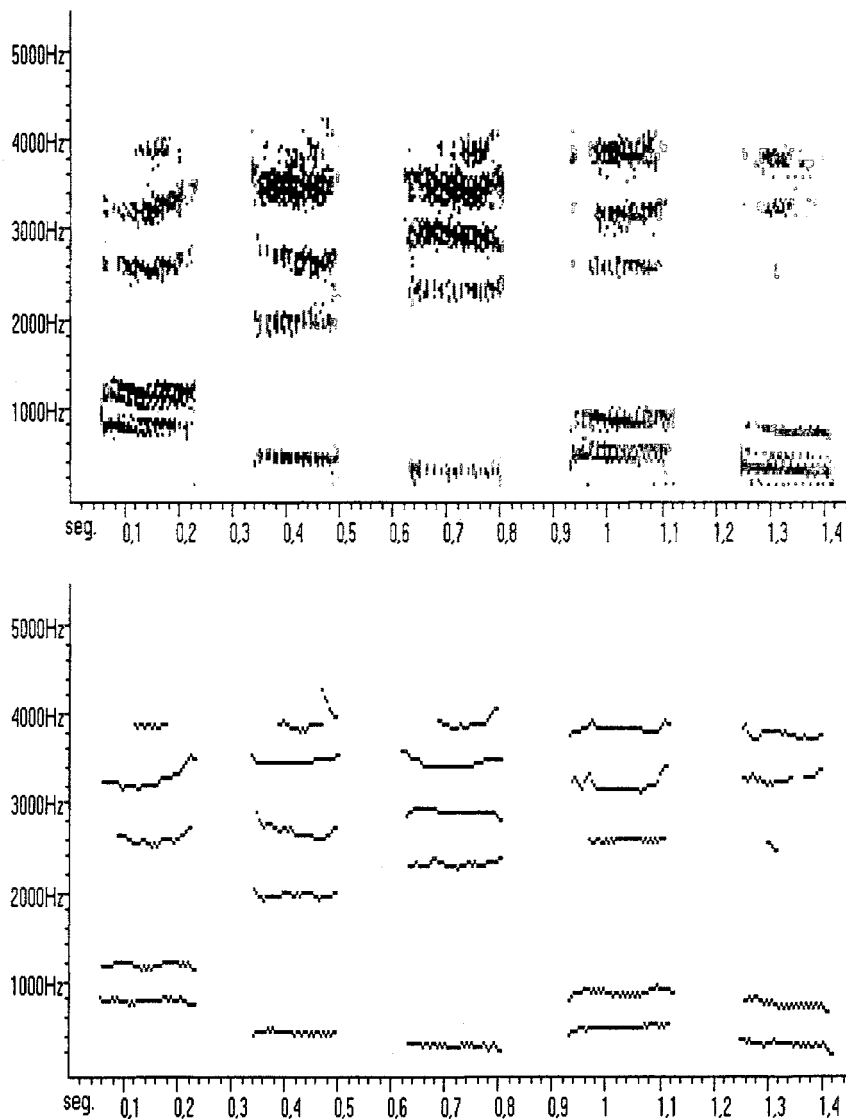
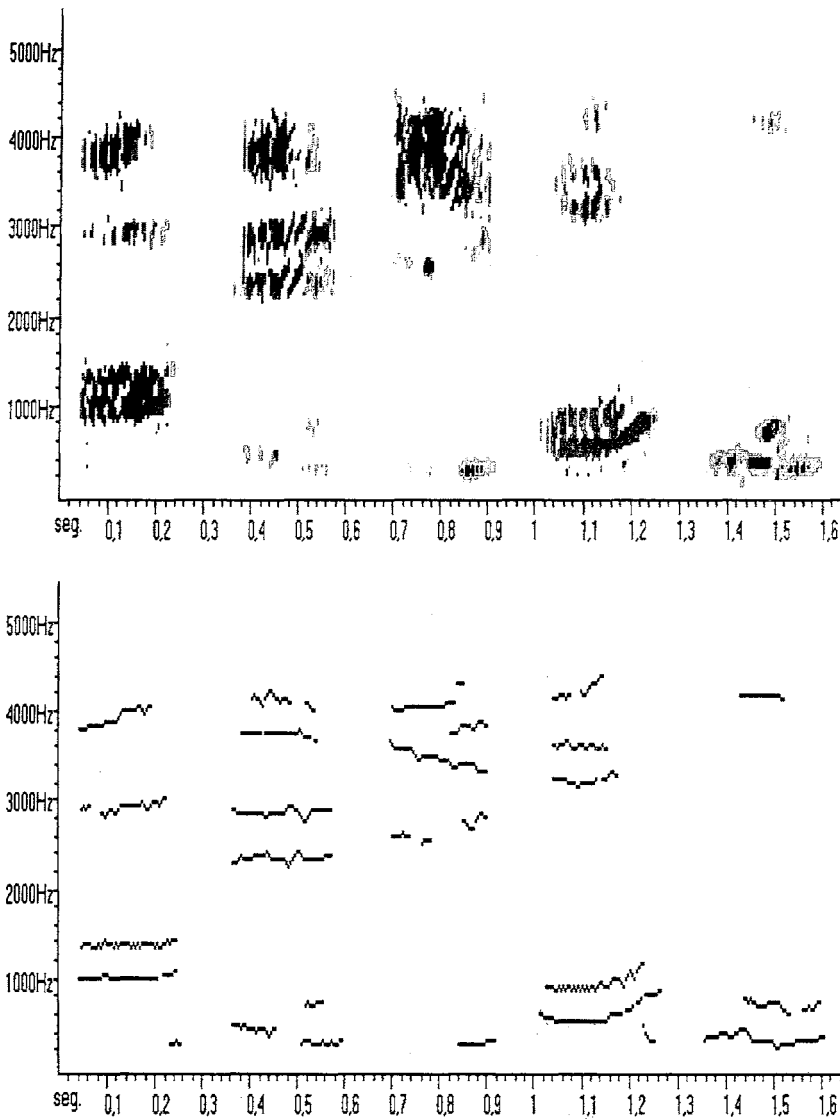


Figura 74. Espectro de la secuencia 'a e i o u' realizado por un hombre mayor.



*Figura 75. Espectro de la secuencia 'a e i o u' realizado por una mujer mayor.*

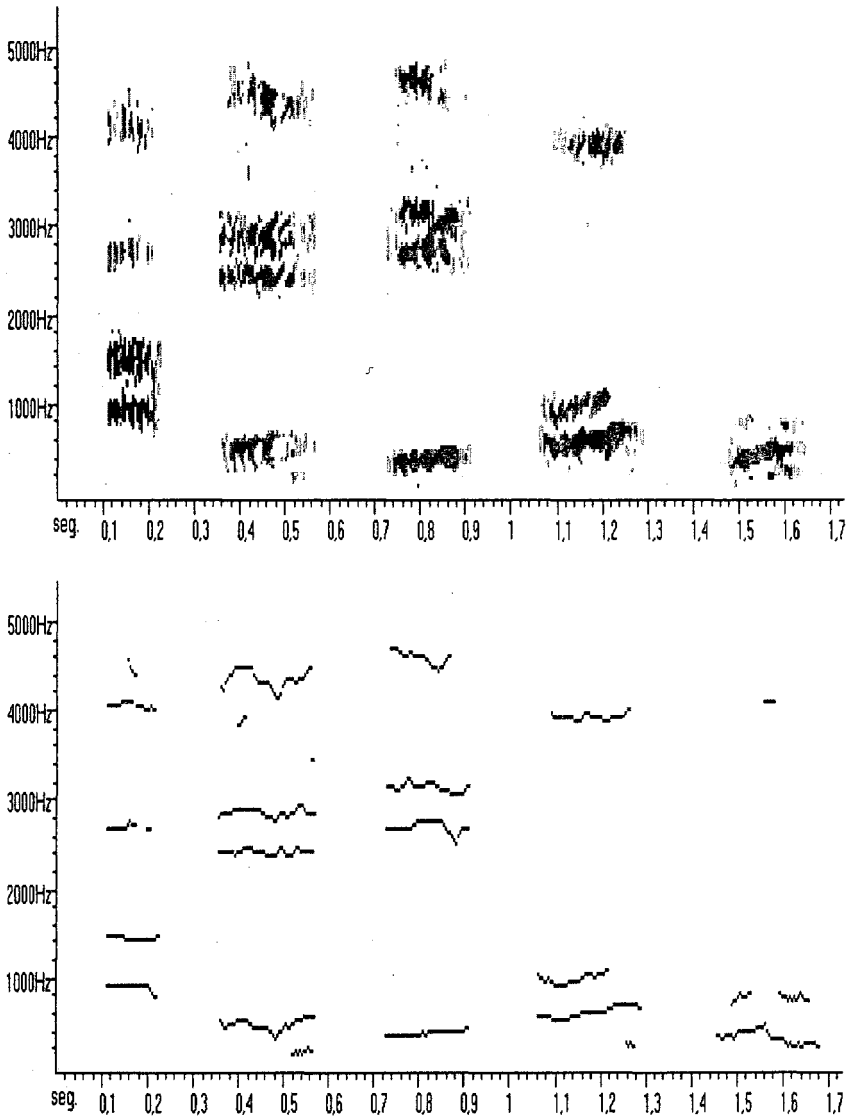


Figura 76. Espectro de la secuencia 'a e i o u' realizado por una mujer mayor.

#### 4. CONCLUSIONES

Es patente la importancia que tiene el uso de los espectros para el estudio o investigación en la fonética y fonología. Por ello, proporcionar métodos nuevos que realcen los espectros aporta aspectos muy interesantes para su uso como herramientas de trabajo.

Hemos presentado un conjunto de espectros que abarcan la mayoría de los alófonos que existen en la lengua castellana. Resulta una base de conocimiento completa para su utilización. Su característica más relevante es que se han utilizado métodos originales para la eliminación de ruido, quedando perfectamente aislados los formantes y demás características acústicas que caracterizan a los diferentes alófonos.

A cada espectro se le acompaña una extracción automática de formantes con mediciones de instantes temporales y de frecuencias para su mejor interpretación.

#### 5. REFERENCIAS

- [Bri88] E. O. Brigham, *The Fast Fourier Transform and its Applications*, Prentice-Hall, Gran Bretaña, 1988.
- [Coh89] L. Cohen, "Time-Frequency Distributions - A Review", *Proc. IEEE*, vol. 77(7), julio 1989, pp. 941-981.
- [Cha95] V. R. Chari & C. Y. Espy-Wilson, "Adaptative Enhancement of Fourier Spectra", *IEEE Trans. Speech and Audio Processing*, vol. 3(1), enero 1995, pp. 35-39.
- [Int95] F. D'Introno, E. Teso y R. Weston, *Fonética y Fonología actual del español*, Cátedra, Madrid, 1995.
- [Koe46] W. Koenig, H. K. Dunn & L. Y. Lacy, "The Sound Spectrograph", *JASA*, vol. 18(1), julio 1946, pp. 19-49.

- [Mar84] E. Martínez Celdrán, *Fonética*, Teide, Barcelona, 1984.
- [Mar87] J. Martí Roca, “FFT como herramienta de análisis en fonética”, *Estudios de fonética experimental*, mayo 1987.
- [Mar90C] E. Martínez Celdrán, “Una utilidad en fonética: la carta de formantes por ordenador”, *Estudios de fonética experimental*, vol. 4, 1990, pp.179-193.
- [Qui93] A. Quilis, *Tratado de fonología y fonética españolas*, Gredos, Madrid, 1993.
- [Sch95] P. Schmid & E. Barnard, “Robust, N-Best Formant Tracking”, *Proc. EUROSPEECH'95*, septiembre 1995, pp. 737-740.
- [Tho94] T. G. Thomas, P. C. Pandey & S. D. Agashe, “A PC-Based Multi-resolution Spectrograph”, *Inst. Electronics & Telecom. Engrs.*, vol. 40(2 & 3), marzo-junio 1994, pp.105-108.