

Reconocimiento Automático del Habla

2023-2024

Producción de la voz



DEPARTAMENT DE SISTEMES
INFORMÀTICS I COMPUTACIÓ



MIARFID-RAH mcastro@dsic.upv.es

Producción de la voz

El proceso de comunicación

La voz humana

Fonética: articulatoria y acústica

Fonética articulatoria

Aparato fonador humano

Mecanismo de producción del Habla

Relación entre sonidos y configuraciones del aparato fonador

Fonética acústica

Vocales

Consonantes explosivas orales

Consonantes explosivas nasales

Consonantes fricativas

Africadas

Líquidas

Otros fenómenos fonológicos

Estructura lingüística del habla

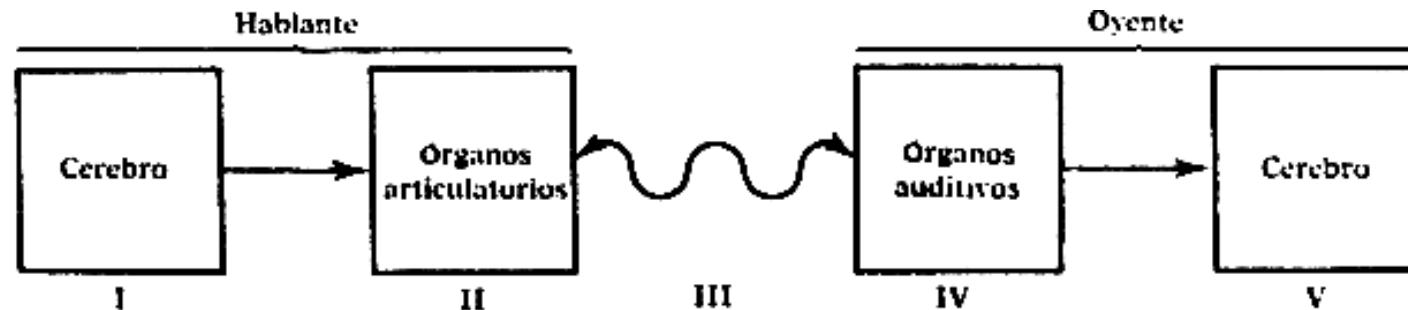
El proceso de comunicación

Sistemas de comunicación

Un sistema de comunicación consta de:

- Un **emisor** o fuente de información : selecciona *signos* de un *alfabeto* para formar un *mensaje*;
- Un **transmisor**: *codifica* el mensaje siguiendo un conjunto de reglas transformándolo de la representación original a otra apta para su transmisión.
- Un **canal**: medio material usado para la transmisión de la información . Todo canal puede venir afectado por *ruido*, esto es, puede introducir defectos que ocasionan una pérdida de información .
- Un **receptor**: *descodifica* el mensaje para devolverlo a su representación original;
- Un **destino**, que recibe el mensaje.

En la *comunicación lingüística* tenemos:

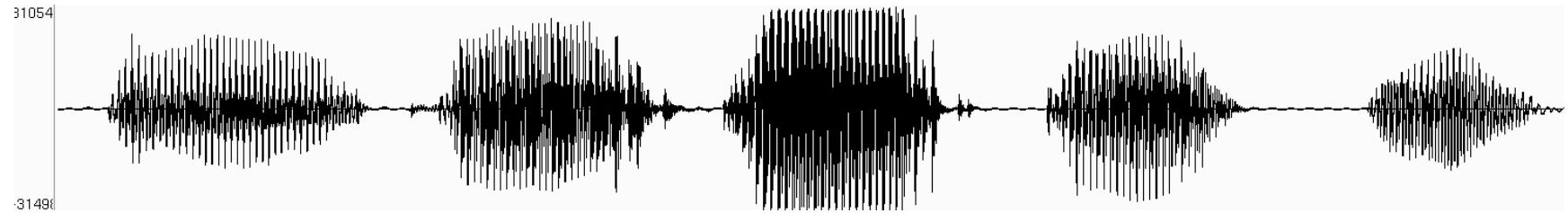


Etapas en el proceso de comunicación hablada

- **Emisor:** cerebro del hablante.
- **Transmisor:** órganos articulatorios del hablante que generan ondas sonoras.
- **Canal:** aire (y, posiblemente, otras etapas).
- **Receptor:** aparato auditivo del receptor.
- **Destino:** cerebro del oyente.

La voz humana

¿Cómo se producen las ondas sonoras del habla?



Señal acústica: amplitud a lo largo del tiempo (pronunciación de “ieaou”).

La señal acústica es producida por el aparato fonador y se transmite mediante ondas de presión propagadas por el aire.

Fonética

La fonética es la ciencia que estudia los sonidos del habla humana.

Los sonidos producidos en el habla pueden transcribirse usando un alfabeto fonético.

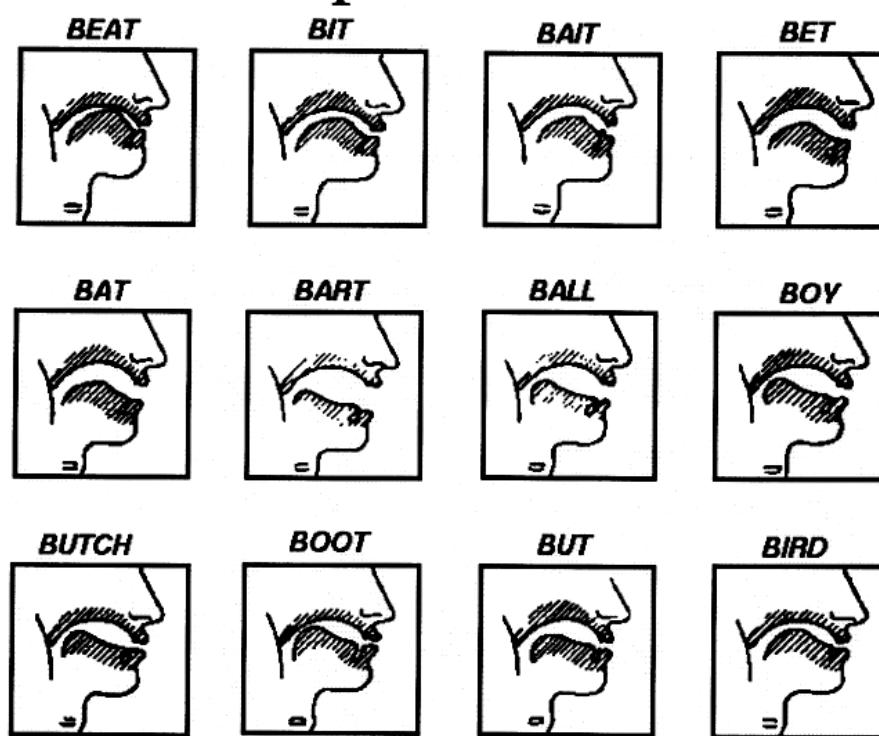
Vocales	abocinadas			anterior		central		posterior	
cerradas	(y	ɥ	u)	i	y	i	ɥ	ɯ	u
semicerradas	(ø	o)	e	ø				ɤ	ɔ
cerradas						θ			
semiabiertas	(œ	ɔ)			ɛ	œ			
abiertas		(ɒ)				æ		a	
						a		ɑ	ɒ

	ilabial	labio-dental	dental, retrofleja alveolar o post-alveolar	palato-alveolar	palatal	velar uvular	labio-palatal	labio-velar	faringea glotal
nasales	m	ŋ	n	ɳ	j	ŋ			
occlusivas	p b	t d	t̪ d̪	c ɟ	k g	q G	kpgb		?
fricativas: centrales	ɸ β	f v	θ ð s z	ʂ ʐ	ç ɿ	x ɣ	χ ɻ	w	h ɦ
aproximantes					j	ɥ	ɥ		
con aire									
laterales:			ɬ l						
expirado					ɺ				
fricativas									
aproximantes									
vibrantes:			r				R		
múltiples			r̩				R̩		
simples			ɺ̩						
sin aire									
eyectivas	p'	b	t'	d		k'			
inyectivas						g			
clics:									
centrales	ʘ		ʇ	ʖ					
laterales									

¿Cómo se pronuncian los fonemas? ¿Cómo se perciben los fonemas? Hay dos aproximaciones a la ciencia fonética: fonética articulatoria y fonética acústica.

Fonética articulatoria

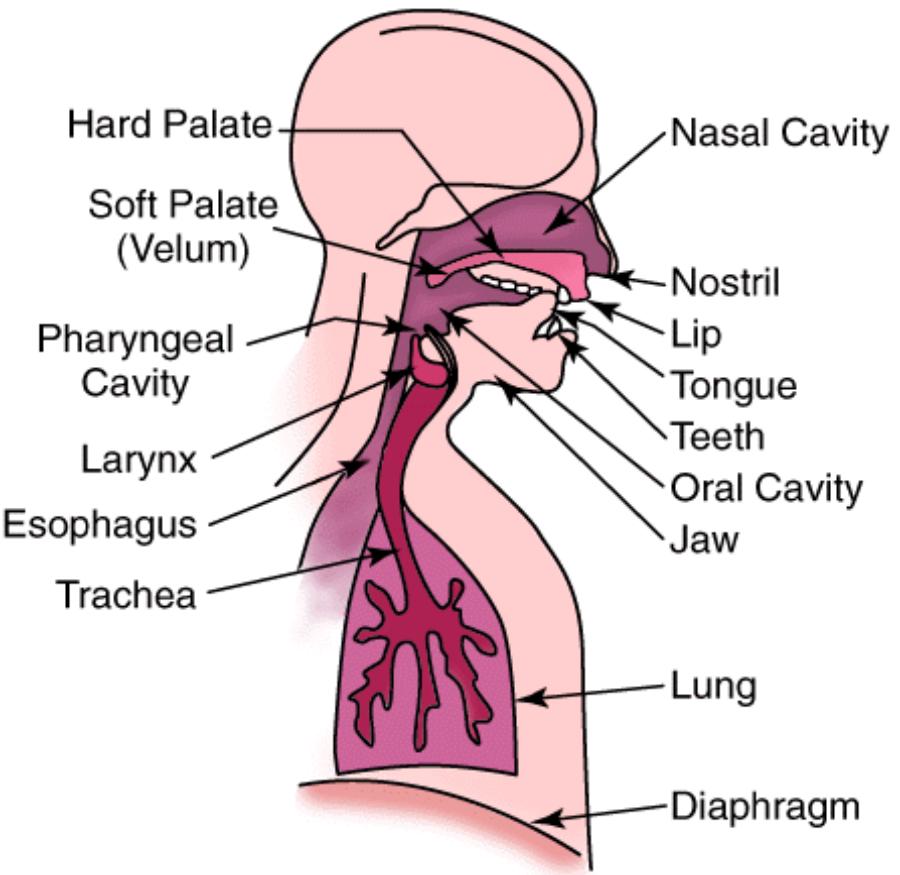
Estudia las propiedades del mecanismo fisiológico de producción de los sonidos y caracteriza los sonidos mediante descripciones del estado del aparato fonador durante la emisión del sonido.



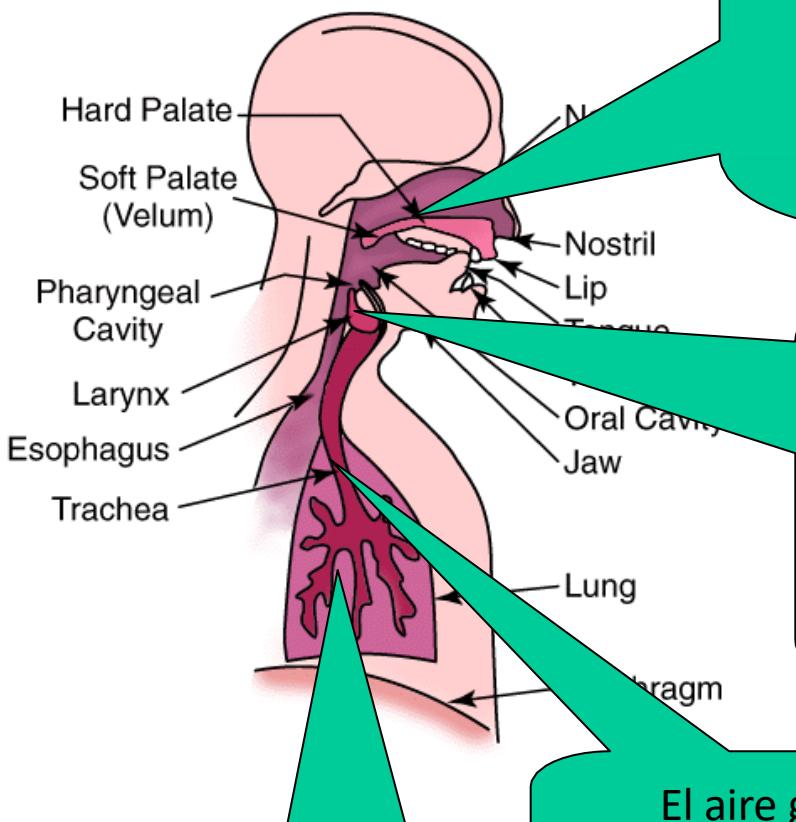
Cada posición de los órganos articulares da un lugar a un sonido determinado. Cualquier modificación origina otro diferente.

Aparato fonador humano

- Órganos Respiratorios. (Sistema Subglotal):
 - Pulmones, Bronquios y Traquea.
Fuente de Energía.
- Órganos Fonadores (Cavidades Glóticas):
 - Laringe, Cuerdas Vocales
y Resonadores (nasal, bucal y faríngeo).
- Órganos articulatorios (Sistema Supraglotal)
 - Paladar, Lengua, Dientes y
Labios.
- Voz: Onda acústica radiada cuando los pulmones expulsan el aire y el flujo resultante es perturbado por alguna constricción en el tracto vocal.



Mecanismo de producción del Habla



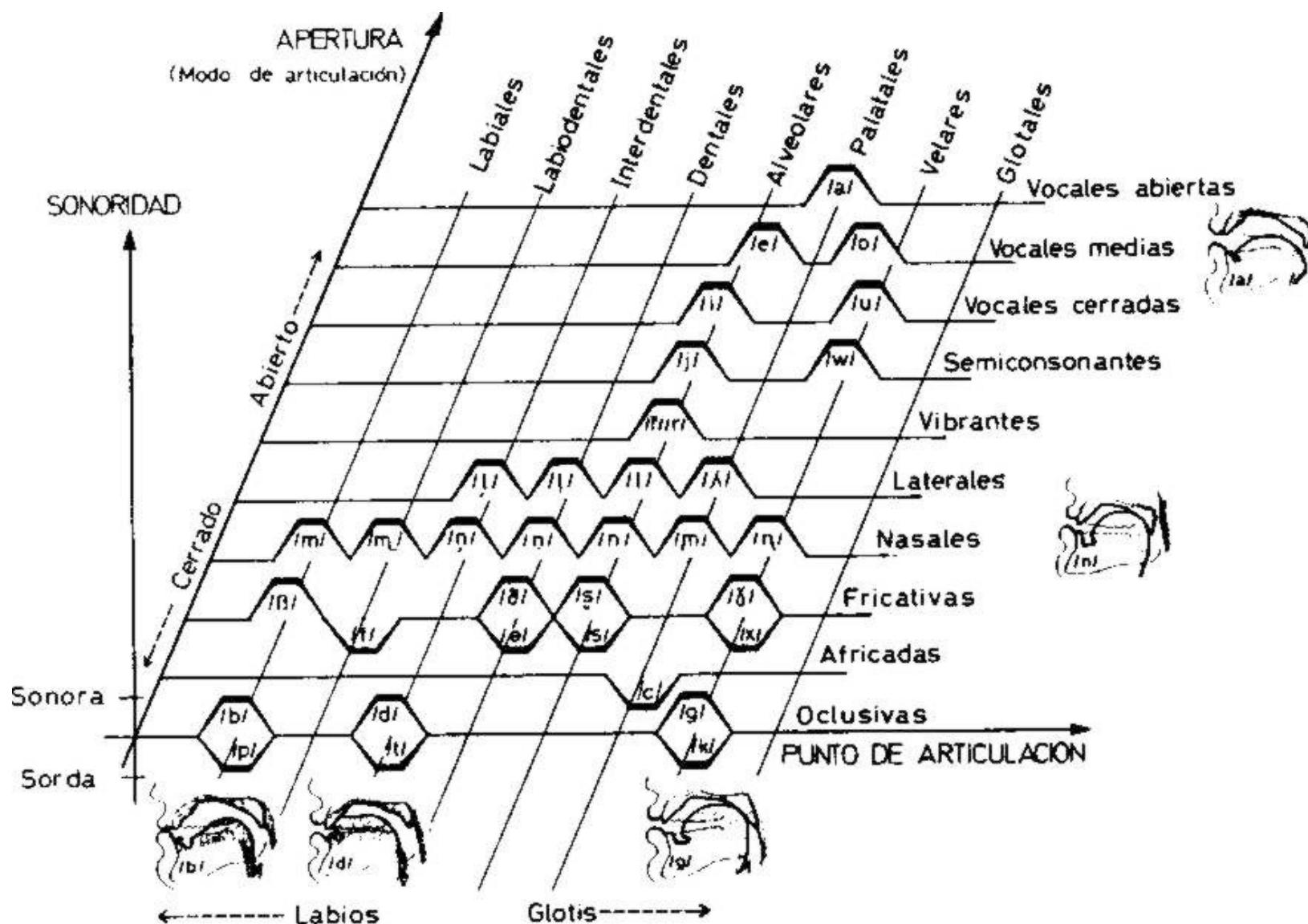
El aire es expulsado con fuerza desde los pulmones

El aire gana velocidad en la tráquea

Sonidos sonoros: los órganos articulatorios, paladar, lengua, dientes, etc. conforman espectralmente el sonido al adoptar la posición apropiada.
Sonidos sordos: El aire encuentra alguna oposición a su paso en algún punto del tracto vocal.

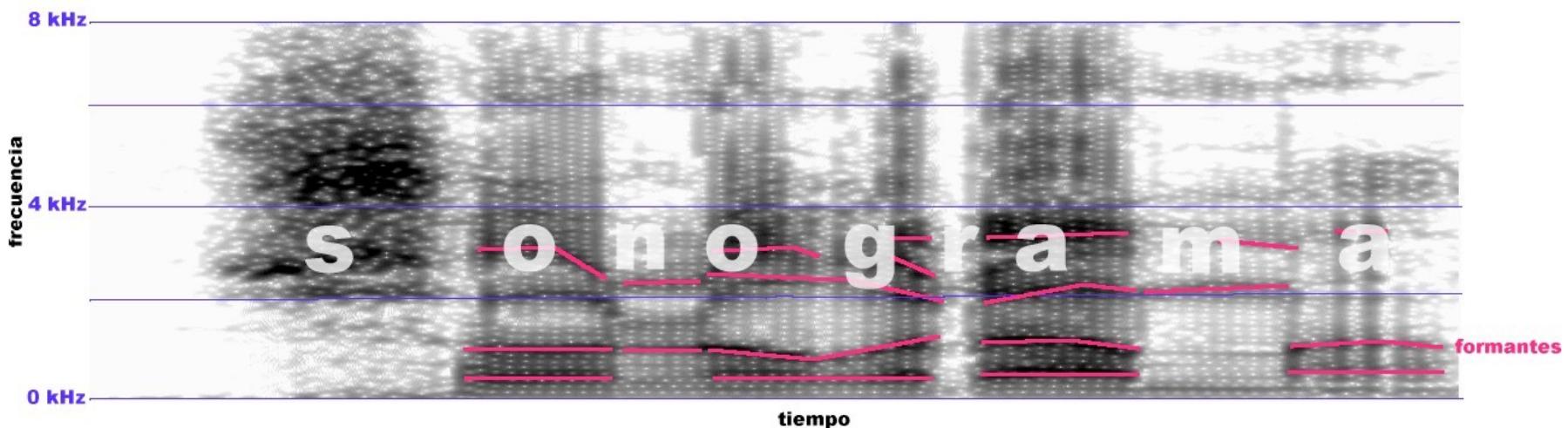
Sonidos sonoros: El aire a su paso hace vibrar las cuerdas vocales. Éstas se encuentran tensas cerrando la glotis.
Sonidos sordos: El aire fluye sin obstáculos a través de la glotis. Ésta se encuentra abierta con las cuerdas vocales relajadas.

He aquí la relación entre sonidos y configuraciones del aparato fonador para el español.



Fonética acústica

Estudia las propiedades de la forma de onda producida: duración, energía, etc. Requiere de instrumental de análisis para obtener medidas de esas propiedades.



Un **espectrograma o sonograma** es una representación **tridimensional** de la señal acústica: representación tiempo-frecuencia de la energía del sonido

- Se dibuja la distribución de energía del sonido en función del tiempo y de las frecuencias.
- El espectrograma se utiliza mucho para analizar la señal de voz.

Vocales

Las vocales en español son fácilmente diferenciables (al menos más fácilmente que en otras lenguas)... pero no aportan tanta información como la consonantes:

- e_ _e__o _i_ _o_a_e_ _o e_ _i_i_i_ _e e__e__e_

Se producen por la vibración de las cuerdas vocales, que es filtrada por la boca/nariz en una posición prácticamente fija.

Su duración es larga en comparación a otros sonidos.

Vocales

Las vocales en español son fácilmente diferenciables (al menos más fácilmente que en otras lenguas)... pero no aportan tanta información como la consonantes:

- _l t_xt_ s_n v_c_l_s n_ _s d_f_c_l d_ _nt_nd_r

Se producen por la vibración de las cuerdas vocales, que es filtrada por la boca/nariz en una posición prácticamente fija.

Su duración es larga en comparación a otros sonidos.

Vocales

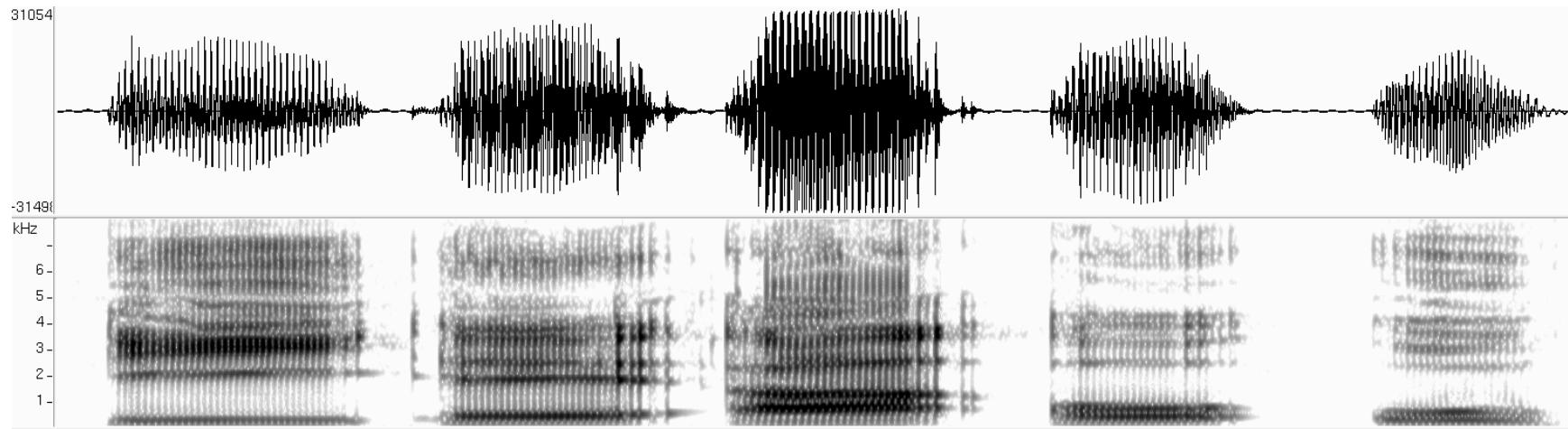
Las vocales en español son fácilmente diferenciables (al menos más fácilmente que en otras lenguas)... pero no aportan tanta información como la consonantes:

- e_ _e__o _i_ _o_a_e_ _o e_ _i_i_i_ _e e__e__e_
- _l t_xt_ s_n v_c_l_s n_ s d_f_c_l d_ nt_nd_r

Se producen por la vibración de las cuerdas vocales, que es filtrada por la boca/nariz en una posición prácticamente fija.

Su duración es larga en comparación a otros sonidos.

Los formantes son bandas oscuras en el sonograma:



pronunciación de “*ieaou*”.

En español, las vocales pueden aparecer en secuencias. Los diptongos son transiciones suaves de una vocal a otra. Normativamente:

- Diptongos crecientes: /i, u/ + /e, a, o/. La /i/ y la /u/ son márgenes silábicos.
- Diptongos decrecientes: /e, a, o/ + /i, u/. La /i/ y la /u/ son márgenes silábicos.
- /i/ + /u/.
- /u/ + /i/.

Además, aparecen otras secuencias: /e, a, o/ + /e, a, o/.

Consonantes explosivas orales

Son sonidos de transición , no continuos, con presión sobre una constricción total (/b/, /p/ en labios; /d/, /t/ tras los dientes; /g/, /k/ cerca del velo).

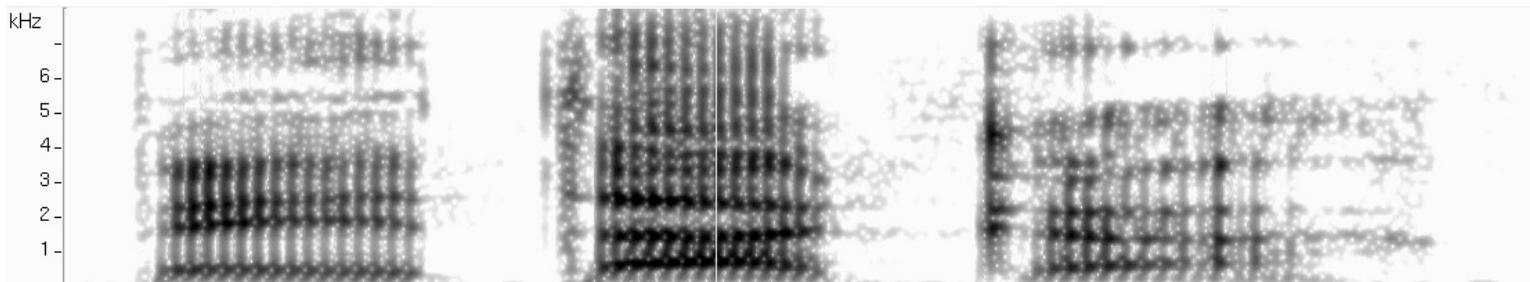
p, b, t, d, k, g

El término se debe a que el momento más audible es el de una explosión. Son difíciles de distinguir.

Desde el punto de vista articulatorio se las denomina oclusivas orales, pues se caracterizan por un cierre momentáneo del canal bucal.

Los fonemas explosivos se dividen acústicamente en **sordos y sonoros**.

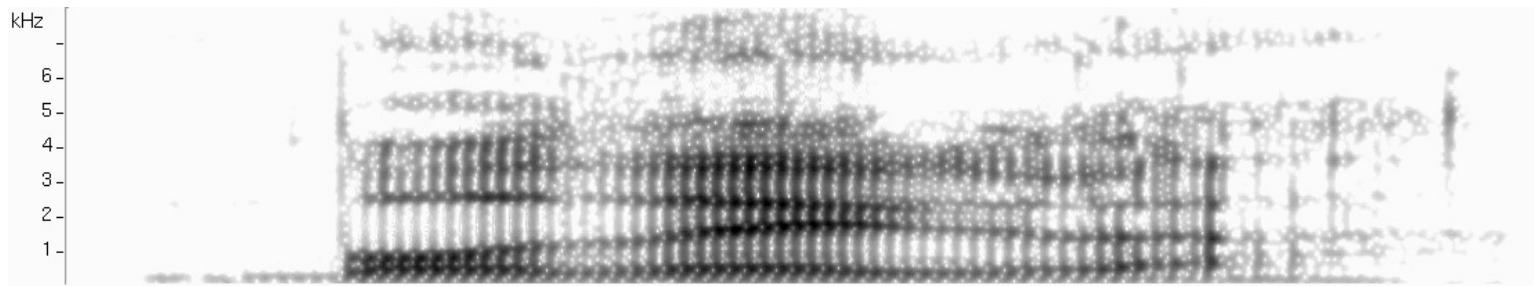
- sordos: p, t, k.



pronunciación de “*petaca*” /*petaka*/.

En el sonograma se aprecia una zona de silencio seguida de un estallido (*burst*) claramente diferenciado del segmento vocálico. Toda la información discriminativa está en el *burst*. Al pronunciar una frase, las regiones de silencio suelen corresponder a los fonemas explosivos, y no a pausas entre palabras (que no suele haberlas).

- sonoros: b, d, g.



pronunciación de “*bodega*” /*bodega*/.

En el sonograma no se aprecian regiones de silencio, aunque sí de disminución de energía.

En el segmento consonántico se aprecia el trazo de la transición de formantes de las vocales que flanquean al fonema explosivo.

Desde el punto de vista articulatorio se dividen en

- labiales (/p, b/),
- dentales (/t, d/) y
- velares (/k, g/).

En posición prenuclear (al principio de una sílaba, antes de vocal) se manifiestan plenamente. En forma postnuclear (al final de sílaba) se manifiestan de formas muy diversas, llegando incluso a desaparecer:

/doktór/ /dogtór/ /doχtór/ /doutór/ /dotór/

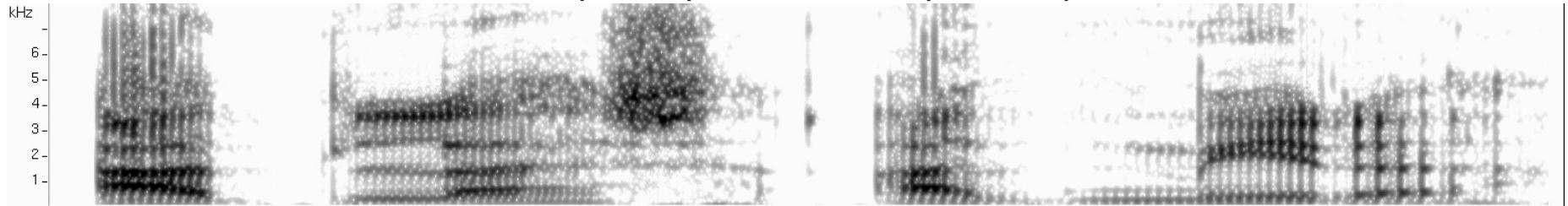
Es habitual que en posición postnuclear se manifiesten como fricativos. El resultado de la neutralización de las explosivas son los archifonemas /B, D, G/. (Un archifonema se produce cuando dos fonemas pierden sus rasgos distintivos).

apnea /áBnea/ ábside /áBside/



Pronunciación de “apnea ábside”

atlas /áDlas/ admira /aDmíra/



Pronunciación de “*atlas admira*”

acta /áGta/ signo /síGno/



Pronunciación de “*acta signo*”

Consonantes explosivas nasales

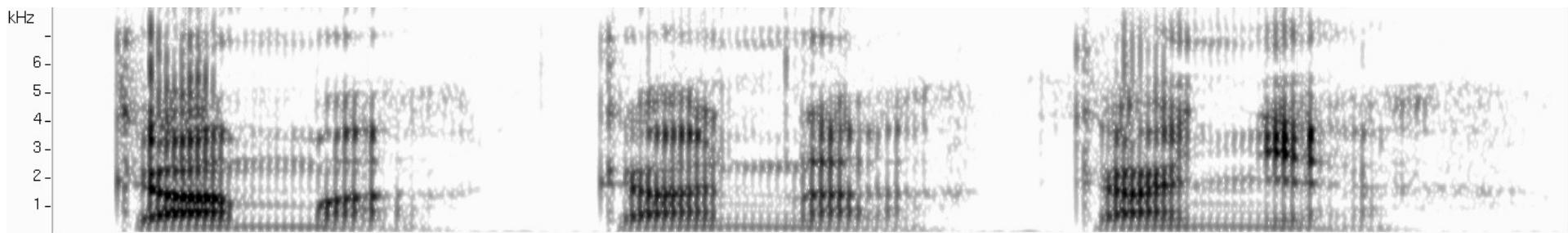
La cavidad oral está ocluida (en /m/, los labios; en /n/ tras los dientes). El velo desciende para que el aire pase por la nariz.

Hay tres explosivas nasales en español: /m, n, η/.

cama /káma/

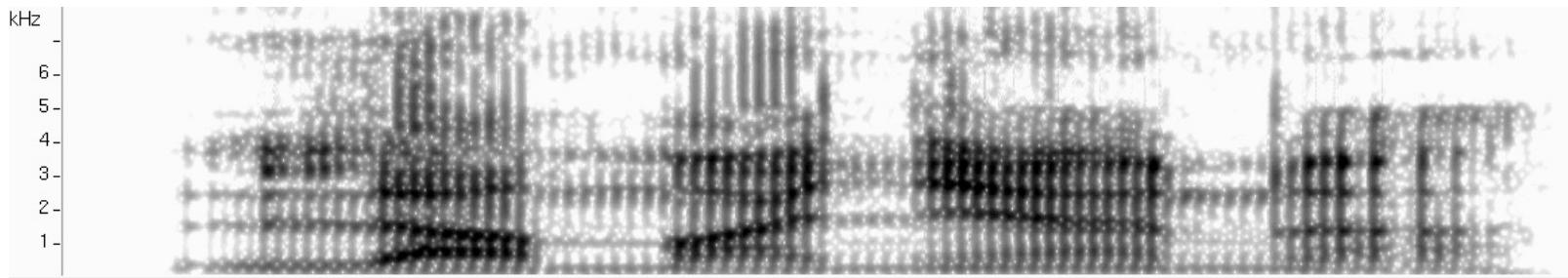
cana /kána/

caña /káηa/



En posición prenuclear se manifiestan plenamente.

- La /m/ presenta tres formantes con independencia de la vocal silábica.
- La /n/ presenta los formantes 1 y 3 y pocas veces el formante 2.
- La /ŋ/ suele presentar sólo el primer formante y aparecer en blanco la zona de altas frecuencias..



Pronunciación de “*la mañana*”

Consonantes fricativas

Se produce una fricción del aire al pasar por una estrechez entre órganos articulatorios. Se caracterizan por el ruido de fricción y modifican los formantes vocálicos contiguos.

- Fricativas de resonancias bajas (sonoras).

- /β/ (labial)

Se diferencia de /b/ en la presencia de zonas que se aproximan a los formantes vocálicos.

bomba /bób̥ba/ boba /bóβa/

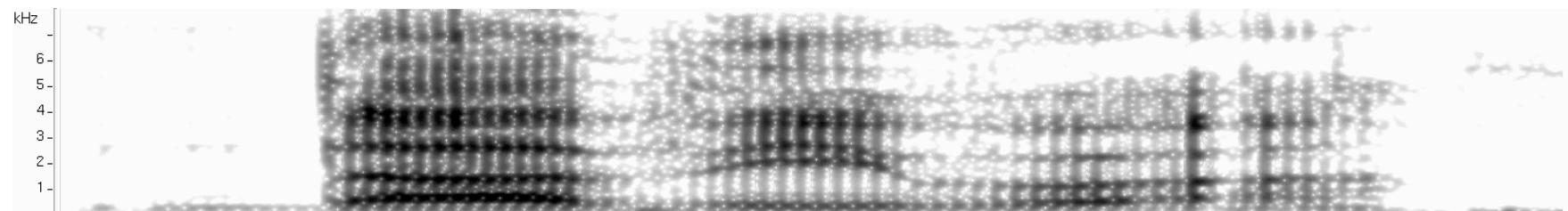


Pronunciación de “*bomba boba*”

- /ð/ (dental)

Se diferencia de /d/ en la presencia de zonas que se aproximan a los formantes vocálicos.

dádiva /dáðiβa/



Pronunciación de “dádiva”.

- velar /γ/

Se diferencia de /g/ en la presencia de zonas que se aproximan a los formantes vocálicos.

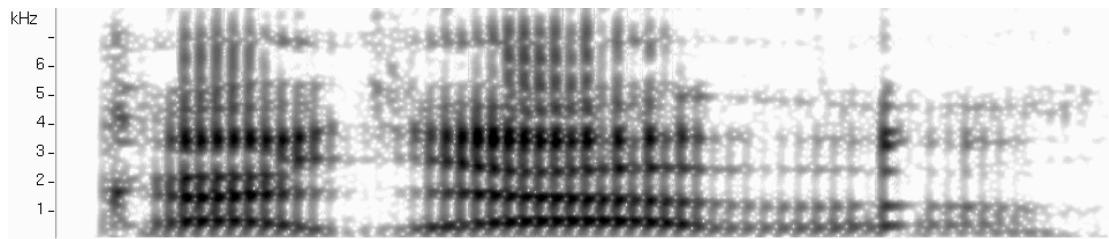
venga /bέŋga/ vega /bέγa/



Pronunciación de “venga vega”.

- /j/ (palatal)

cayado /cajádo/

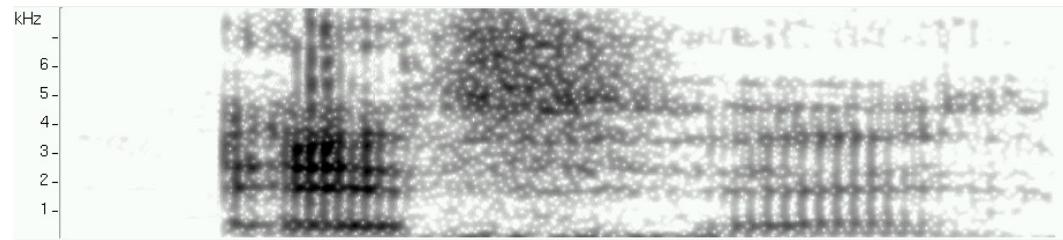


Pronunciación de “cayado”.

- Fricativas de resonancias altas (sordas).

Un flujo estable de aire se vuelve turbulento en una región constrictiva. La energía se concentra en frecuencias altas y es de naturaleza no periódica.

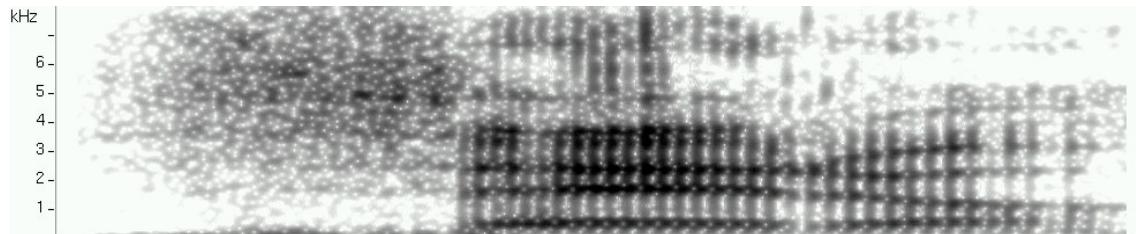
- /f/ (labial) Intensidad débil.



Pronunciación de “efe”

- /θ/ (dental)

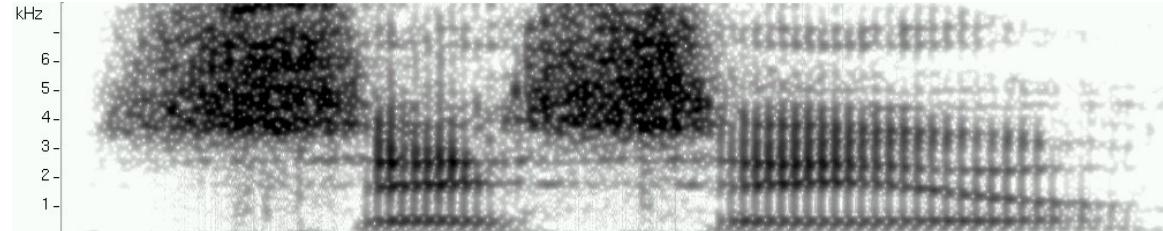
Su frecuencia varía mucho (/θa/ presenta un rango de frecuencias muy diferente de /θu/). Intensidad débil.



Pronunciación de “*cero*”

– /s/ (alveolar)

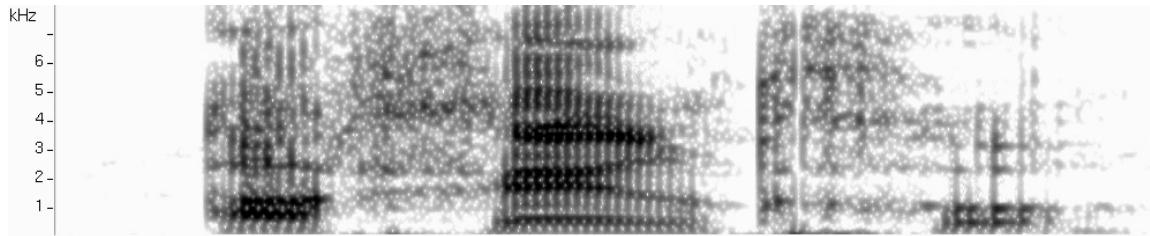
Ruido muy intenso en alta frecuencia. Presenta muchas realizaciones diferentes.



Pronunciación de “seseo”

– /χ/ (velar)

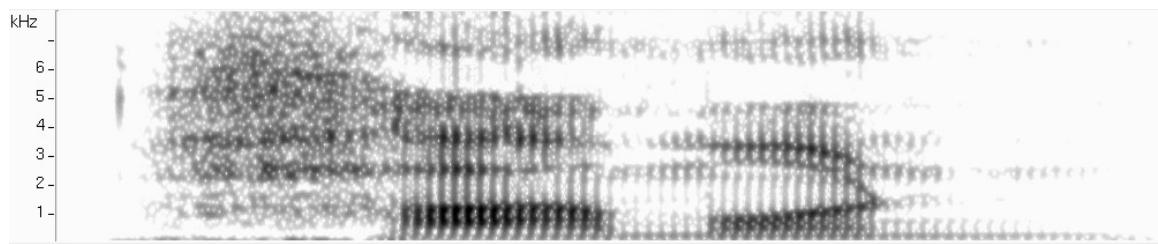
Es vibrante, por lo que presenta estriaciones verticales en el espectro.



Pronunciación de “ajenjo”

(En Chile hay un alófono de /χ/ ante /i/ y /e/: /ç/)

- /h/



Pronunciación de “jamón” (como /hamón/)

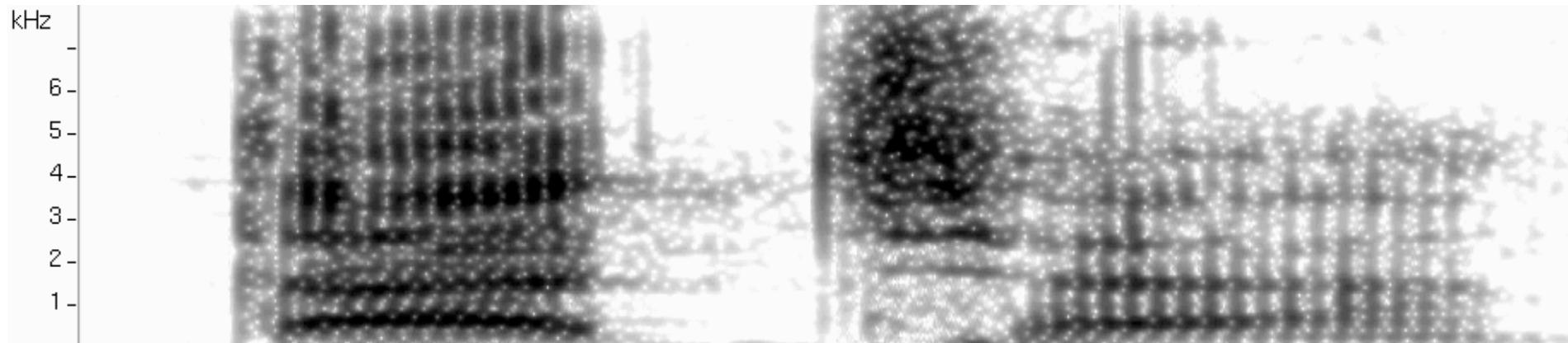
(Propio del acento andaluz y de algunos países latinoamericanos.) Breve sonido turbulento. Muy débil acústicamente.

Africadas

Aparece un momento interrupto y otro constrictivo en su realización.

- Africadas sordas.

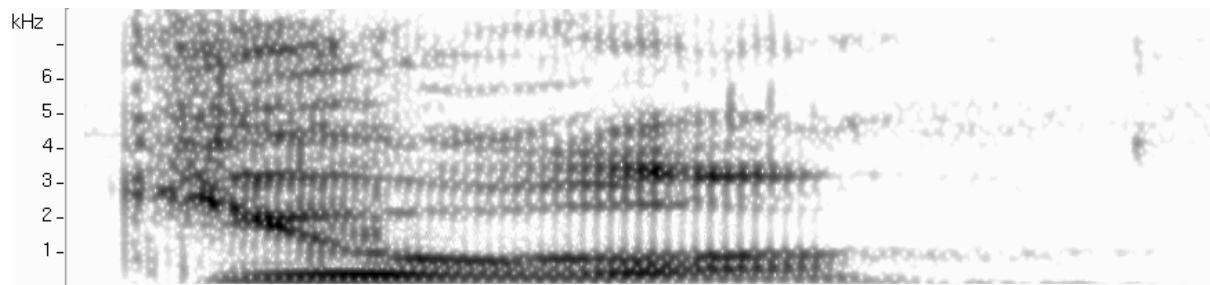
/tʃ/ (tacha /tátʃa/).



Pronunciación de “tacha”.

- Africadas sonoras.

/χ/ (yugo /χúγo/).



Pronunciación de “yugo”.

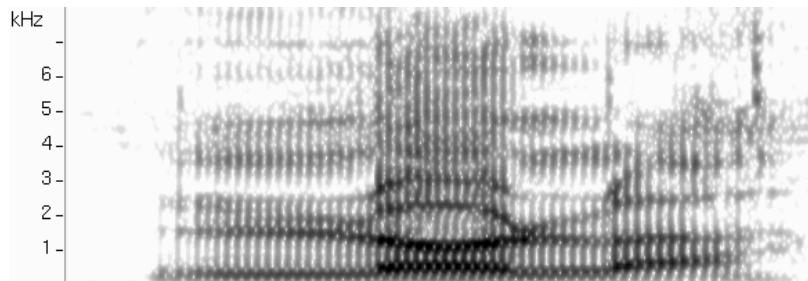
Muchas diferencias en su realización (acentos).

En ocasiones (tras consonante nasal, como en “cónyuge” (/kónχuxε/), no presenta momento interrupto.

Líquidas

- Laterales.

- /l/



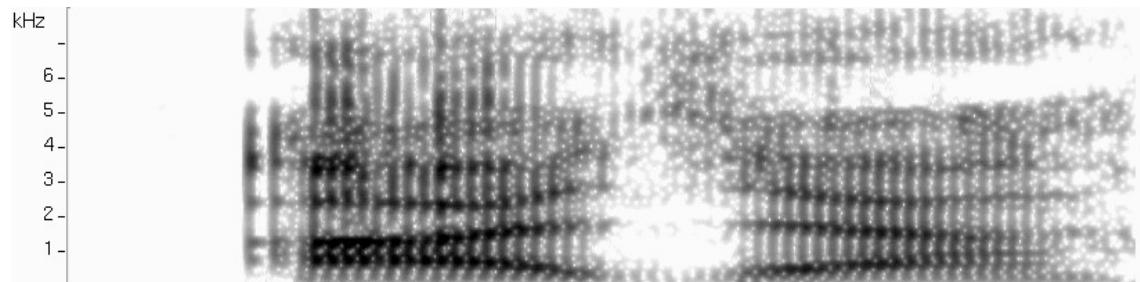
Pronunciación de “lola”.

Aparecen formantes similares a los vocálicos.

Presenta muchos alófonos:

- * /l/ alveolar (ala /ála/),
- * /l/ linguodental (alto /álto/),
- * /l/ linguointer dental (alza /álθa/),
- * /l/ linguopalatalizada (colcha /kólʃa/).

- /λ/

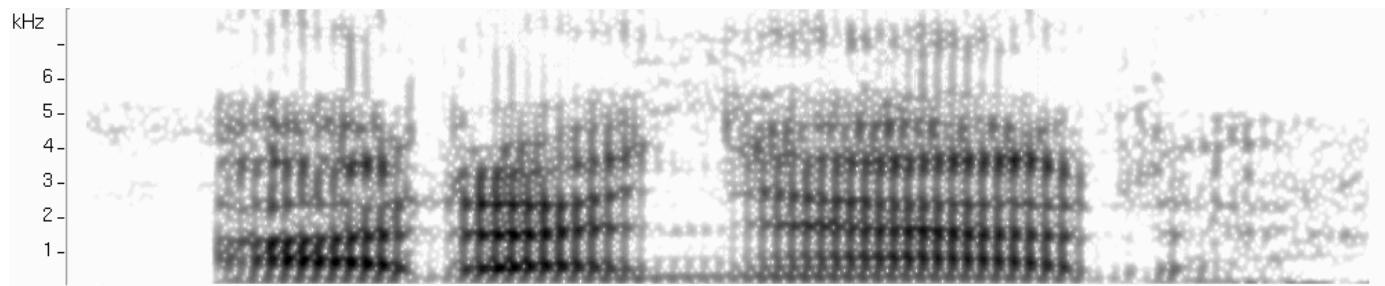


pronunciación de “*allá*”.

Los dos primeros formantes que presenta son de frecuencia ligeramente inferior a los de /i/ y el tercero algo superior.

- Vibrantes.

- /r/ simple.



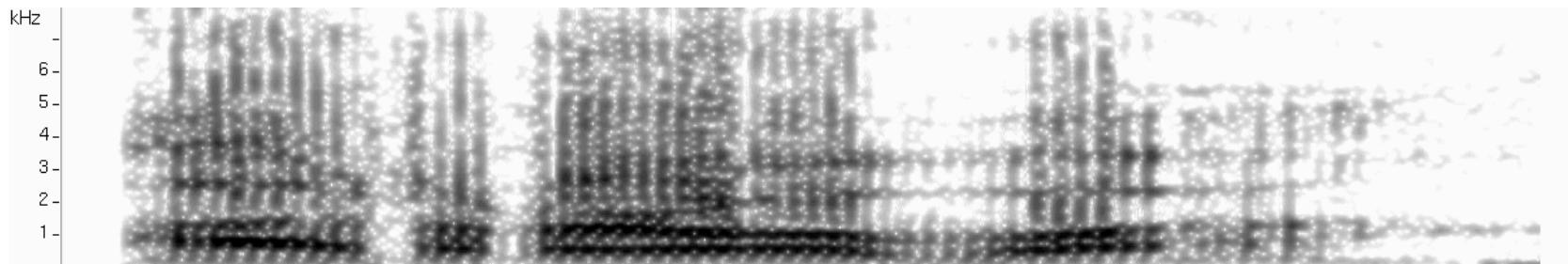
Pronunciación de “arañar”.

Interrumpo muy breve.

Presenta una variante africada /ɾ/.

- /r̚/ múltiple.

Más larga que /r/. Una media de tres interrupciones.



Pronunciación de “arroba”.

Presenta una variante asibilada /r̚/ y una variante fricativa /ɹ/ (y otra faríngea en Puerto Rico).

El elemento esvarabático es la presencia de un elemento vocálico que hace que “prado”, por ejemplo, se realice como “parado”.

Otros fenómenos fonológicos

Omitimos de esta introducción el estudio de fenómenos como el acento prosódico y la entonación , aunque, naturalmente, influyen en la percepción .

La prosodia informa sobre:

- intención de la pronunciación (pregunta, orden, sentencia),
- énfasis (centra la atención en una parte específica),
- resolución de ambigüedades sintácticas/semánticas,
- estado anímico del hablante.

La información prosódica enriquece la pronunciación de las frases con:

- entonación ,
- pausas,
- acento,
- ritmo.

Estructura lingüística del habla

Además de sus características acústicas, las señales de voz se pueden caracterizar en términos de su **estructura lingüística**.

- La estructura lingüística se refiere a las regularidades recurrentes en el lenguaje hablado según lo describen las teorías lingüísticas, tales como cuáles son los componentes básicos del habla y cómo se organizan.
- Las descripciones lingüísticas proporcionan un medio para la interpretación, conceptualización y comunicación sistemáticas de los fenómenos relacionados con el habla.

Unidades elementales del lenguaje hablado

En términos de organización fonética básica, el habla puede verse como una organización jerárquica de unidades elementales:

- En el nivel más bajo de la jerarquía, están los **phones**, que se consideran realizaciones físicas de los **fonemas**.
- Las secuencias de fonemas se organizan en sílabas.
- Y las sílabas forman **palabras** (donde cada palabra consta de una o más sílabas).
- Luego, una o más palabras forman **frases**.

Bibliografía

- Antonio Quilis: Fonética acústica de la lengua española. Biblioteca Románica Hispánica. Editorial Gredos. Madrid. 1981.
- Antonio Quilis, José A. Fernández: Curso de Fonética y fonología españolas para estudiantes angloamericanos. CSIC, Instituto Miguel de Cervantes. 1979.
- Lawrence Rabiner, Biing-Hwang Juang: Fundamentals of speech recognition. Prentice Hall. 1993.
- Francisco Casacuberta, Enrique Vidal: Reconocimiento automático del habla. Marcombo. 1987.
- María José Castro, Salvador España, Andrés Marzal, Ismael Salvador. Transcripción ortográfico-fonéticos para el castellano. Procesamiento del Lenguaje Natural, 27:241–245, septiembre 2001.
- Antonio Ríos Mestre. La transcripción Fonética automática del diccionario electrónico de formas simples flexivas del español: estudio fonológico en el léxico. Estudios de lingüística Española, 4, 1999.