

LA VOZ HUMANA

Federico Miyara

1. CONCEPTOS PRELIMINARES

1.1. Comunicación y lenguaje

Los sistemas de comunicación transportan información. Nos proponemos estudiar un sistema de comunicación específico, el de la comunicación a través de señales de voz, es decir señales acústicas tradicionalmente emitidas y recibidas por seres humanos en forma oral. Algunos de nuestros objetivos son la representación, análisis, modificación, mejora de la relación señal/ruido, generación artificial de mensajes vocales inteligibles para el ser humano y el reconocimiento automático de mensajes vocales pronunciados por seres humanos.

Históricamente, desde la Antigua Grecia se han realizado intentos por generar voces artificiales. En muchos casos eran simplemente juegos de tuberías conectadas a un locutor humano, en otros auténticos ingenios acústicos capaces de producir sonoridades vocálicas [1]. El desarrollo de la telefonía a principios del siglo XX motivó intensas investigaciones sobre las propiedades de la voz y la audición con el fin de mejorar la calidad de la comunicación telefónica. El proceso continuó y hoy en día las tecnologías existentes permiten, por ejemplo, disponer de sistemas de comunicación oral hombre-máquina [2].

En todo sistema de comunicación hay varios componentes: *emisor*, *receptor*, *mensaje*, *código*, *canal* y *contexto* [3]. Es necesario conocer algunos aspectos de cada uno de ellos para poder integrar sistemas que funcionen de manera eficaz y eficiente. En nuestro caso el emisor es el conjunto integrado por el cerebro que “piensa” el mensaje y el aparato fonatorio que lo “traduce” a una emisión acústica. El receptor es el aparato auditivo que recibe la onda sonora y la transforma en impulsos nerviosos que luego son interpretados por el cerebro.¹ El mensaje es la idea a comunicar. El código es el lenguaje hablado. La combinación del mensaje y el código constituyen la *señal*. El canal puede ser el medio en el cual se propaga la onda sonora (en general el aire) o un medio de transmisión electrónico que constituye en sí mismo otro subsistema de comunicación cuyas propiedades son bien conocidas y que se aproxima en muchos casos (aunque no siempre) a la idealidad. El contexto puede tener un sinnúmero de componentes, que van desde factores puramente subjetivos o psicológicos, como el interés, la atención, la motivación hasta factores físicos tales como respuesta en frecuencia, interferencias, distorsiones, ruido, etc.

¹ Por supuesto, también hay un sistema de comunicación local en cualquiera de los dos procesos mencionados, ya que de la generación del mensaje a su emisión final hay una transferencia de información.

1.2. Algunos conceptos sobre lenguaje [3]

La *lengua* es un sistema de signos lingüísticos que permiten la comunicación en una comunidad. Es un sistema pues cada uno de sus elementos tiene *entidad propia* y *entidad relativa* a su posición o relación con los otros elementos. Es un código de signos. Tiene carácter social, ya que es común a una sociedad.

El *habla* es el acto de seleccionar los signos de entre los disponibles y organizarlos a través de ciertas reglas. Materializa el código. Es individual, vale decir que cambia de un individuo a otro. Los signos pueden corresponder al lenguaje escrito o al oral.

El *lenguaje* es un *sistema* articulado ya que los sonidos y otros componentes se integran entre sí. Está formado por *signos lingüísticos*, nombre que recibe la señal en el lenguaje.

El lenguaje tiene modalidades regionales llamadas *dialectos*.

Un *signo* es algo que reemplaza a otra cosa para comunicarla en un *mensaje*.

Los signos lingüísticos se clasifican en dos tipos: *significado* y *significante*. El *significado* es el concepto mental, idea o contenido a comunicar. El *significante* es la *imagen*, ya sea *gráfica* o *acústica* que se le asigna.

La relación entre significado y significante es arbitraria o convencional, aunque no necesariamente discrecional: involucra acuerdos tácitos, explícitos o normativos en una comunidad lingüística.

En el lenguaje escrito, el significante es la *grafía escrita*, formada por combinaciones de *letras*, en tanto que en el lenguaje hablado es su *realización acústica* mediante la *palabra hablada*.

Las *palabras* son los elementos *libres* mínimos del lenguaje. La *sintaxis* es el conjunto de reglas para la coordinación de las palabras en frases u oraciones. En su versión escrita las palabras están formadas por letras o *grafemas*, es decir unidades gráficas mínimas, y, en el caso oral, por *fonemas*.

Los *fonemas* son la unidad fónica ideal mínima del lenguaje. Se materializan a través de los sonidos, pero de una manera no unívoca. Las variantes de los fonemas se denominan *alófonos*.

Los *monemas* son unidades mínimas con significado, que puede ser *gramatical*, dando origen a los *morfemas*, o *léxico*, representado por los *lexemas*. Los morfemas tienen relación con la gramática, o la forma de organizar o dar estructura a las categorías básicas del lenguaje (genero, número, tiempo o persona de los verbos, etc.), mientras que los lexemas se refieren a significados externos al lenguaje mismo.

Las palabras constan de al menos un monema, siendo las más comunes *bimone-máticas*, que incluyen un lexema y un morfema. En la tabla siguiente se dan dos ejemplos en los que se identifican los componentes de la palabra.²

Tabla 1. Ejemplos de monemas, grafemas y fonemas

Palabra	Monemas		Grafema	Fonemas
	Lexema	Morfema		
Gato	Gat	o	G, a, t, o	/g/, /a/, /t/, /o/
Amaban	Ama	ban	A, m, a, b, a, n	/a/, /m/, /a/, /b/, /a/, /n/

² Notar que los fonemas se representan entre barras, por ejemplo /b/. Su realización es un sonido.

1.3. Fonología y fonética [3,4]

La *Fonología* estudia los *fonemas*, es decir el modelo fónico convencional e ideal del lenguaje. La *Fonética*, en tanto, se refiere a los *sonidos* en el habla, incluyendo su producción acústica y los procesos físicos y fisiológicos de emisión y articulación involucrados.

Así, la Fonología es el estudio de los sonidos de la lengua en cuanto a su carácter simbólico o de representación mental. Procede detectando regularidades o recurrencias en los sonidos del lenguaje hablado y sus combinaciones, y haciendo abstracción de las pequeñas diferencias debidas a la individualidad de cada hablante y de características *suprasegmentales* como la entonación, el acento (*tónico*, es decir por aumento de la intensidad y *agógico*, por aumento de la duración), etc. Cada uno de los sonidos abstractos así identificados es un *fonema*. Uno de los objetivos de la fonología es acotar al máximo la cantidad de fonemas requeridos para representar cada idioma de una manera suficientemente precisa.

La Fonética estudia experimentalmente los mecanismos de producción y percepción de los sonidos utilizados en el habla a través del análisis acústico, articulatorio y perceptivo. Se ocupa, por consiguiente, de las realizaciones de los fonemas.

2. La voz humana

2.1. Breve anatomía del aparato fonatorio

La voz humana se produce voluntariamente por medio del aparato fonatorio. Éste está formado por los pulmones como fuente de energía en la forma de un flujo de aire, la *laringe*, que contiene las *cuerdas vocales*, la *faringe*, las *cavidades oral* (o bucal) y *nasal* y una serie de elementos articulatorios: los *labios*, los *dientes*, el *alvéolo*, el *paladar*, el *velo del paladar* y la *lengua* (figura 1).

Las *cuerdas vocales* son, en realidad, dos membranas dentro de la laringe orientadas de adelante hacia atrás (figura 2). Por adelante se unen en el *cartílago tiroideo* (que puede palparse sobre el cuello, inmediatamente por debajo de la unión con la cabeza; en los varones suele apreciarse como una protuberancia conocida como *nuez de Adán*). Por detrás, cada una está sujeta a uno de los dos *cartílagos aritenoides*, los cuales pueden separarse voluntariamente por medio de músculos. La abertura entre ambas cuerdas se denomina *glotis*.

Cuando las cuerdas vocales se encuentran separadas, la glotis adopta una forma triangular. El aire pasa libremente y prácticamente no se produce sonido. Es el caso de la respiración.

Cuando la glotis comienza a cerrarse, el aire que la atraviesa proveniente de los pulmones experimenta una turbulencia, emitiéndose un ruido de origen aerodinámico conocido como *aspiración* (aunque en realidad acompaña a una espiración o exhalación). Esto sucede en los sonidos denominados “aspirados” (como la h inglesa).

Al cerrarse más, las cuerdas vocales comienzan a vibrar a modo de lengüetas, produciéndose un sonido tonal, es decir periódico. La frecuencia de este sonido depende de varios factores, entre otros del tamaño y la masa de las cuerdas vocales, de la tensión que se les aplique y de la velocidad del flujo del aire proveniente de los pulmones. A mayor tamaño, menor frecuencia de vibración, lo cual explica por qué en los varones, cuya glotis es en promedio mayor que la de las mujeres, la voz es en general más grave. A mayor tensión la frecuencia aumenta, siendo los sonidos más agudos. Así, para lograr

emitir sonidos en el registro extremo de la voz es necesario un mayor esfuerzo vocal. También aumenta la frecuencia (a igualdad de las otras condiciones) al crecer la velocidad del flujo de aire, razón por la cual al aumentar la intensidad de emisión se tiende a elevar espontáneamente el tono de voz.

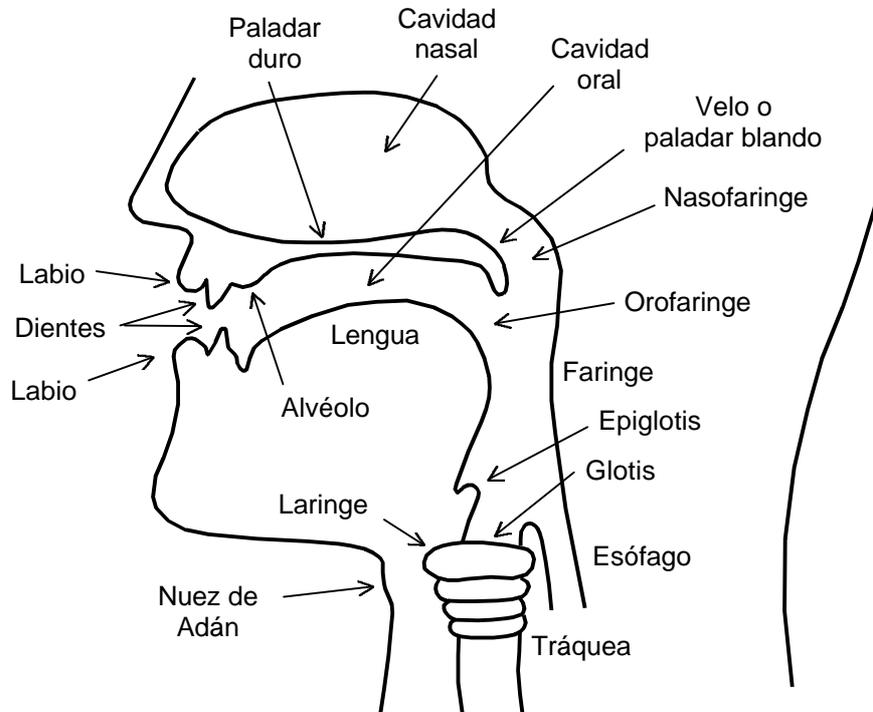


Figura 1. Corte esquemático del aparato fonatorio humano.

Finalmente, es posible obturar la glotis completamente. En ese caso no se produce sonido. Sobre la glotis se encuentra la *epiglotis*, un cartílago en la faringe que permite tapar la glotis durante la deglución para evitar que el alimento ingerido se introduzca en el tracto respiratorio. Durante la respiración y la fonación (emisión de sonido) la epiglotis está separada de la glotis permitiendo la circulación del flujo de aire. Durante la deglución, en cambio, la laringe ejecuta un movimiento ascendente de modo que la glotis apoya sobre la epiglotis.

La porción que incluye las cavidades faríngea, oral y nasal junto con los elementos articulatorios se denomina genéricamente *cavidad supraglótica*, en tanto que los espacios por debajo de la laringe, es decir la tráquea, los bronquios y los pulmones, se denominan *cavidades infraglóticas*.

Varios de los elementos de la cavidad supraglótica se controlan a voluntad, permitiendo modificar dentro de márgenes muy amplios los sonidos producidos por las cuerdas vocales o agregar partes distintivas a los mismos, e inclusive producir sonidos propios. Todo esto se efectúa por dos mecanismos principales: el *filtrado* y la *articulación*.

El *filtrado* actúa modificando el espectro del sonido. Tiene lugar en las cuatro cavidades supraglóticas principales: la faringe, la cavidad nasal, la cavidad oral y la cavidad labial. Las mismas constituyen resonadores acústicos que enfatizan determinadas bandas frecuenciales del espectro generado por las cuerdas vocales, conduciendo al concepto de *formantes*, es decir una serie de picos de resonancia ubicados en frecuen-

cias o bandas de frecuencia que, según veremos, son bastante específicas para cada tipo de sonido.

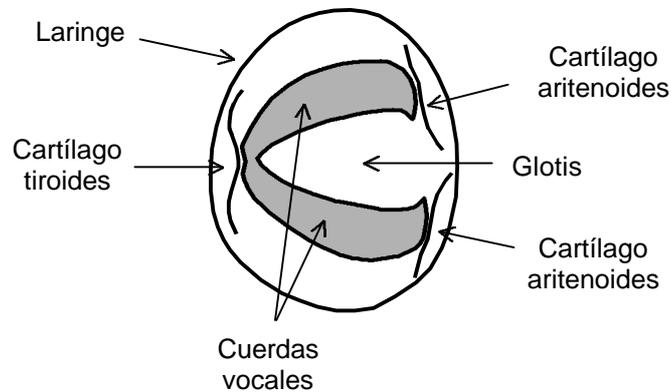


Figura 2. Corte esquemático de la laringe según un plano horizontal.

La *articulación* es una modificación principalmente a nivel temporal de los sonidos, y está directamente relacionada con la emisión de los mismos y con los fenómenos transitorios que los acompañan. Está caracterizada por el lugar del tracto vocal en que tiene lugar, por los elementos que intervienen y por el modo en que se produce, factores que dan origen a una clasificación fonética de los sonidos que veremos luego.

2.2. Clasificación de los sonidos de la voz

Los sonidos emitidos por el aparato fonatorio pueden clasificarse de acuerdo con diversos criterios que tienen en cuenta los diferentes aspectos del fenómeno de emisión. Estos criterios son:

- a) Según su carácter vocálico o consonántico.
- b) Según su oralidad o nasalidad
- c) Según su carácter tonal (sonoro) o no tonal (sordo)
- d) Según el lugar de articulación
- e) Según el modo de articulación
- f) Según la posición de los órganos articulatorios
- g) Según la duración

2.2.1. Vocales y consonantes

Desde un punto de vista mecanoacústico, las *vocales* son los sonidos emitidos por la sola vibración de las cuerdas vocales sin ningún obstáculo o constricción entre la laringe y las aberturas oral y nasal. Dicha vibración se genera por el principio del oscilador de relajación, donde interviene una fuente de energía constante en la forma de un flujo de aire proveniente de los pulmones. Son siempre sonidos de carácter tonal (cuasi-periódicos), y por consiguiente de espectro discreto. Las *consonantes*, por el contrario, se emiten interponiendo algún obstáculo formado por los elementos articulatorios. Los

sonidos correspondientes a las consonantes pueden ser tonales o no dependiendo de si las cuerdas vocales están vibrando o no. Funcionalmente, en el castellano las vocales pueden constituir palabras completas, no así las consonantes.

2.2.2. Oralidad y nasalidad

Los fonemas en los que el aire pasa por la cavidad nasal se denominan *nasales*, en tanto que aquéllos en los que sale por la boca se denominan *orales*. La diferencia principal está en el tipo de resonador principal por encima de la laringe (cavidad nasal y oral, respectivamente). En castellano son nasales sólo las consonantes “m”, “n”, “ñ”

2.2.3. Tonalidad

Los fonemas en los que participa la vibración de las cuerdas vocales se denominan *tonales* o, también, *sonoros*.³ La tonalidad lleva implícito un espectro cuasi periódico.⁴ Como se puntualizó anteriormente, todas las vocales son tonales, pero existen varias consonantes que también lo son: “b”, “d”, “m”, etc. Aquellos fonemas producidos sin vibraciones glotales se denominan *sordos*. Varios de ellos son el resultado de la turbulencia causada por el aire pasando a gran velocidad por un espacio reducido, como las consonantes “s”, “z”, “j”, “f”.

2.2.4. Lugar y modo de articulación (consonantes)

La *articulación* es el proceso mediante el cual alguna parte del aparato fonatorio interpone un obstáculo para la circulación del flujo de aire. Las características de la articulación permitirán clasificar las consonantes. Los órganos articulatorios son los labios, los dientes, las diferentes partes del paladar (alvéolo, paladar duro, paladar blando o velo), la lengua y la glotis. Salvo la glotis, que puede articular por sí misma, el resto de los órganos articula por oposición con otro. Según el lugar o punto de articulación se tienen fonemas:

Bilabiales: oposición de ambos labios

Labiodentales: oposición de los dientes superiores con el labio inferior

Linguodentales: oposición de la punta de la lengua con los dientes superiores

Alveolares: oposición de la punta de la lengua con la región alveolar

Palatales: oposición de la lengua con el paladar duro

Velares: oposición de la parte posterior de la lengua con el paladar blando

Glotales: articulación en la propia glotis

A su vez, para cada punto de articulación ésta puede efectuarse de diferentes modos, dando lugar a fonemas:

³ Debe tenerse cuidado con la palabra “sonoro”, ya que en Psicoacústica es un concepto asociado a la sonoridad, es decir la sensación de intensidad. Desde este punto de vista la “s” es sonora, ya que tiene una intensidad importante, pero no es tonal.

⁴ En realidad las vocales sólo son periódicas cuando el locutor hace el esfuerzo de emitir un sonido sin inflexiones. La entonación implica una variación de la frecuencia.

Oclusivos: la salida del aire se cierra momentáneamente por completo

Fricativos: el aire sale atravesando un espacio estrecho

Africados: oclusión seguida por fricación

Laterales: la lengua obstruye el centro de la boca y el aire sale por los lados

Vibrantes: la lengua vibra cerrando el paso del aire intermitentemente

Aproximantes: La obstrucción muy estrecha que no llega a producir turbulencia

Los fonemas oclusivos (correspondientes a las consonantes “b” inicial o postnasal,⁵ “c”, “k”, “d”, “g” inicial, postnasal o postlateral,⁶ “p”, “t”) también se denominan a veces *explosivos*, debido a la liberación repentina de la presión presente inmediatamente antes de su emisión. Pueden ser sordos o sonoros, al igual que los fricativos (“b” postvocálica, postlateral y postvibrante, “g” postvocálica y post vibrante, “f”, “j”, “h” aspirada, “s”, “y”, “z”). Sólo existe un fonema africado en castellano, correspondiente a la “ch”. Los laterales (“l”, “ll”) a veces se denominan líquidos, y son siempre sonoros. Los dos fonemas vibrantes del castellano (consonantes “r”, “rr”) difieren en que en uno de ellos (“r”) se ejecuta una sola vibración y es intervocálico, mientras que en el otro (“rr”) es una sucesión de dos o tres vibraciones de la lengua. Finalmente, los fonemas aproximantes (la “i” y la “u” cerradas que aparecen en algunos diptongos) son a veces denominados semivocales, pues en realidad suenan como vocales. Pero exhiben una diferencia muy importante: son de corta duración y no son prolongables.

En la tabla 2 se indican las consonantes clasificadas según el lugar y el modo de articulación, la sonoridad y la oro-nasalidad. En algunos casos una misma consonante aparece en dos categorías diferentes, correspondiente a las diferencias observadas.

Tabla 2. Clasificación de las consonantes de la lengua castellana según el lugar y el modo de articulación y la sonoridad.

Lugar de articulación	Modo de articulación								
	Oral							Nasal	
	Oclusiva		Fricativa		Africada	Lateral	Vibrante	Aproximante	Sonora
	Sorda	Sonora	Sorda	Sonora	Sorda	Sonora	Sonora	Sonora	
Bilabial	p	b, v		b, v				w	m
Labiodental			f						
Linguodental			z	d					
Alveolar	t	d	s	y	ch	l	r, rr		n
Palatal				(y)	(ch)	ll		i	ñ
Velar	k	g	j	g					
Glotal			h						

⁵ La “b” castellana (al igual que la “v”, que, contrariamente a lo que se suele creer, no tiene diferencias fonéticas con la “b”) corresponde en realidad a dos fonemas diferentes. El primero, cuando aparece al principio de una palabra (“b” inicial) o después de una nasal (“b” postnasal). El segundo, después de una vocal, de una vibrante o de una lateral (“b” postvocálica, postvibrante o postlateral).

⁶ La “g”, al igual que la “b”, corresponde a dos fonemas diferentes. El primero, cuando está al principio de una palabra (“b” inicial) o después de una nasal o lateral (“b” postnasal o postlateral). El segundo, después de una vocal o de una vibrante (“b” postvocálica o postvibrante).

2.2.5. Posición de los órganos articulatorios (vocales)

En el caso de las vocales, la articulación consiste en la modificación de la acción filtrante de los diversos resonadores, lo cual depende de las posiciones de la lengua (tanto en elevación como en profundidad o avance), de la mandíbula inferior, de los labios [3,7,8] y del paladar blando. Estos órganos influyen sobre los formantes, permitiendo su control.

Podemos clasificar las vocales según la posición de la lengua como se muestra en la tabla 3.

Tabla 3. Clasificación de las vocales castellanas según la posición de la lengua.

Posición vertical	Tipo de vocal	Posición horizontal (avance)		
		Anterior	Central	Posterior
Alta	Cerrada	i		u
Media	Media	e		o
Baja	Abierta		a	

Otra cualidad controlable es la *labialización*, es decir el hecho de que se haga participar activamente los labios. Las vocales labializadas, también definidas como *redondeadas*, son las que redondean los labios hacia adelante, incrementando la longitud efectiva del tracto vocal. La única vocal labializada en el castellano es la “u”.

En otros idiomas, como el francés, el portugués, el catalán y el polaco, así como en lenguas no europeas como el guaraní o el hindi, existe también el matiz de oralidad o nasalidad [4,9]. En las vocales orales el velo (paladar blando) sube, obturando la nasofaringe, lo cual impide que el aire fluya parcialmente por la cavidad nasal. En las vocales *nasalizadas* (u *oronasales*) el velo baja, liberando el paso del aire a través de la nasofaringe. Se incorpora así la resonancia nasal.

2.2.6. Duración

La duración de los sonidos, especialmente de las vocales, no tiene importancia a nivel semántico en el castellano, pero sí en el plano expresivo, a través de la *agogia*, es decir el énfasis o acentuación a través de la duración. En inglés, en cambio, la duración de una vocal puede cambiar completamente el significado de la palabra que la contiene.

2.3. El alfabeto fonético internacional

El castellano es un idioma cuya escritura es eminentemente fonética, ya que salvo pocos casos, hay correspondencia entre grafema y fonema. No todos los idiomas tienen esta característica. El inglés es un caso quizás extremo, a tal punto que George Bernard Shaw ha creado posibles ortografías alternativas para algunas palabras basándose en la forma en que sus fonemas aparecen escritos en otras palabras [5]. Estas extrañas ortografías y el análisis correspondiente se muestran en la tabla 4.

Tabla 4. Ortografías alternativas de George Bernard Shaw para dos palabras inglesas [5].

Palabra	Ortografía alternativa (según Shaw)	Fonema	Palabra en la que se usa la ortografía alternativa	Escritura en el Alfabeto Fonético Internacional
fish	ghoti	GH	enough	[ɪˈnʌf]
		O	women	[ˈwɪmɪn]
		TI	nation	[ˈneɪʃn]
potato	ghoughpteighbteau	GH	hiccough	[ˈhɪcʌp]
		OUGH	though	[ðʰʊ]
		PT	pteranodon	[traˈnɪdon]
		EIGH	neighbour	[ˈneɪbɹ]
		BT	debt	[det]
		EAU	bureau	[bjʊˈrɪʊ]

Se ha compilado un extenso conjunto de símbolos fonéticos conocido como el *Alfabeto Fonético Internacional* (International Phonetic Alphabet, IPA)⁷ que contiene una gran cantidad de fonemas de los diversos idiomas, y que permite representar de una manera inequívoca los fonemas independientemente del idioma. El subconjunto correspondiente al idioma castellano se indica en la tabla 5.

Tabla 5. Los fonemas del alfabeto fonético internacional utilizados en la lengua castellana [6, 11].

Fonemas castellanos					
Sonido	Ejemplo	Sonido	Ejemplo	Sonido	Ejemplo
[p]	p aso	[θ]	zor z al, láp iz	[ɲ]	ma ñ ana, ñ o ñ o
[b]	b ase, v ena	[s]	so l o, co s a	[dʒ]	y o, Y ape y ú
[β]	labo r , lava r	[x]	gi r o, ja r abe	[j]	bie n , bi o lógo
[t]	tr e s, can t o	[tʃ]	he ch o, Ch ubut	[w]	h u eso, bu it re
[d]	d ama, and a r	[r]	ar d er, ja r abe		
[ð]	ce d ro, ver d ad	[rr]	per r o, ro j o	[a]	a ma
[k]	ca s o, dis c o	[l]	lo a ble, fie l	[e]	es e ra, ve r
[g]	gu l a, go m a	[λ]	ll a nto, ca l le	[i]	vi n e, ir i s
[γ]	agu a , neg r o	[m]	ma m á, ámb a r	[o]	lo r o, po s
[f]	fi n o, t i fón	[n]	ne n e, jo v en	[u]	bu l ra, hurac á n

⁷ En Internet se encuentran disponibles tipos (fonts) del alfabeto fonético internacional para utilizar con los procesadores de texto [11].

Bibliografía y referencias

1. Pelton, Gordon E., "Voice Processng". MacGraw-Hill. Singapore, 1992.
2. Fletcher, Harvey, "Speech and Hearing in Communication". American Institute of Physics. Woodbury, USA, 1995.
3. Fuentes, J. L., "Gramática moderna de la lengua española". Bibliográfica Internacional. Madrid, España, 1988
4. Reccasens i Vives, Daniel, "Fonètica i Fonologia". Biblioteca Universitària 18. Enciclopèdia Catalana. Barcelona, España, 1993.
5. "Webster's New Universal Unabridged Dictionary". Barnes & Noble. New York, USA, 1992
6. Gonzalez, Mike (compilador), "Collins Concise Spanish-English English-Spanish Dictionary" Webster's New World Book. Prentice Hall Press. New York, USA, 1985.
7. Syrdal, A.; Bennet, R.; Greenspan, S. "Applied Speech Technology" CRC Press. Boca Raton, FL, USA, 1995
8. Pythoud, Christophe. "Online Phonetics Course".
Internet: <http://www.unil.ch/ling/phonetique/api-eng.html>
9. Miyara, Alberto. Comunicación personal. 15/6/99.
10. International Phonetic Association. Internet: <http://www2.arts.gla.ac.uk/IPA/ipa.html>
11. Tipos del IPA. Internet: <http://www.sil.org/computing/fonts/encore-ipa.html>