

Llisterri, J. (1996). Los sonidos del habla. In C. Martín Vide (Ed.), *Elementos de lingüística*. (pp. 67-128).
Barcelona: Octaedro.

[http://liceu.uab.cat/~joaquim/publicacions/
Llisterri_96_Sonidos_Habla.pdf](http://liceu.uab.cat/~joaquim/publicacions/Llisterri_96_Sonidos_Habla.pdf)

LOS SONIDOS DEL HABLA

por JOAQUIM LLISTERRI

INTRODUCCIÓN

El objetivo de este capítulo es describir los mecanismos implicados en la producción, transmisión y recepción de los sonidos del habla, destacando la interrelación entre las características articulatorias, acústicas y perceptivas de los mismos. Se muestra también cómo el estudio de tales mecanismos permite una clasificación de los elementos que configuran la estructura sonora del lenguaje. En la primera parte se examina el carácter interdisciplinar de la fonética, entendida como una de las posibles aproximaciones al estudio de la comunicación hablada. Los tres apartados siguientes se dedican al análisis articulatorio, acústico y perceptivo de los elementos segmentales, describiendo a continuación los elementos suprasegmentales. Para acabar, se introduce brevemente el campo de las tecnologías del habla, ámbito en el que es posible aplicar los conocimientos fonéticos al desarrollo de sistemas que faciliten la comunicación oral entre el hombre y los sistemas informáticos.

1. LA FONÉTICA Y EL ESTUDIO DE LA LENGUA

El presente apartado define la fonética y los ámbitos con los que se relaciona, resaltando su carácter interdisciplinar y su metodología propia. Se introducen los tres campos principales a partir de los cuales se organiza el resto del capítulo —fonética articulatoria, acústica y perceptiva— y las principales áreas de aplicación de la fonética. Tras una reflexión sobre el carácter esencialmente oral de la lengua, se justifica la necesidad de disponer de un sistema de representación del habla —transcripción fonética— que permita superar las inadecuaciones de la ortografía.

1.1. DEFINICIÓN DE LA FONÉTICA

EL CAMPO DE ESTUDIO DE LA FONÉTICA

Entre todas las modalidades comunicativas de que dispone la especie humana, el habla es seguramente la más natural y una de las más eficaces para la transmisión de

la información. Puesto que se trata de un fenómeno complejo, el estudio del habla suele abordarse desde diferentes perspectivas; la intersección de todas ellas constituye un campo de trabajo conocido como *ciencias del habla* (*speech sciences*), cuyo objeto de estudio es la *comunicación oral* (*speech communication*). En este ámbito confluyen diversas disciplinas, desde las ciencias cognitivas hasta la informática, incluyendo, naturalmente, la lingüística y la fonética. La interdisciplinariedad se justifica considerando que la comprensión global del papel del habla en la comunicación exige conocer las bases biológicas de los mecanismos de producción y percepción del sonido, su procesamiento cognitivo para convertirlo en información lingüísticamente relevante, y también las condiciones externas en las que se da este proceso, tanto las referentes al entorno físico como las que vienen dadas por factores culturales y sociales.

Enmarcada en este ámbito más amplio, la *fonética* tiene como objetivos determinar el modo en que los sonidos del habla se emplean con fines comunicativos en las lenguas naturales y explicar los mecanismos que condicionan tanto su producción como su percepción. Por tanto, la fonética persigue desarrollar un modelo que dé cuenta de la producción y la percepción de los sonidos del habla en el proceso de la comunicación. Naturalmente, este modelo no debe estar en contradicción con los conocimientos sobre el funcionamiento de los sistemas implicados en el procesamiento del lenguaje. Por otra parte, el modelo debe tener igualmente en cuenta los condicionantes pragmáticos y sociolingüísticos que operan sobre el acto de hablar y debería ser coherente con propuestas más globales elaboradas en el ámbito de la lingüística. No es ilógico, finalmente, pensar que tal modelo puede aportar ideas sobre cómo mejorar la producción artificial del habla, su reconocimiento y su comprensión por parte de sistemas informáticos que actualmente ya realizan estas tareas.

Todo ello explica que el estudio de la fonética no se realice aisladamente de las demás ramas del saber que aspiran a explicar y modelar la comunicación humana. Aun así, puede considerarse que la fonética constituye una disciplina autónoma, que tradicionalmente ha mantenido y sigue manteniendo estrechas relaciones con la lingüística.

LA FONÉTICA COMO CIENCIA EXPERIMENTAL Y COMO ÁMBITO DE TRABAJO INTERDISCIPLINAR

Debido a la naturaleza de su objeto de estudio, la fonética es esencialmente experimental, aunque no por ello hay que pensar que carece de una base teórica. Más bien al contrario, ya que se ha empezado definiendo su objetivo como la elaboración de un modelo —es decir, una teoría— del proceso de producción y percepción de los sonidos del habla. Al referirse a la *experimentalidad* de la fonética se quiere señalar que, a diferencia de ciertos enfoques en otras disciplinas de la lingüística, la estrategia para abordar el estudio del habla no se basa en la introspección, sino en datos obtenidos en situaciones comunicativas lo más naturales posibles, analizados con métodos que permiten un acercamiento a las propiedades físicas de la señal sonora, a los mecanismos de producción o a los de recepción; las conclusiones extraídas se basan en un número estadísticamente relevante de enunciados realizados por hablantes diferentes.

Sin embargo, debe destacarse que no es el uso de herramientas como los ordenadores o la estadística lo que confiere a la fonética un carácter experimental, sino la misma concepción de la disciplina como un campo en el que se procede a la recogida de

datos guiada por una hipótesis y en el que se valida o invalida esta hipótesis según los resultados obtenidos mediante una metodología determinada de forma rigurosa.

El fonetista debe aprender a menudo recursos de trabajo propios de otras áreas experimentales como por ejemplo la fisiología, la acústica, el procesamiento de señales, la informática, la estadística o la psicología, por lo que su campo de trabajo puede calificarse, como se ha señalado anteriormente, de *interdisciplinar*.

FONÉTICA GENERAL, FONÉTICA DE LAS LENGUAS PARTICULARES Y APLICACIONES DE LA FONÉTICA

Aunque la fonética tenga, como se acaba de ver, un objetivo unitario, pueden distinguirse diversos enfoques: puesto que existen mecanismos biológicos comunes de producción y percepción del habla, la *teoría fonética —fonética general—* se interesa por la influencia de estos factores en la configuración del inventario de sonidos de que disponen las lenguas del mundo para llevar a cabo su función comunicativa y también por elaborar el modelo de producción y percepción del habla al que se ha aludido anteriormente. En cambio, la *fonética descriptiva —fonética de las lenguas particulares—* se ocupa de estudiar cómo una determinada lengua hace uso de estos mecanismos generales y de describir el inventario de sonidos de cada lengua en términos articulatorios, acústicos y perceptivos. También existe un interés por averiguar los mecanismos de cambio en los inventarios y en las propiedades de los sonidos de las diversas lenguas, por lo que puede hablarse de *fonética diacrónica —o fonética histórica—* por oposición a la *fonética sincrónica*.

Hay que contar, además, con campos específicos de *aplicación de la fonética*, que mencionamos brevemente a continuación. Cuando el objetivo es la definición de la norma de pronunciación de una lengua dada suele hablarse de *ortología*; la definición de la norma no sólo es un aspecto importante en la estandarización de una lengua, sino que su conocimiento también forma parte de la preparación de profesionales de la expresión oral, sea en los medios de comunicación, sea en el teatro y el cine o bien en la formación de los intérpretes de conferencia.

En cambio, la *ortofonía* aplica los conocimientos de fonética a la reeducación de personas que sufren alguna alteración en sus producciones sonoras; éste es un campo de trabajo que entra de lleno dentro de la *logopedia —reeducación de personas afectadas por perturbaciones del lenguaje y del habla—* y que se relaciona igualmente con la *foniatría*, disciplina médica que trata específicamente los trastornos de la voz, y también con los estudios generales de patología del lenguaje; dentro del campo de las alteraciones de la comunicación hablada, la fonética se relaciona igualmente con la *audiología*, centrada en el estudio y el tratamiento de las alteraciones de la audición.

Es posible, por otra parte, aplicar los resultados de la investigación en fonética a la corrección de la pronunciación de lenguas extranjeras, y en este caso suele hablarse de *corrección fonética*, siendo ésta una especialidad que forma parte de la metodología de la enseñanza de lenguas extranjeras. La fonética es igualmente importante en el campo de la enseñanza de la lengua materna, especialmente en la enseñanza de la lectura y la escritura.

Existen aún otros ámbitos de aplicación de la fonética, y quizás uno de los que recientemente ha adquirido más importancia es el de la *tecnología del habla*, al que se dedica el último apartado de este capítulo. Finalmente, vale la pena señalar el uso de

los conocimientos sobre el habla en la identificación de personas a partir de sus voces; esto se lleva a cabo, por ejemplo, en el ámbito judicial y constituye una especialidad conocida como *fonética forense*.

FONÉTICA ARTICULATORIA, FONÉTICA ACÚSTICA Y FONÉTICA PERCEPTIVA

Tras definir la fonética como el estudio de los sonidos del habla utilizados con fines comunicativos en las lenguas naturales, parece lógico que al establecer divisiones en el interior de la fonética se recurra al clásico esquema de la comunicación, basado en la existencia de un emisor, de un mensaje que se transmite a través de un medio y de un receptor al que llega este mensaje. Los sonidos del habla pueden considerarse desde el punto de vista de su producción situándose en el terreno del emisor —fonética articulatoria—, de su transmisión en tanto que constituyen una señal sonora —fonética acústica—, o de su percepción, colocándose en el lugar del receptor —fonética auditiva o perceptiva. En la figura 1 se esquematiza esta triple perspectiva:

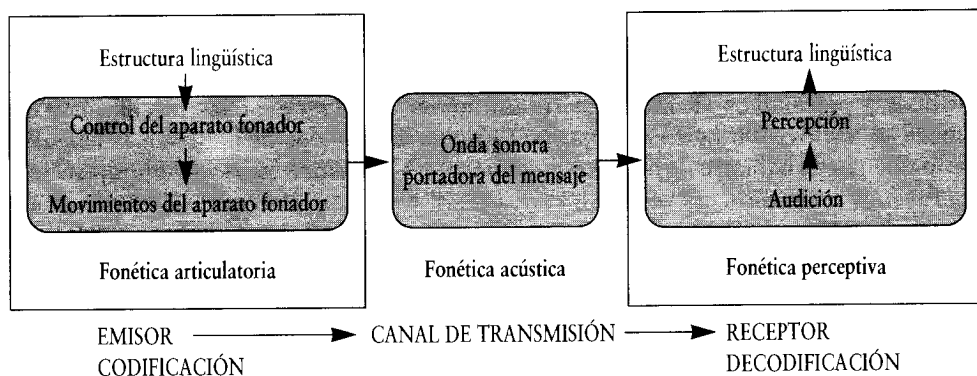


Figura 1: El proceso de la comunicación y la división de la fonética.

La *fonética articulatoria* se ocupa del estudio de la producción de los sonidos del habla, tanto de su clasificación en función de las diversas partes del aparato fonador, como de la elaboración de modelos que expliquen globalmente la producción del habla como conjunto de movimientos coordinados de los articuladores guiados por la necesidad de emitir una señal portadora de una determinada estructura lingüística.

La *fonética acústica* trata las propiedades físicas de los sonidos del habla, considerándolos en tanto que ondas sonoras transmisoras de un mensaje y relacionándolas con el modo en que se producen. El interés por la fonética acústica se ha visto aumentado con el surgimiento de las técnicas de síntesis y reconocimiento del habla, tema que se aborda al final del capítulo.

La *fonética perceptiva* tiene una doble vertiente: por una parte se ocupa del procesamiento auditivo de los sonidos del habla —*fonética auditiva*—, examinando cómo las ondas sonoras llegan hasta el oído y se transforman allí en impulsos nerviosos que

se transmiten al cerebro; por otra, debe estudiar también cómo se da una interpretación fonética a estos impulsos, asimilándolos a los sonidos de una lengua conocida y recuperando el mensaje transportado por la onda sonora, proceso que propiamente constituye la percepción. En este campo se trabaja también para proponer modelos globales de la percepción del habla que puedan relacionarse tanto con las limitaciones que impone el sistema auditivo como con la información lingüística presente en la señal sonora.

Es importante darse cuenta de que ninguna de estas tres ramas de la fonética es independiente de las otras. En apartados sucesivos se expone de qué modo las propiedades acústicas de los sonidos se explican por la configuración adoptada por el aparato fonador en el momento de su producción, estableciéndose así una estrecha relación entre la fonética articulatoria y la fonética acústica. Por otra parte, tanto la audición como la percepción de los sonidos del habla vienen determinadas por sus características acústicas, de forma que el oído puede concebirse como un sofisticado analizador de ondas sonoras, estableciéndose así una correlación entre la fonética acústica y la perceptiva. Estas dependencias se explican por el hecho de que tanto la producción como la percepción del habla están dirigidas a un mismo fin: la comunicación entre un emisor y un receptor a través de señales sonoras mediante las que se transmite un mensaje.

1.2. LA LENGUA COMO OBJETO ORAL

LENGUA ORAL Y LENGUA ESCRITA

Aunque tal como se ha planteado hasta ahora el dominio de la fonética ha quedado patente que su objeto de estudio es la comunicación hablada, es necesaria una breve reflexión sobre la concepción habitual de la lengua. Si se piensa en la importancia otorgada a la ortografía en la enseñanza, en la idea tan arraigada de que dominar la lengua equivale a un buen conocimiento del registro escrito, en la actitud ante el alfabetismo, en las épocas de la historia de la humanidad en que la escritura era privilegio de unos pocos, o en los antiguos métodos de enseñanza de lenguas basados en la traducción de textos, es fácil percatarse de que nuestra imagen de la lengua se basa en la escritura.

Sin embargo, como recuerda P. Delattre,¹ las lenguas son ante todo **objetos orales**. Esta idea viene apoyada por una serie de argumentos recogidos por J. Murillo² entre los que cabe mencionar al menos dos: por una parte, en la historia de la humanidad la lengua oral es anterior a la lengua escrita y existen aún lenguas que carecen de una forma escrita fijada, sin que se den lenguas naturales que no posean una manifestación oral; por otra, en el desarrollo del niño la lengua oral es anterior a la escrita y se adquiere muy difícilmente pasada una determinada edad crítica, mientras que la escritura puede aprenderse en cualquier momento.

1. «Sur les origines celtiques de la prononciation française», *Le bayou* 62 (1945): 316-318; en *Studies in French and comparative phonetics*, pp. 215-217 (p. 216), Mouton, La Haya, 1966.

2. «Langue orale, langue écrite» en M. Pendants (ed.), *L'écrit*, pp. 23-28, Barcelona: Institut de Ciències de l'Educació, Universitat Autònoma de Barcelona, 1979.

LA TRANSCRIPCIÓN FONÉTICA

Las relaciones entre sonido y grafía

Puesto que, como se ha indicado, el objeto de estudio de la fonética es la lengua hablada, es preciso disponer de un instrumento para representarla. Para ello es esencial distinguir desde el principio entre la *representación ortográfica* y la *representación fonética* de la lengua. Como es sabido, la ortografía proporciona las reglas que permiten asociar sonidos a grafías; en determinadas lenguas, como por ejemplo el castellano, la mayoría de sonidos tienen una representación escrita única, mientras que en otras como el inglés las relaciones entre los sonidos y las grafías que los representan son muy complejas. En conjunto, pueden establecerse cinco tipos de correspondencia entre sonido y *grafema*, entendiendo éste como la unidad de representación gráfica de la lengua:

- a un sonido le corresponde un único grafema y viceversa:
 - ◆ el grafema <l> y el sonido [l] (como en «la») en castellano.
- un mismo grafema representa dos sonidos:
 - ◆ el grafema <c> representa a los sonidos [k] (como en «casa») y [θ] (como en «cine») en castellano.
- un mismo sonido puede representarse mediante dos grafías:
 - ◆ el sonido [x] puede representarse con los grafemas <j> (como en «jamón») y <g> (como en «genio») en castellano.
- dos grafías representan un único sonido:
 - ◆ los grafemas <gu> representan el sonido [g] (como en «guerra») en castellano.
- una grafía no representa ningún sonido:
 - ◆ el caso del grafema <h> en castellano.

Justamente a causa de esta falta de biunivocidad entre la representación gráfica y la realización fonética es necesario un sistema que nos permita reflejar sin ningún tipo de ambigüedades la realidad de la lengua oral. Este sistema se conoce con el nombre de *alfabeto fonético*, y se utiliza para la representación simbólica de los elementos segmentales y suprasegmentales del habla en la *transcripción fonética*.

La transcripción fonética

Disponer de un procedimiento de transcripción fonética es útil en todos los casos en que se requiere representar la lengua hablada. En dialectología puede fijarse la representación fonética sin las ambigüedades de la grafía; en los diccionarios monolingües y bilingües es útil reflejar la pronunciación estándar de cada palabra junto con sus variantes; al estudiar las patologías del habla es preciso reproducir las realizaciones al-

teradas para las que la ortografía no ofrece a menudo alternativas; y, en conjunto, la transcripción fonética puede emplearse en cualquier análisis lingüístico en el que se requiera disponer de una representación lo más cercana posible a las realizaciones sonoras de la lengua.

Sin embargo, la transcripción fonética presenta también limitaciones: en primer lugar, es extraordinariamente difícil transcribir una lengua que no se conoce debido a las dificultades de segmentar la señal sonora en elementos discretos, es decir, separados unos de otros; por otra, un mismo símbolo puede presentar realizaciones fonéticas diferentes en lenguas distintas, por lo que disponer de una transcripción no garantiza una pronunciación completamente nativa de una lengua desconocida.

Los alfabetos fonéticos

Aun considerando las limitaciones señaladas, en fonética y en lingüística es imprescindible disponer de un *alfabeto fonético*. Su característica principal debe ser la eliminación de las arbitrariedades que se dan en las lenguas naturales en cuanto a la relación entre grafemas y sonidos. Por lo tanto, el principio básico de cualquier alfabeto fonético es que cada sonido debe transcribirse mediante un único símbolo y cada símbolo no debe representar a más de un sonido; este principio debe mantenerse de modo que el conjunto de símbolos permita representar los sonidos de más de una lengua.

El sistema de transcripción fonética más utilizado hoy en día por la comunidad científica es el AFI o Alfabeto Fonético Internacional (IPA, *International Phonetic Alphabet*); sin embargo, en el mundo hispánico se ha utilizado predominantemente el alfabeto fonético de la *Revista de Filología Española (RFE)* propuesto por T. Navarro Tomás,³ por lo que el dominio de la bibliografía sobre el español presupone el conocimiento de este sistema.

El Alfabeto Fonético Internacional

El *Alfabeto Fonético Internacional* es el fruto de los esfuerzos de un grupo de profesores que, en 1886, fundaron en Europa una asociación para la reforma de la didáctica de las lenguas modernas basada en la necesidad de enseñar la lengua hablada. En 1897 nació de este mismo grupo la Asociación Fonética Internacional (IPA, *International Phonetic Association*), responsable del desarrollo y la forma actual del AFI. La última revisión de este sistema de transcripción se llevó a cabo en Kiel en 1989, bajo los auspicios de la Asociación.⁴

El AFI pretende ser un conjunto de símbolos que representan los posibles sonidos de las lenguas del mundo, de modo que pueda utilizarse en todos aquellos casos en que es necesario representar la estructura fonética o fonológica de la lengua. Sin embargo,

3. Alfabeto Fonético de la RFE, *Revista de Filología Española*, II (1915): 74-76.

4. IPA (1949), *The principles of the International Phonetic Association*. Londres: International Society of Phonetic Sciences. Versión revisada: *Journal of the International Phonetic Association*, 23,1 (1993).

únicamente se emplean símbolos diferentes para representar aquellos sonidos que son distintivos —es decir, que sirven para diferenciar palabras— en las lenguas; como se indica más adelante, las propiedades fonéticas que no son distintivas pueden transcribirse mediante el uso de diacríticos que acompañan a los símbolos. De esta forma se consigue reducir el inventario de símbolos manteniendo la posibilidad de representar características fonéticas detalladas.

Puesto que la clasificación de los símbolos se lleva a cabo atendiendo a las características del modo de producción de los sonidos que representan, el Alfabeto Fonético Internacional se describe tras haber presentado las nociones necesarias de fonética articuladora en el próximo apartado.

2. LA PRODUCCIÓN DE LOS SONIDOS DEL HABLA

La producción de los sonidos del habla puede concebirse como un proceso en el cual una representación lingüística —el mensaje— se convierte en una onda sonora mediante la acción del aparato fonador. Éste realiza un conjunto coordinado de movimientos que tienen como finalidad la producción de una señal portadora de toda la información necesaria para que el receptor interprete el mensaje inicialmente pensado por el emisor.

En este apartado se presenta en primer lugar el funcionamiento del aparato fonador en la producción del habla, estableciendo las nociones que permiten una clasificación de los sonidos de las lenguas desde el punto de vista articulatorio. Tal clasificación se refleja en el Alfabeto Fonético Internacional, descrito al final del apartado.

2.1. EL PROCESO DE PRODUCCIÓN

El conjunto de los órganos implicados en la producción del habla suele denominarse *aparato fonador*. Es interesante darse cuenta, en primer lugar, de que en el aparato fonador no hay órganos específicamente dedicados a la producción del habla, sino que consta del mismo conjunto de órganos que intervienen en la alimentación o la respiración. Sin embargo, no debe olvidarse que, en última instancia, y aunque aquí no aparece ninguna referencia al tema, todo el proceso de producción está controlado por determinadas áreas especializadas del cerebro.

En conjunto el aparato fonador puede dividirse en tres partes, tal como se esquematiza en la figura 2:

- los *mecanismos subglóticos*, que proporcionan el aire necesario para la producción del sonido,
- la *laringe*, responsable de la fonación o producción de la voz, y
- las *cavidades supraglóticas*, en las que se lleva a cabo la articulación.

El conjunto de la laringe y las cavidades supraglóticas suele denominarse *tracto vocal*.

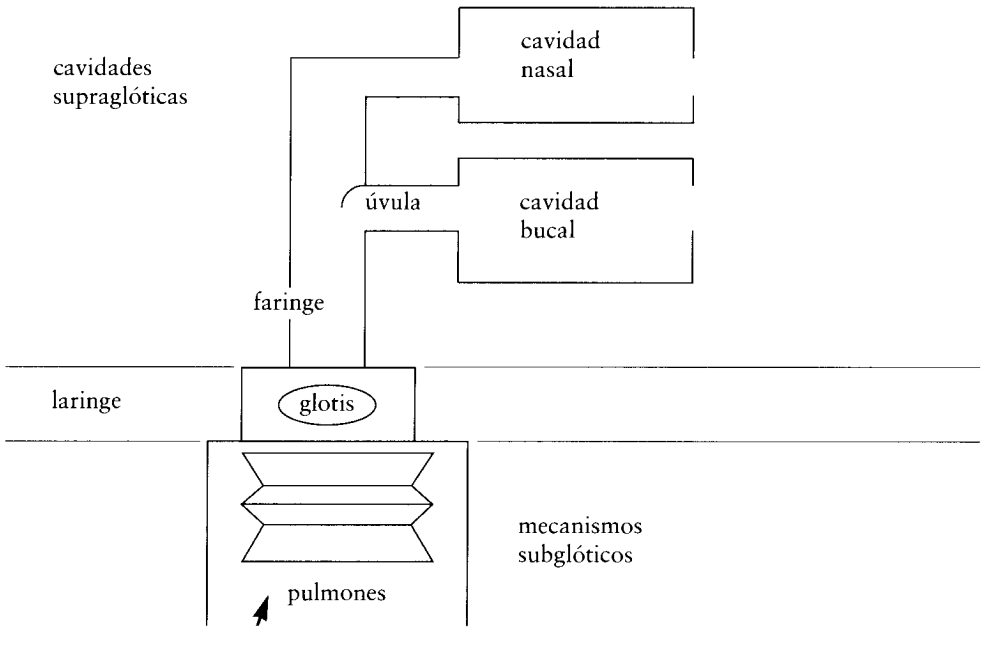


Figura 2: Representación esquematizada del aparato fonador.

El proceso de producción de los sonidos del habla se inicia con la *respiración*. Los pulmones son responsables de la creación de una corriente de aire que atraviesa la laringe y que, a su paso por la glotis, pone en vibración las cuerdas vocales si se encuentran juntas, dando lugar a la *fonación* o producción de la voz. A continuación la corriente de aire llega a las cavidades supraglóticas y, según la posición de la úvula, se reparte entre la cavidad nasal y la bucal o se concentra únicamente en la cavidad bucal, donde es modificada mediante el proceso conocido como *articulación*. En términos físicos el aparato fonador constaría esencialmente de un generador de energía —los pulmones—, un sistema vibratorio —la laringe— y una cavidad de resonancia —las cavidades supraglóticas.

LOS MECANISMOS SUBGLÓTICOS Y LA RESPIRACIÓN

Para la producción de los sonidos del habla es imprescindible la existencia de una corriente de aire. Los *pulmones* son los responsables de producirla, con la acción conjunta de los músculos implicados en la respiración.

El proceso de la *respiración* puede explicarse como una serie de aumentos y disminuciones sucesivas de la presión del aire contenido en los pulmones, de la que son responsables la musculatura torácica y la abdominal. Cuando aumenta la presión en el interior de la caja torácica, se produce una salida de aire, mientras que, cuando ésta disminuye, el aire entra en los pulmones.

La entrada de aire tiene lugar gracias a un cambio de posición del *diafragma* —un

músculo en la base del tórax que lo separa del abdomen—, el cual baja y adquiere una posición plana; con ello aumenta el volumen de la caja torácica, a lo que contribuye también la acción de los *músculos intercostales*, situados entre las costillas. El resultado es una disminución de la presión del aire en el interior del tórax, que produce la entrada de aire del exterior o *inspiración*. En la *expiración* el aire sale a consecuencia de la relajación del diafragma y de los intercostales que, volviendo a su posición normal, disminuyen el volumen de la caja torácica aumentando al mismo tiempo la presión del aire en el interior de tórax y provocando su salida. La figura 3 muestra la posición relajada de las estructuras respiratorias (a) y los cambios que se producen tanto en la caja torácica como en la pared abdominal y en el diafragma durante la inspiración (b).

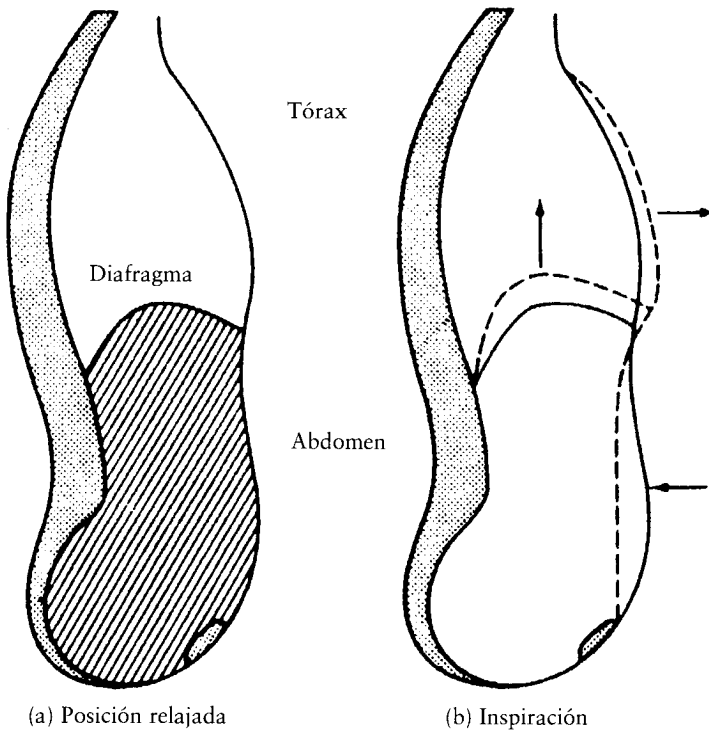


Figura 3: Representación esquematizada de los cambios producidos en la caja torácica, la pared abdominal y el diafragma en posición relajada (a) y en la inspiración (b) (de T. J. Hixon (1973), «Respiratory function in speech», en F. D. Minifie, T. J. Hixon & F. Williams (eds.), *Normal aspects of speech, hearing and language*, p. 110. Prentice-Hall, Englewood Cliffs, N. J.).

Durante la producción del habla se debe llevar a cabo un control de la respiración, proceso que normalmente se realiza inconscientemente y ligado a la necesidad de oxigenar la sangre. Al hablar, las inspiraciones son rápidas y breves, mientras que el tiempo de espiración es mayor, debiéndose controlar la velocidad y la presión de salida del aire mediante el diafragma y los demás músculos implicados en la respiración.

Los mecanismos utilizados para la salida de la corriente de aire dan lugar a una clasificación de los sonidos del habla que se estudia en el apartado 2.2.

LA FONACIÓN

La *fonación* es el proceso mediante el cual se produce la voz, aprovechando el aire que procede de los pulmones. Tiene lugar en la laringe, y en él intervienen principalmente las cuerdas vocales.

La *laringe* es un conjunto de cartílagos y un hueso, denominado hioides, situado en la base de la lengua. Uno de los cartílagos más importantes es el tiroideo, formado por dos láminas que forman un ángulo más agudo en el hombre que en la mujer, provocando la prominencia conocida como manzana de Adán o nuez. En la cavidad formada por esta estructura cartilaginosa se encuentran dos pliegues musculares recubiertos de una mucosa que reciben el nombre de *cuerdas vocales*. La denominación de «cuerdas» proviene de la imagen de la laringe que se obtuvo en las primeras observaciones realizadas ya en el siglo XVIII con la ayuda de un espejo, aunque en realidad constituyen dos masas musculares situadas a ambos lados de la laringe. El espacio situado entre los dos pliegues recibe el nombre de *glotis*. La figura 4 muestra esquemáticamente la estructura de la laringe desde diversas perspectivas.

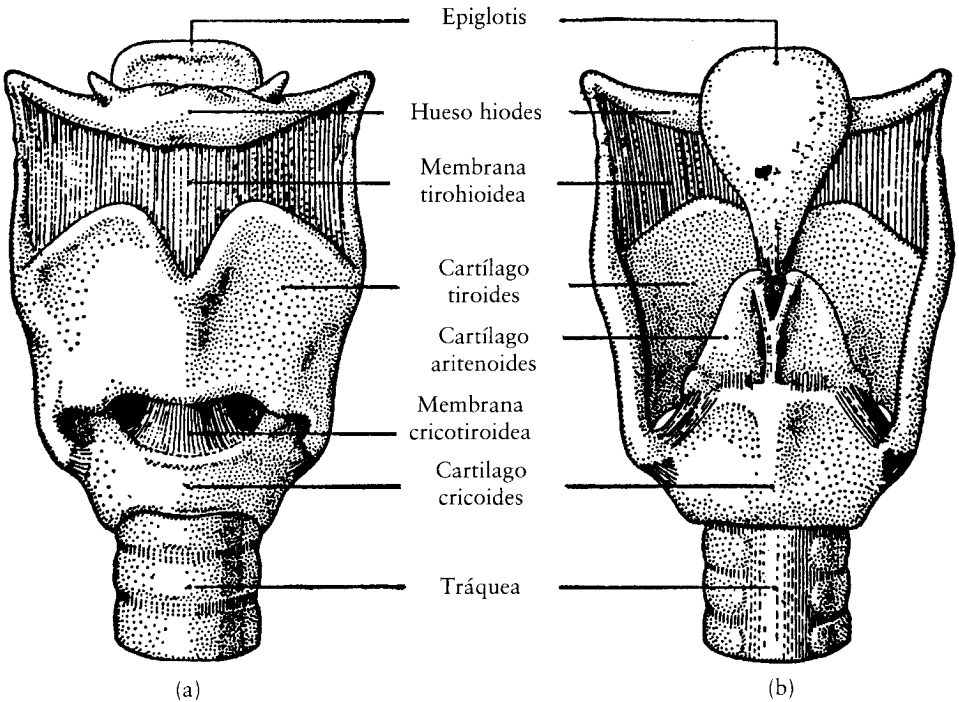
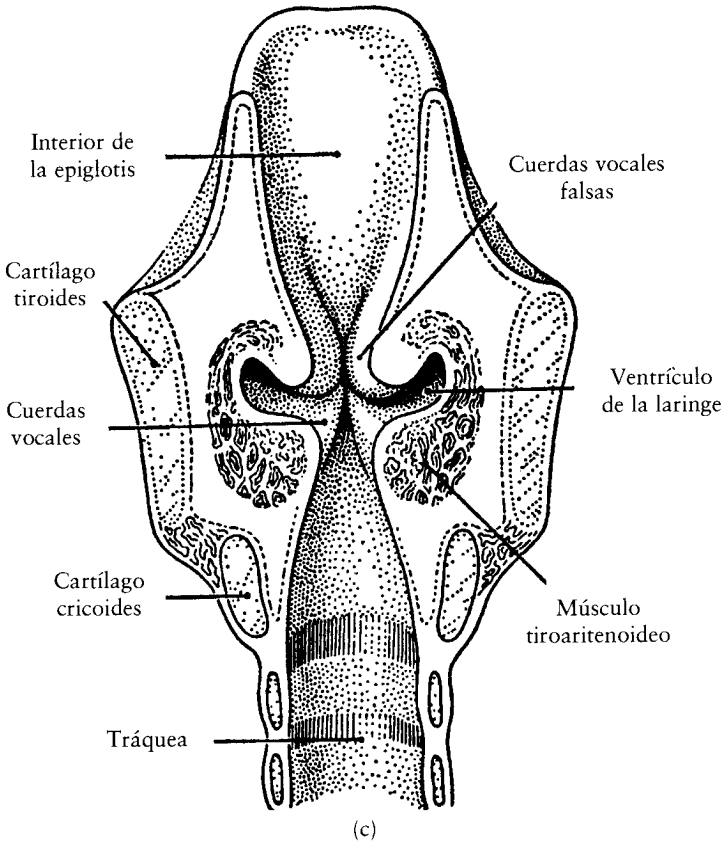


Figura 4: Representación esquematizada de la estructura de la laringe: (a) visión anterior; (b) visión posterior; (c) sección frontal (de G. J. Borden & K. S. Harris (1980), *Speech science primer. Physiology, acoustics and perception of speech*, p. 76. Williams & Wilkins, Baltimore).



La *voz* es el resultado de la vibración de las moléculas de aire provocada por el movimiento de las cuerdas vocales cuando la corriente de aire procedente de los pulmones pasa a través de la glotis. Si la glotis se encuentra abierta —es decir, las cuerdas vocales están separadas— el aire puede salir libremente. Si, por el contrario, las cuerdas vocales se encuentran juntas debido a la acción de los cartílagos *aritenoides* que, entre otros, controlan sus movimientos, la presión del aire expulsado por los pulmones tiende a abrirlas de modo que el aire pueda salir al exterior; sin embargo, una vez ha pasado una cierta cantidad de aire, la propia elasticidad del músculo que forma las cuerdas vocales tiende a cerrarlas de nuevo; al cierre de las cuerdas vocales contribuye también el denominado efecto Bernoulli, consistente en la succión resultado del descenso de la presión que se produce cuando la corriente de aire aumenta su velocidad al atravesar un espacio estrecho como el que queda en la glotis entre los dos pliegues vocálicos. Sin embargo, como el aire sigue ascendiendo desde los pulmones, se produce una presión suficiente para volver a separar las cuerdas vocales y dejar pasar nuevamente una cierta cantidad de aire, tras lo cual la glotis vuelve a cerrarse por la acción combinada de los dos elementos mencionados. Este proceso tiene lugar repetidamente —a una velocidad de unas cien veces por segundo en el hombre—, de modo que el re-

sultado son sucesivas bocanadas de aire que salen de la glotis como consecuencia del ciclo de abertura y cierre de las cuerdas vocales. La figura 5 presenta esquemáticamente el ciclo vibratorio de las cuerdas vocales.

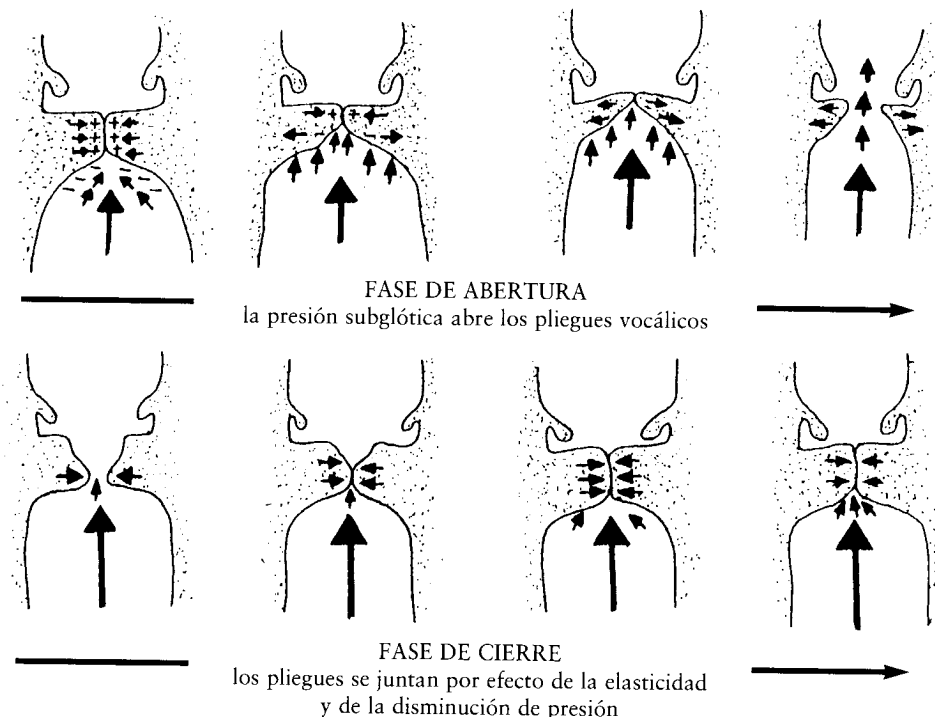


Figura 5: Representación esquematizada del ciclo vibratorio de las cuerdas vocales (de C. R. Schneiderman (1984), *Basic anatomy and physiology in speech and hearing*, p. 76. College-Hill Press, San Diego).

En el apartado 3 se muestran las características acústicas del resultado de la vibración de las cuerdas vocales. Puesto que para producir la voz se combina la elasticidad muscular y el efecto de un descenso de la presión del aire en la glotis, la teoría que explica el movimiento de las cuerdas vocales se conoce como *teoría mioelástica-aerodinámica* de la fonación.

La posición de la glotis y los ajustes de las cuerdas vocales que se realizan gracias a la acción de los cartílagos cricoides y aritenoides dan lugar a diversos tipos de fonación que se explican en el apartado 2.2.

LA ARTICULACIÓN

La *articulación* es el proceso por el cual se modifica la corriente de aire procedente de los pulmones y la laringe en las cavidades supraglóticas como consecuencia de los

cambios de volumen y de forma de estas cavidades. Tal como se observa en la representación de la figura 6, el conjunto de las cavidades supraglóticas puede dividirse en tres partes: la faringe, la cavidad bucal y la cavidad nasal.

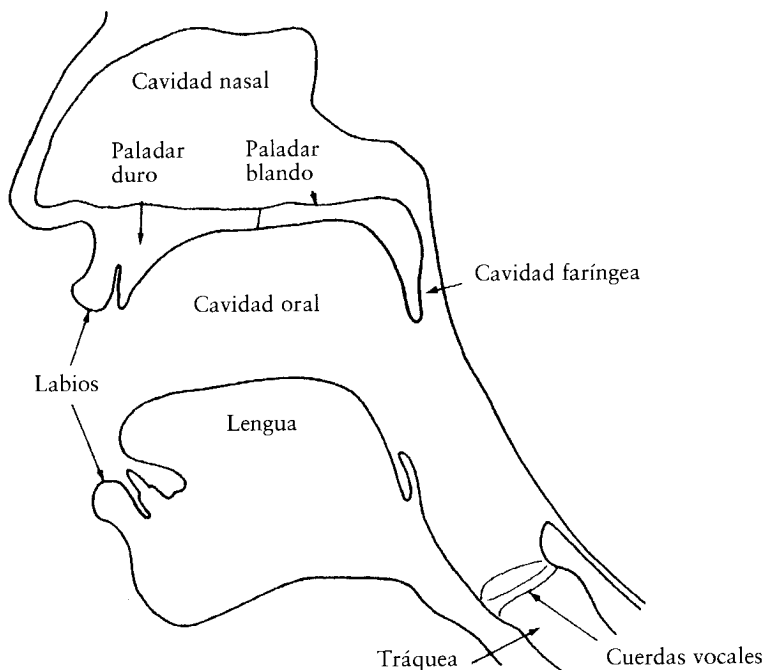


Figura 6: Representación esquematizada de las cavidades supraglóticas (de R. G. Daniloff [1973] «Normal articulation processes», en F. D. Minifie, T. J. Hixon & F. Williams [eds.], *Normal aspects of speech, hearing and language*, p. 170. Prentice-Hall, Englewood Cliffs, N. J.).

LA FARINGE

La *faringe* es una cavidad situada entre la laringe y el velo del paladar, formada por paredes musculares flexibles; la forma y el volumen de esta cavidad puede alterarse en la producción de diversas clases de sonidos.

LA CAVIDAD BUCAL

La *cavidad bucal* consta de partes fijas —como la mandíbula superior— y de partes móviles —como por ejemplo la lengua o la mandíbula inferior—, que son responsables de algunas de las principales modificaciones que sufre la corriente de aire procedente de la laringe. Está delimitada por la mandíbula superior y la inferior y por la pared faríngea y los labios en los extremos.

En la parte superior de la cavidad bucal se encuentra el paladar, en el que se distingue la *úvula* —apéndice móvil del paladar blando que abre y cierra el paso del aire hacia la cavidad nasal—, el *paladar blando* o *velo del paladar* y el *paladar duro*, en el que puede distinguirse una zona *prepalatal* y una *postpalatal*. A continuación se hallan los *alveolos*, una prominencia situada entre el paladar duro y la parte posterior de los *incisivos*. Las dos mandíbulas sirven de apoyo a los *dientes*, divididos en incisivos, caninos, premolares y molares. En la parte exterior de la cavidad bucal se encuentra el *labio inferior* y el *labio superior*. Finalmente, la *lengua* es un conjunto de músculos dividido habitualmente en raíz —la parte posterior—, *dorso* —la parte central, en la que puede distinguirse el *predorso* en la parte anterior y el *posdorso* en la parte posterior—, *lámينا* y el *ápice* o extremo. La figura 7 esquematiza las principales partes de la cavidad bucal. Como se observará en el apartado 2.2., la clasificación articulatoria de los sonidos se lleva a cabo mediante la definición de las partes fijas y móviles de la cavidad bucal que intervienen en su producción.



Figura 7: Representación esquematizada de la cavidad bucal (de P. Ladefoged [1975] *A course in phonetics*, p. 17. Harcourt, Brace, Jovanovich, Nueva York, 1993, 3.ª ed.).

LA CAVIDAD NASAL

La *cavidad nasal* está constituida por dos tubos que conectan la faringe a las fosas nasales; según la posición de la úvula el aire puede o no llegar a la cavidad nasal, que se convierte en una cavidad adicional de resonancia. Aunque su tamaño y forma sea diferente en cada persona, no es una cavidad cuyas características puedan alterarse du-

rante la articulación. Como veremos, el paso de aire por la cavidad nasal es también uno de los criterios en la clasificación articulatoria de los sonidos.

LA REPRESENTACIÓN ARTICULATORIA DE LOS SONIDOS DEL HABLA

La manera tradicional de representar la articulación de los sonidos del habla es mediante un *corte sagital* en el que se reproduce la posición de los órganos implicados en la producción de un segmento, tal como se muestra en la figura 8.

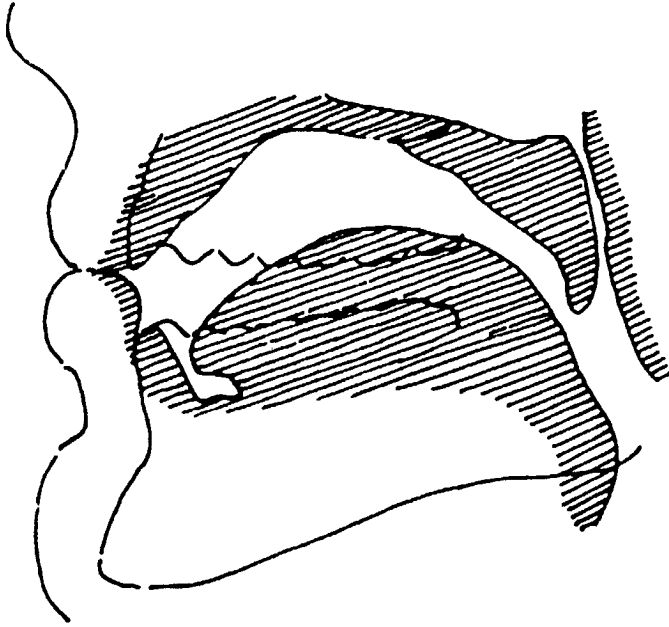


Figura 8: Corte sagital representando la realización de la consonante nasal bilabial [m] del español (de A. Quilis & J. A. Fernández [1964], *Curso de fonética y fonología españolas para estudiantes angloamericanos*, p. 112. Consejo Superior de Investigaciones Científicas, Madrid, 1982, 10.ª ed.).

En las primeras etapas de la fonética estas representaciones se basaban en la propia intuición de los investigadores sobre la posición del aparato fonatorio. Posteriormente se llevaron a cabo a partir de esquematizaciones de radiografías, y más tarde se impuso la *cinerradiografía*, que permite la filmación del tracto vocal en movimiento. Otros métodos como la *palatografía* hacen posible visualizar el contacto entre la lengua y el paladar, y en este caso también se ha pasado de las representaciones estáticas a técnicas como la *electropalatografía* dinámica, mediante las que puede observarse la evolución del contacto linguopalatal en una secuencia; un ejemplo de los resultados obtenidos se muestra en la figura 9.

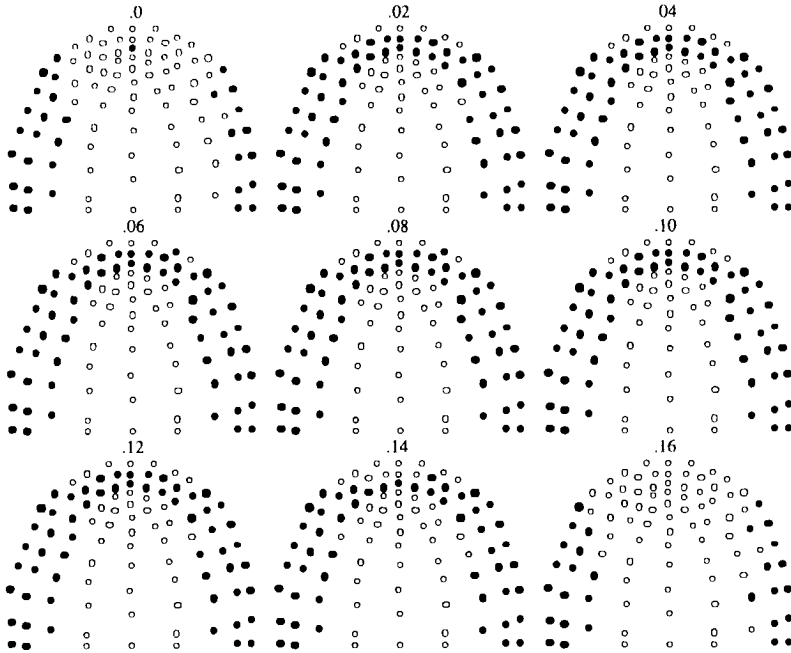


Figura 9: Secuencia electropalatográfica mostrando —de izquierda a derecha y de arriba abajo— la evolución del contacto linguoalveolar en la consonante [n] del inglés en el contexto [ana]. Los puntos marcados en negro indican la zona de contacto, y la diferencia temporal entre cada secuencia es de 20 ms. (de D. Byrd [1994], «Palatogram reading as a phonetic skill: a short tutorial», *Journal of the International Phonetic Association*, 24(1): 112).

Existen también métodos para el estudio de la respiración y del flujo de salida del aire en el habla, para la visualización y filmación del movimiento de las cuerdas vocales, y sistemas que combinan la información procedente de varias fuentes a fin de obtener información lo más completa posible sobre el proceso de producción.

2.2. LA CLASIFICACIÓN ARTICULATORIA DE LOS SONIDOS DEL HABLA

La *clasificación articulatoria* de los sonidos del habla se lleva a cabo, como se ha señalado anteriormente, partiendo de la configuración que adopta cada una de las partes en las que se divide el aparato fonador. Por ello, siguiendo a Ladefoged (1971, 1975), pueden establecerse cuatro criterios de clasificación: en función de la utilización de la corriente de aire producida por los mecanismos subglóticos, del estado de la glotis, de la configuración de la cavidad oral y nasal y del lugar del tracto vocal en el que se produce la articulación.

Antes de abordar la clasificación articulatoria hay que tener en cuenta que en la producción de muchos sonidos se produce una *constricción*, es decir, un acercamiento entre

dos partes de las cavidades supraglóticas —por ejemplo entre el dorso de la lengua y el paladar duro o entre el ápice de la lengua y los alveolos— que puede dar lugar a un cierre total del paso del aire, en cuyo caso se habla de *oclusión*, o a un estrechamiento del canal del paso del aire, con lo que se produce una *fricción*. Cuando tiene lugar un cierre total, la presión del aire que llega de los pulmones y la laringe aumenta progresivamente, de modo que al cabo de un cierto tiempo el canal se abre produciéndose como resultado una *explosión*. Este mecanismo básico de cierre o estrechamiento y abertura es el que se utiliza en la producción de la mayoría de los sonidos consonánticos.

CLASIFICACIÓN ARTICULATORIA DE LOS SONIDOS EN FUNCIÓN DE LA CORRIENTE DE AIRE

Para establecer una clasificación articulatória de los sonidos del habla en función de la corriente de aire producida por los pulmones se consideran dos elementos: por una parte, el lugar donde se produce la corriente de aire o el lugar en el que esta corriente se ve interrumpida; por otra, el mecanismo de entrada y salida del aire. Con estos criterios es posible definir las cuatro clases de sonidos que se presentan en la tabla 1:

Producción / obstrucción de la corriente de aire	Mecanismo de entrada / salida del aire	Descripción	Clasificación
Pulmonar (corriente de aire producida por los pulmones y obstrucción en el tracto vocal)	Egresivo (con salida del aire)	El aire sale de los pulmones controlado por los mecanismos respiratorios.	Oclusivo
Glotal (obstrucción producida en la glotis)	Egresivo (con salida del aire)	La glotis se encuentra cerrada y se produce una elevación de la mandíbula inferior que comprime el aire de la cavidad que queda entre la glotis y el lugar de constricción; la abertura de la glotis suele ir precedida de una explosión.	Eyectivo
Glotal (obstrucción producida en la glotis)	Ingresivo (con entrada del aire)	Además de producirse una constricción en la cavidad bucal se produce un movimiento de descenso de la laringe y de la mandíbula inferior con la glotis cerrada.	Implosivo
Velar (obstrucción producida en el velo del paladar)	Ingresivo (entrada del aire)	Se producen dos constricciones simultáneas en la cavidad bucal: una entre el dorso de la lengua y el velo del paladar y otra en el lugar de articulación del sonido; con ello tiene lugar una rarefacción que produce una entrada de aire en el momento de la explosión.	Clics

Tabla 1: Clasificación articulatória de los sonidos del habla en función de la corriente de aire (de Ladefoged, 1971, 1975).

CLASIFICACIÓN ARTICULATORIA DE LOS SONIDOS
EN FUNCIÓN DEL ESTADO DE LA GLOTIS

La clasificación articuladora de los sonidos en función del estado de la glotis y de los diversos ajustes de las cuerdas vocales puede resumirse, siguiendo a Ladefoged (1971), tal como aparece en la tabla 2:

Estado de la glotis	Clasificación
Glotis abierta, sin vibración de las cuerdas vocales.	Sordo
Glotis cerrada, con vibración de las cuerdas vocales, que se mantienen juntas por la acción de los cartílagos aritenoides.	Sonoro
Glotis abierta durante e inmediatamente después de la constricción en las cavidades supraglóticas.	Aspirado
La glotis se cierra inmediatamente después de la explosión en las cavidades supraglóticas.	No aspirado
Una parte de las cuerdas vocales se encuentra separada por la acción de los cartílagos aritenoides mientras que la otra se encuentra junta, de manera que durante la vibración de las cuerdas vocales se escapa una gran cantidad de aire.	Murmurado (<i>Breathy</i>)
Las cuerdas vocales se encuentran cerradas por acción de los cartílagos aritenoides, excepto una pequeña parte que se encuentra en vibración.	Laringalizado (<i>Creaky</i>)
Las cuerdas vocales se mantienen juntas o muy poco abiertas con la excepción de la parte que queda entre los cartílagos aritenoides.	Susurrado (<i>Whisper</i>)
Las cuerdas vocales se mantienen juntas impidiendo el paso del aire y dejándolo salir bruscamente.	Oclusión glotal

Tabla 2: Clasificación articuladora de los sonidos del habla en función del estado de la glotis (de Ladefoged, 1971, 1975).

Algunas de las posibilidades mencionadas se esquematizan en la figura 10, en la que las zonas sombreadas indican la glotis abierta y el movimiento de las cuerdas vocales se indica mediante flechas horizontales.

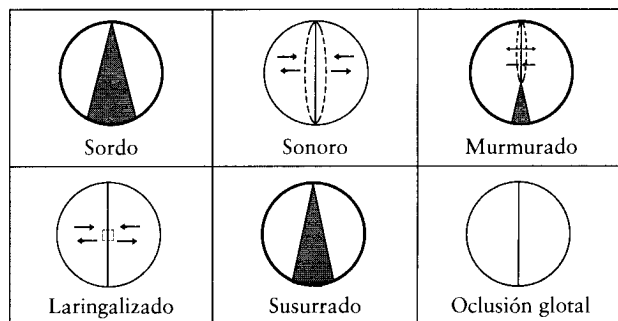


Figura 10: Esquemización del estado de la glotis en diversos tipos de fonación (de R. Daniloff, G. Schuckers & L. Feth (1980), *The physiology of speech and hearing*, p. 174. Prentice-Hall, Englewood Cliffs, N.J.).

CLASIFICACIÓN ARTICULATORIA DE LOS SONIDOS EN FUNCIÓN DE LOS PROCESOS ORONASALES

Una tercera posibilidad de clasificación articulatória reside en el modo como la corriente de aire se distribuye en las cavidades supraglóticas en función de las características del canal por el que se realiza su salida al exterior. Este tipo de clasificación responde a lo que tradicionalmente se conoce como *modo de articulación*. Se resume a continuación en la tabla 3:

Procesos oronasales	Clasificación
El velo del paladar se eleva provocando que la úvula obstruya el paso del aire hacia la cavidad nasal	Oral
El velo del paladar desciende y el aire puede pasar libremente hacia la cavidad nasal	Nasal
Cierre completo del paso del aire	Oclusivo (<i>Stop</i>)
Estrechamiento del canal de paso del aire de modo que se produce una turbulencia	Fricativo
Cierre completo del paso del aire seguido de un estrechamiento del canal que produce una turbulencia	Africado
Aproximación de dos articuladores sin que se llegue a dar un estrechamiento tal que se produzca una turbulencia	Aproximante
Vibración repetida de un articulador contra otro	Vibrante múltiple (<i>Trill</i>)
Contacto momentáneo de un articulador móvil con uno fijo	Vibrante simple (<i>Tap o Flap</i>)
Salida del aire por ambos lados de la cavidad bucal	Lateral
Salida del aire por el centro de la cavidad bucal	Central

Tabla 3: Clasificación articulatória de los sonidos del habla en función del modo de articulación (de Ladefoged, 1971, 1975).

CLASIFICACIÓN ARTICULATORIA EN FUNCIÓN DEL LUGAR DE ARTICULACIÓN

Tal como se ha mencionado, la articulación se produce normalmente por la aproximación o el contacto entre un articulador fijo y uno móvil. Según la zona en la que tenga lugar este contacto se establece una clasificación de los sonidos en función del *lugar* —o *punto*— *de articulación*, resumida en la tabla 4:

Lugar de articulación y contacto entre los articuladores	Clasificación
Labio superior y labio inferior	Bilabial
Labio inferior e incisivos superiores	Labiodental
Ápice o dorso de la lengua y la parte posterior de los incisivos superiores	Dental
Ápice de la lengua situado entre los incisivos inferiores y superiores	Interdental
Elevación del ápice de la lengua hacia la parte posterior de los alvéolos	Retroflejo
Parte anterior de la lengua y alvéolos	Alveolar
Parte anterior del dorso de la lengua y paladar duro	Palatal

Parte posterior de la lengua y velo del paladar	Velar
Parte posterior de la lengua y úvula	Uvular
Raíz de la lengua y pared faríngea	Faríngeo
Cierre de la glotis	Glotal

Tabla 4: Clasificación articulatoria de los sonidos del habla en función del lugar de articulación (de Ladefoged, 1971, 1975).

La combinación entre el modo y el lugar de articulación es la manera más habitual de describir articulatoriamente los sonidos del habla. Así, se habla de sonidos fricativos palatales para indicar que se produce un estrechamiento en el canal del paso del aire entre el dorso de la lengua y el paladar duro, o de aproximantes velares para señalar que se da una aproximación entre la parte posterior de la lengua y el velo del paladar sin que llegue a producirse una fricción. En determinadas ocasiones hay autores que prefieren precisar el lugar de articulación mencionando los dos articuladores: se habla por ello de ápico-dentales, indicando que es justamente el ápice de la lengua el que entra en contacto con la parte posterior de los incisivos, o de dorsopalatales, a fin de mostrar que es el dorso de la lengua la parte que entra en contacto o se aproxima al paladar. La descripción articulatoria se completa con la mención del estado de la glotis, de modo que un sonido puede definirse, por ejemplo, como fricativo palatal sordo o, con mayor precisión, como fricativo dorsopalatal sordo. Cabría añadir aún su carácter oral o nasal y la salida lateral o central del aire, aunque en la práctica sólo se especifique el primer parámetro en las nasales y el segundo en las laterales.

LA CLASIFICACIÓN ARTICULATORIA DE LAS VOCALES

La clasificación articulatoria de las vocales se realiza mediante tres características: la configuración de los labios —que determina la diferencia entre *redondeadas* y *no redondeadas*—, la posición del punto de máxima constricción entre la lengua y la parte superior de la cavidad bucal —distinguiéndose las vocales *anteriores*, *centrales* y *posteriores*— y la distancia en el punto de máxima constricción entre la lengua y la parte superior de la cavidad bucal, que divide las vocales en *abiertas* y *cerradas*, con la posibilidad de establecer grados intermedios.

CARACTERÍSTICAS ARTICULATORIAS SECUNDARIAS

Por influencia de los sonidos precedentes o siguientes se producen modificaciones adicionales en la articulación de un sonido, fenómeno que se conoce como *coarticulación*. Así, una consonante no labial puede labializarse si está en contacto con una vocal redondeada como [o] o [u], o una consonante alveolar como [n] puede dentalizarse en contacto con una dental como [t]. El resultado de estas modificaciones suele ser una asimilación. En la tabla 5 se resumen las principales modificaciones que afectan a la articulación de un sonido del habla:

Modificación	Resultado
Redondeamiento o protrusión labial	Labialización
Elevación de la parte anterior de la lengua	Palatalización
Elevación de la parte posterior de la lengua	Velarización
Desplazamiento hacia la parte posterior de la raíz de la lengua	Faringalización
Salida del aire por la cavidad nasal	Nasalización

Tabla 5: Características articulatorias secundarias (de Ladefoged, 1971, 1975).

2.3. EL ALFABETO FONÉTICO INTERNACIONAL

En la figura 11 se presenta el cuadro actual de símbolos del *Alfabeto Fonético Internacional*. Como puede observarse, cada consonante se define por el lugar y el modo de articulación y por la sonoridad; se distinguen diez modos de articulación —aunque eyectivo e implosivo corresponden estrictamente hablando a modos de salida del aire— y once lugares; la sonoridad se indica colocando el símbolo a la izquierda de la casilla si éste representa una consonante sorda o a la derecha si representa una sonora. Las combinaciones de lugar y modo consideradas imposibles se señalan con recuadros sombreados. Con esta tabla es sencillo saber qué símbolo debe usarse para representar una determinada consonante: basta buscar la intersección entre el modo y el lugar de articulación que la define.

Respecto al modo de articulación africado, el AFI recurre al uso de una ligadura juntando el símbolo que representa la oclusiva al que representa la fricativa. El AFI incluye además una lista de símbolos adicionales que pueden utilizarse para la representación de los sonidos no incluidos en el cuadro principal.

CONSONANTES (PULMONARES)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retrofleja	Palatal	Velar	Uvular	Faringea	Glotal
Oclusiva	p b		t d			ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ	n			ɳ	ɲ	ŋ	ɴ		
Vibrante simple	ʙ		r						ʀ		
Vibrante multiple				ɾ		ɽ					
Fricativa	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Fricativa lateral			ɬ ɮ								
Aproximante		ʋ	ɹ			ɻ	j	ɰ			
Aproximante lateral			l			ɭ	ʎ	ʟ			

CONSONANTES (NO PULMONARES)

Clics	Implosivas sonoras	Eyectivas
◌̩ Bilabial	◌̩ Bilabial	◌̩ como en
◌̪ Dental	◌̪ Dental/alveolar	◌̪ Bilabial
◌̫ (Post) alveolar	◌̫ Palatal	◌̫ Dental/alveolar
◌̬ Palatoalveolar	◌̬ Velar	◌̬ Velar
◌̭ Lateral alveolar	◌̭ Uvular	◌̭ Fricativa alveolar

SÍMBOLOS ADICIONALES

◌̮ Fricativa labio-velar sorda	◌̯ ◌̰ Fricativas alveolo-palatales
◌̩ Aproximante labio-velar sonora	◌̪ Vibrante simple lateral alveolar
◌̫ Aproximante labio-palatal sonora	◌̬ Simultáneamente ◌̭ y X
◌̭ Fricativa epiglótica sorda	Las africadas y las dobles articulaciones pueden representarse, si es necesario, mediante dos símbolos unidos por una ligadura:
◌̮ Fricativa epiglótica sonora	
◌̯ Oclusiva epiglótica	

$k̠p̠$ $t̠s̠$

Figura 11: Símbolos consonánticos (pulmonares y no pulmonares) y símbolos adicionales utilizados en el Alfabeto Fonético Internacional (de IPA (1993), «IPA chart, revised to 1993», *Journal of the International Phonetic Association*, 23(1)).

Tal como se aprecia en la figura 12, las vocales se representan en un cuadro que recoge las dimensiones anterior-posterior y abierto-cerrado anteriormente descritas. Las vocales redondeadas o labializadas se sitúan a la derecha de cada par en el que puede darse esta posibilidad, y las no redondeadas a la izquierda. Cabe señalar que el AFI simboliza las vocales que no forman el núcleo de una sílaba con un diacrítico, tal como puede verse en la figura 13.

VOCALES

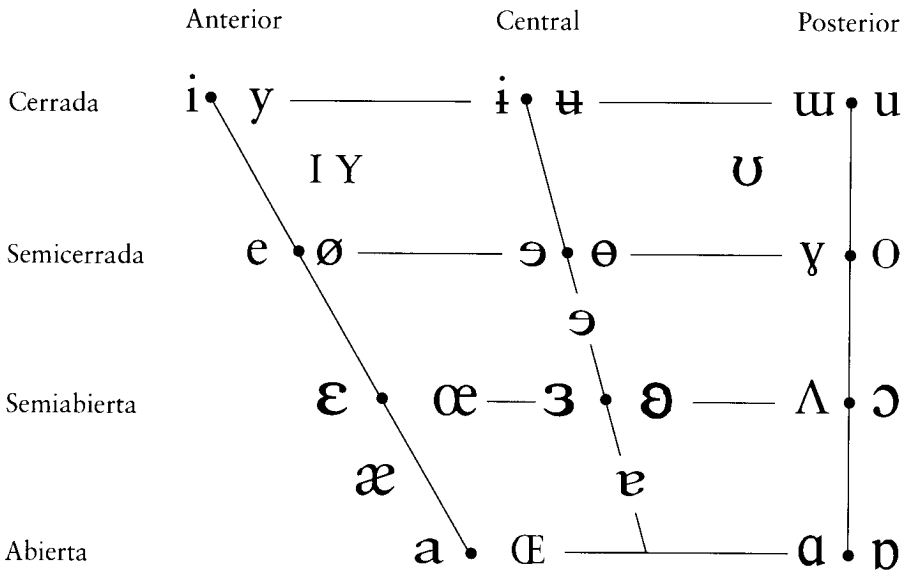


Figura 12: Símbolos vocálicos utilizados en el Alfabeto Fonético Internacional (de IPA [1993], «IPA chart, revised to 1993», *Journal of the International Phonetic Association*, 23[1]).

Por otra parte, el AFI incluye una serie de *diacríticos* que permiten expresar las características articulatorias secundarias o las modificaciones en la producción de un determinado sonido por efecto del contexto. Por ejemplo, existen diacríticos para indicar que una consonante sorda se ha convertido en sonora por influencia de una consonante adyacente —como sucede en el caso de [s] en la palabra castellana «desde»— o para señalar que una consonante alveolar tiene carácter dental por ir seguida de una consonante de este tipo —como [n] en la palabra castellana «canto».

DIACRÍTICOS

◌̚	Ensonorizada	ɳ̚ ɰ̚	◌̤	Murmurada	ɸ̤ ɹ̤	◌̪	Dental	ɸ̪ ɹ̪
◌̚	Sonorizada	ɳ̚ ɰ̚	◌̥	Laringalizada	ɸ̥ ɹ̥	◌̺	Apical	ɸ̺ ɹ̺
◌̚	Aspirada	ɳ̚ʰ ɰ̚ʰ	◌̦	Linguolabial	ɸ̦ ɹ̦	◌̻	Laminar	ɸ̻ ɹ̻
◌̚	Menos redondeada	ɸ̚ ɹ̚	◌̜	Labializada	ɸ̜ ɹ̜	◌̣	Nasalizada	ɸ̣ ɹ̣

DIACRÍTICOS (cont.)

◌̥	Menos redondeada	ç	j	Palatalizada	t ^j d ^j	n	Explosión nasal	d ⁿ
◌̤	Adelantada	ɥ	ɣ	Velarizada	t ^ɣ d ^ɣ	l	Explosión lateral	d ^l
◌̠	Retrotraída	ɨ	ɤ	Faringalizada	t ^ɤ d ^ɤ	◌̚	Sin explosión audible	d ^{◌̚}
◌̞	Centralizada	ɘ	~	Velarizada o faringalizada	ɫ			
◌̚	Medio centralizada	ɛ̥	◌̚	Más cerrada	ɛ̥ (ɨ̥ = Fricativa alveolar sonora)			
◌̚	Silábica	ɨ̥	◌̚	Más abierta	ɛ̥ (β̥ = Aproximante bilabial sonora)			
◌̚	No silábica	ɛ̥	◌̚	Raíz de la lengua adelantada	ɛ̥			
◌̚	Rotacidad	ə̥	◌̚	Raíz de la lengua retrotraída	ɛ̥			

Figura 13: Diacríticos utilizados en el Alfabeto Fonético Internacional (de IPA [1993]), «IPA chart, revised to 1993», *Journal of the International Phonetic Association*, 23[1]).

3. LA ESTRUCTURA ACÚSTICA DE LOS SONIDOS DEL HABLA

El resultado del funcionamiento de los mecanismos implicados en la producción del habla descritos en el apartado anterior es una onda sonora que se transmite del emisor al receptor. Esta onda —que, como se explica más adelante, no es más que la transmisión de variaciones de presión en las moléculas de aire— transporta toda la información necesaria para que el receptor pueda hacer una interpretación del mensaje, recuperando en cierto modo las unidades lingüísticas que el emisor ha codificado. La fonética acústica estudia las propiedades de la señal sonora en tanto que portadora de información sobre las unidades lingüísticas que forman el mensaje.

En este apartado se describen las propiedades acústicas de los sonidos del habla, introduciendo para ello algunas nociones básicas de acústica; éstas permiten definir las principales características de los sonidos del habla y, consideradas conjuntamente con el modo de funcionamiento del aparato fonador, permiten clasificarlos. Finalmente, las propiedades acústicas se relacionan con las articulatorias a fin de mostrar la dependencia de las primeras con respecto a las segundas.

3.1. LAS CARACTERÍSTICAS ACÚSTICAS DE LOS SONIDOS DEL HABLA

LA ONDA SONORA

En el apartado 2.1. se ha explicado que el resultado del paso del aire por la glotis es la emisión de una serie de sucesivas bocanadas de aire al ritmo de abertura y cierre de las cuerdas vocales. Con ello se consigue la producción de una onda sonora con unas características determinadas, que se exponen en este apartado. Para que se produzca la onda sonora las moléculas de aire deben entrar en vibración, y éste es justamente el resultado de su paso a través de los pliegues vocálicos.

La vibración de una molécula de aire puede representarse mediante el esquema de la figura 14.

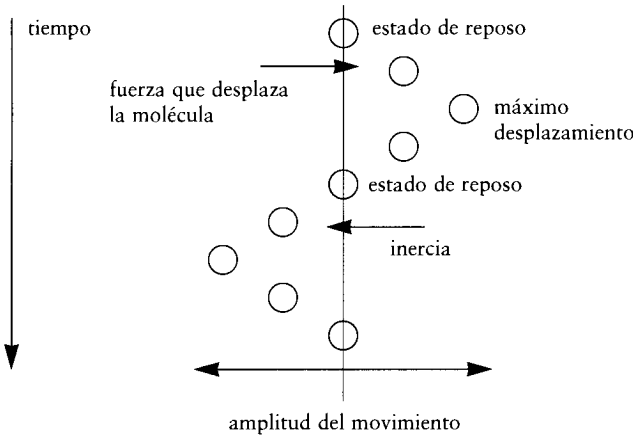


Figura 14: Representación esquematizada del movimiento vibratorio de una molécula de aire.

Al ser empujada por una fuerza, una molécula se desplaza y se aleja de su posición inicial de reposo; por su elasticidad tiende a volver a la posición de partida, pero debido a la inercia del movimiento se aleja de ella, desplazándose hacia el lado contrario; a continuación vuelve a la posición inicial, y como consecuencia de la fuerza que la ha empujado sigue alejándose de la posición de reposo. Este movimiento es similar al vaivén de un péndulo o de un columpio y, al igual que en estos casos, el desplazamiento respecto a la posición de reposo disminuye en amplitud a medida que cesa el efecto de la fuerza inicial. La *amplitud* del movimiento se define como la distancia entre la posición de reposo y el punto de máximo desplazamiento, mientras que el *tiempo* que dura el movimiento depende de la fuerza aplicada inicialmente y de la resistencia que ofrezca el medio en el que se produce la vibración. El movimiento de vaivén se repite diversas veces, tal como puede verse en la figura 15.

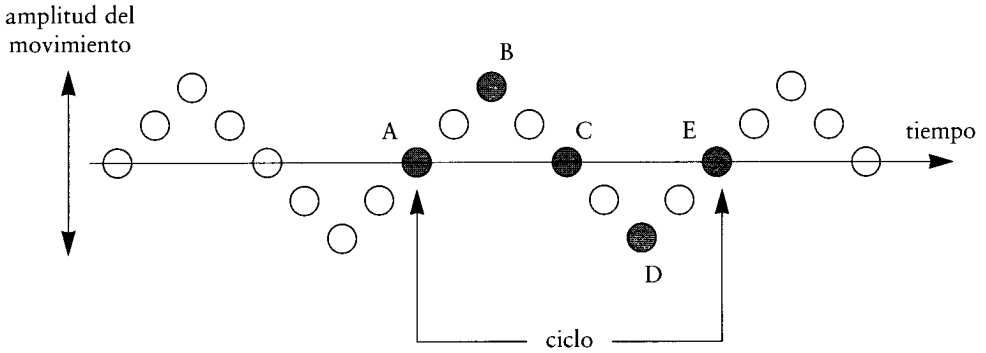


Figura 15: Representación esquematizada del movimiento vibratorio de una molécula de aire a lo largo del tiempo.

La *frecuencia* de un movimiento vibratorio como el representado en la figura 15 se define en función del número de veces que se va de la posición de reposo (A) al punto de separación máxima respecto a esta posición (B), se vuelve al punto de reposo (C), se llega de nuevo al punto de separación máxima (D) y se vuelve a la posición de reposo (E), es decir, en el número de *ciclos* por unidad de tiempo.

La vibración de las moléculas de aire que pasan por la glotis a través de las cuerdas vocales en movimiento tiene esta misma estructura, aunque la situación es algo más compleja. Puesto que hay más de una molécula, el movimiento de una desplaza a las demás, de modo que se crean zonas en las que las moléculas se encuentran muy juntas —con una presión elevada o zonas de compresión— y zonas en las que las moléculas se encuentran muy separadas entre sí —zonas de baja presión o rarefacción—. Si en el esquema de la figura 15 se representa la presión en el eje vertical y en lugar de una molécula se considera globalmente la corriente de aire, se obtiene una representación de los fenómenos aerodinámicos que tienen lugar al producirse una onda sonora.

La onda sonora es pues el resultado de la vibración de las moléculas del aire, y puede definirse en función de su amplitud, su frecuencia y del tiempo durante el cual se lleva a cabo el movimiento. Sin embargo, las representaciones observadas hasta ahora corresponden a las de los llamados sonidos o *tonos puros* —por ejemplo un silbato o el que se escucha al descolgar un teléfono—. La onda sonora producida en la glotis corresponde al tipo de sonidos denominados *compuestos* o *complejos*, consistentes en la superposición de varios tonos puros. En la figura 16 se muestra cómo puede obtenerse una onda sonora compuesta (d) a partir de la suma de tres tonos puros (a), (b) y (c); en ella, se han substituido las moléculas por una línea continua que indica su movimiento y su concentración en las zonas de presión y rarefacción mencionadas.

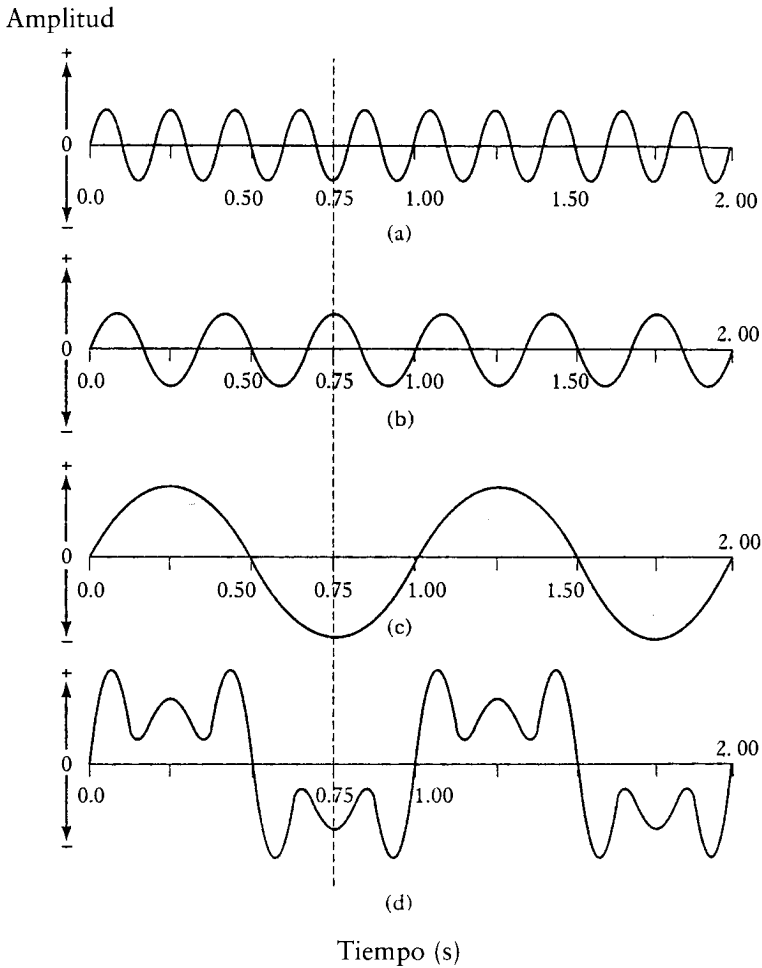


Figura 16: Obtención de una onda sonora compuesta (d) a partir de la suma de tonos puros (a, b y c) (de P. B. Denes & E. N. Pinson (1963), *The speech chain; the physics and biology of spoken language*. Anchor Press/Doubleday, Garden City, N.Y. W. H. Freeman, Nueva York, 1993, 2.ª ed., p. 33).

La obtención de una onda compuesta partiendo de varias simples no es más que un proceso consistente en sumar las amplitudes. Como puede verse en la figura 16, cada una de las ondas simples (a), (b) y (c) tiene una frecuencia propia, resultado de las repeticiones de ciclos en una misma unidad de tiempo. La onda sonora simple de frecuencia más baja entre todas las que forman una compuesta —(c) en el caso de la figura 16— se denomina *fundamental* (abreviado F_0), y las demás se definen como *armónicos*. La unidad de medida de la frecuencia es el *Hercio* (Hz), que equivale a un ciclo por segundo; por tanto, las ondas de la figura tienen una frecuencia de 5 Hz, 3 Hz y 1 Hz. El tipo de representación utilizado se conoce como representación temporal u *oscilograma*, y en ella se observa la *forma de la onda* sonora en su evolución a lo largo del tiempo.

Puede representarse también una onda sonora compuesta prescindiendo del tiempo y considerando únicamente la frecuencia y la amplitud de cada uno de sus armónicos, como aparece en la figura 17.

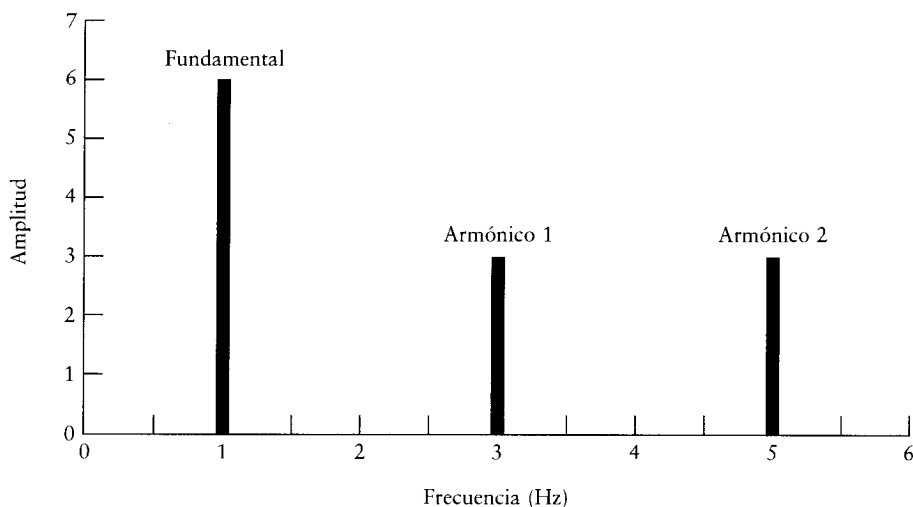


Figura 17: Representación espectral de la onda sonora compuesta (d) mostrada en la figura 16 (de P.B. Denes & E. N. Pinson (1963), *The speech chain; the physics and biology of spoken language*. Anchor Press/Doubleday, Garden City, N.Y. W. H. Freeman, Nueva York, 1993, 2.^a ed., p. 36).

En la representación de la figura 17, el eje horizontal muestra la frecuencia, y el vertical la amplitud: cada línea vertical representa un armónico y, de lo visto hasta ahora, se deduce que las formas de onda (a), (b) y (c) de la figura 16 corresponden a tonos puros, mientras que (d) corresponde a una onda sonora compuesta. El tipo de representación mostrado en la figura 17 recibe el nombre de *espectro*.

La vibración de las cuerdas vocales tiene como resultado acústico un sonido complejo con un espectro similar al representado en la parte superior (a) de la figura 18. La frecuencia del fundamental corresponde a la frecuencia de abertura y cierre de las cuerdas vocales, situándose alrededor de los 100 Hz —es decir, 100 veces por segundo— en los hombres y hacia los 250 Hz en las mujeres.

LA RESONANCIA Y LOS FORMANTES

El espectro de la onda sonora tal como se produce en la glotis presenta pues una forma similar a la que aparece en la parte superior (a) de la figura 18. En cambio, al analizar la misma onda sonora tras su paso por las cavidades supraglóticas —parte inferior (c) de la figura 18— es patente que se han producido modificaciones en la amplitud de cada uno de los armónicos.

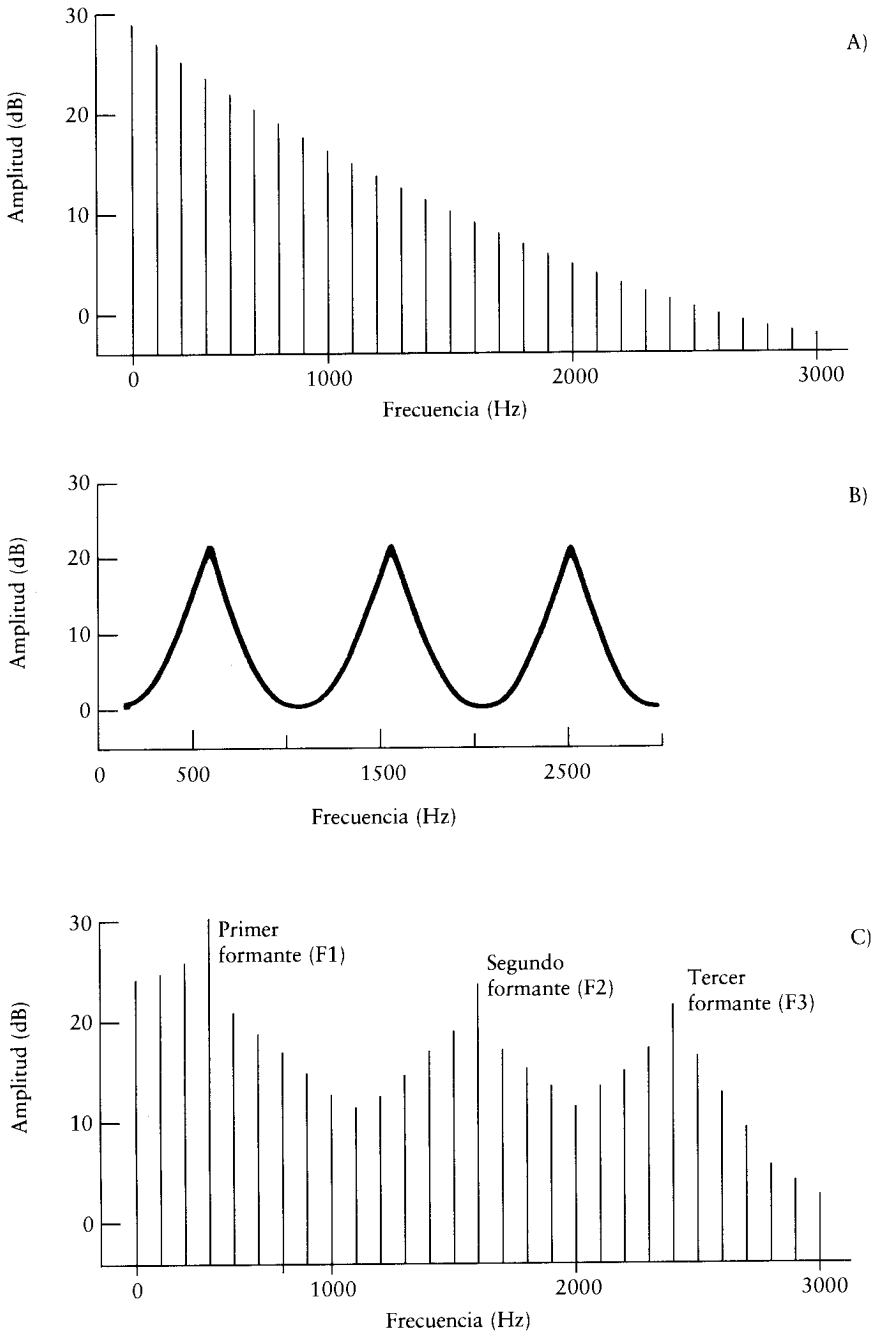


Figura 18: Espectro de la onda sonora en la glotis (a) y espectro de la onda sonora tras su paso por las cavidades supraglóticas (c); (b) representa la función de transferencia del tracto vocal (de G. J. Borden & K. S. Harris [1980], *Speech science primer. Physiology, acoustics and perception of speech*, p. 99. Williams & Wilkins, Baltimore).

Como se aprecia en la figura 18, de un espectro con una pendiente constante en el que la amplitud de los armónicos disminuye progresivamente a medida que aumenta su frecuencia (a), se pasa a un espectro con diferencias de amplitud entre los armónicos (c). Ello es debido al efecto de la *resonancia* que se produce en las cavidades supraglóticas descritas en 2.1. La resonancia es un fenómeno que también se da, por ejemplo, en instrumentos como un violín o un contrabajo; consiste en la modificación de la amplitud de los armónicos de un sonido complejo ejercida por la cavidad en la que se produce la vibración de las moléculas de aire; por regla general, las cavidades pequeñas refuerzan la amplitud de los armónicos de frecuencia alta —y por este motivo percibimos un sonido agudo en el caso del violín—, mientras que las cavidades grandes refuerzan la amplitud de los armónicos de frecuencia baja —con lo que, como en el caso de los contrabajos, obtenemos sonidos percibidos como graves—. Este mismo efecto se produce en las cavidades supraglóticas, de modo que la amplitud de los armónicos de la onda sonora procedente de la glotis queda alterada en función de la posición adoptada para la articulación de cada sonido. Especialmente en el caso de los sonidos periódicos complejos, puede observarse cómo se refuerza la amplitud de grupos de armónicos situados alrededor de una determinada frecuencia, configurándose así los llamados *formantes*.

Puede observarse de forma esquemática este fenómeno en el caso de las vocales, comparando, en la figura 19, la configuración de las cavidades supraglóticas en una vocal anterior como [i] y en una posterior como [u]. En la vocal anterior [i], el adelantamiento de la lengua crea una cavidad anterior al punto de constricción de tamaño relativamente pequeño, mientras que la cavidad posterior a este punto presenta un tamaño mayor. Esto se refleja en un espectro con un primer formante de frecuencia baja —relacionado con la cavidad posterior— y un segundo formante de frecuencia alta —relacionado con la cavidad anterior—. En cambio, en la vocal posterior [u], el punto de constricción, situado en la zona velar, crea dos cavidades grandes, una anterior —a la que contribuye también el adelantamiento de los labios— y otra posterior. Esto se refleja en el espectro de [u], que presenta dos formantes de frecuencia baja. La figura 19 esquematiza la relación entre la configuración articulatoria y el resultado acústico, representado en forma de espectro.

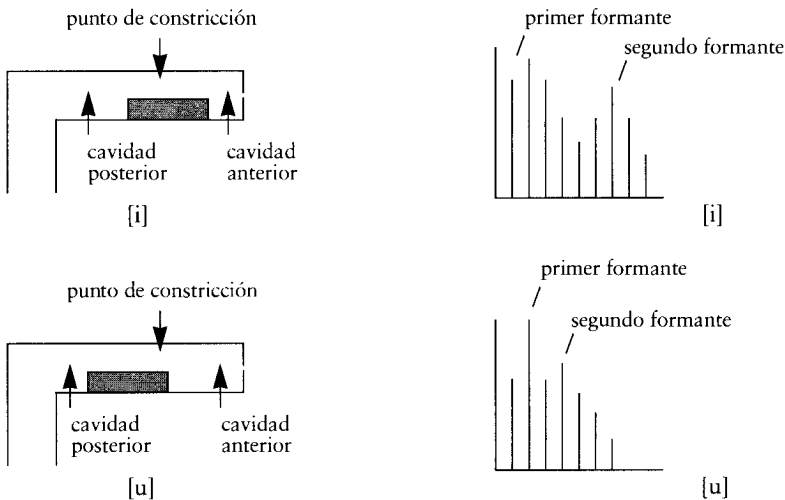


Figura 19: Representación esquemática de la relación entre la configuración articulatoria y las características acústicas de las vocales [i] e [u].

Si desde el punto de vista articulatorio estas dos vocales se distinguen por el punto en que se produce la constricción —es decir, por la aproximación de la lengua hacia la zona palatal o hacia la velar—, desde el punto de vista acústico quedan diferenciadas en función de la frecuencia a la que se encuentran esas zonas de armónicos de amplitud reforzada por acción de la resonancia de las cavidades supraglóticas o *formantes*. Este hecho es característico de todos aquellos sonidos que se producen con resonancia en el tracto vocal como las vocales, las nasales o las laterales.

SONIDOS APERIÓDICOS CONTINUOS E IMPULSIONALES

No todos los sonidos del habla se producen mediante la vibración de las cuerdas vocales. Algunos, como se ha visto anteriormente, se producen con la glotis abierta, por lo que el aire puede salir libremente. La interrupción de la corriente de aire tiene lugar entonces en algún punto del tracto vocal, como en el caso de las oclusivas y las fricativas sordas. En las primeras, tiene lugar un cierre completo del canal por el que pasa el aire, de forma que se da un aumento de presión en la zona posterior al cierre durante la fase conocida como *oclusión*, hasta que los articuladores se abren produciéndose entonces una salida rápida del aire acumulado —*explosión*. La consecuencia acústica de este fenómeno se traduce en un sonido aperiódico de tipo *impulsional*, caracterizado por una importante cantidad de energía sonora en un lapso de tiempo muy corto.

En cambio, en las fricativas sordas la turbulencia en la corriente de aire creada por el estrechamiento en el punto en que se produce la constricción tiene como resultado un sonido que en términos acústicos puede describirse como un sonido aperiódico *continuo*. La falta de periodicidad —debido a que no se da vibración de las cuerdas vocales— se manifiesta en un espectro sin armónicos.

Aun así, en la producción de los sonidos aperiódicos, tanto en los impulsionales como en los continuos, se observa igualmente una influencia del volumen de las cavidades supraglóticas en función del lugar en el que se produce la constricción. Ello se hará patente en el próximo apartado.

La clasificación acústica de los sonidos del habla puede resumirse, atendiendo a todo lo mencionado hasta ahora, de la forma siguiente (tabla 6):

sonidos periódicos	puros o simples		
	compuestos o complejos	vibración de las cuerdas vocales y resonancia	vocales, nasales, laterales
sonidos aperiódicos	impulsionales	cierre y explosión en el tracto vocal	oclusivas
	continuos	fricción en el tracto vocal	fricativas

Tabla 6: Clasificación acústica de los sonidos del habla.

3.2. LA CLASIFICACIÓN ACÚSTICA DE LOS SONIDOS DEL HABLA

EL MODELO DE LA FUENTE Y EL FILTRO

A fin de clasificar acústicamente los sonidos del habla, el aparato fonador puede considerarse en términos de una *fente* —el lugar donde se produce la corriente de aire indispensable para la producción del sonido— y un filtro —el conjunto de cavidades que, por el fenómeno de la resonancia, modifican las características de la fuente—. Existen entonces sonidos con una fuente *periódica* —cuando vibran las cuerdas vocales— y sonidos con fuente *aperiódica*, cuando se trata de *impulsiones* producidas por una explosión en algún punto del tracto vocal o cuando se trata de sonidos aperiódicos *continuos* causados por una fricción. La estructura del aparato fonador humano permite combinar esas dos fuentes simultáneamente: por ejemplo, las consonantes fricativas sonoras se producen mediante una fuente periódica —la vibración de las cuerdas vocales— combinada con una fuente aperiódica —la turbulencia en el lugar de constricción.

Las cavidades supraglóticas, como se ha visto anteriormente, modifican la amplitud de los armónicos de la fuente por efecto de la resonancia. Este efecto puede pensarse también como el de un *filtro* que permite el paso de determinados elementos y, en cambio, obstruye el de otros. Para explicar el proceso de forma matemática se recurre al concepto de *función de transferencia* (o función de filtro) —esquemática en la parte central (b) de la figura 18—, que no es otra cosa que la descripción de las características de las cavidades de resonancia en función de la frecuencia de los armónicos cuya frecuencia amplifican o reducen. Como se expone en el apartado 6.1., la posibilidad de formular matemáticamente las propiedades acústicas del tracto vocal en la producción del habla constituye la base para diversas aplicaciones en tecnologías del habla.

El filtro puede considerarse como *fijo* en los sonidos durante la producción de los cuales no se altera la forma de la cavidad de resonancia —por ejemplo en las vocales— y como *variable* cuando necesariamente tiene que producirse un cambio, como en el caso de las oclusivas, que constan de una fase de oclusión y otra de explosión. Al mismo tiempo, la resonancia puede darse en la cavidad *oral* o en la cavidad *nasal*, en función de la posición de la úvula cerrando o abriendo el paso de la corriente del aire hacia las fosas nasales.

La tabla 7, basada en Landercy y Renard (1977), resume las características acústicas de las principales clases de sonidos del habla en función de las propiedades de la fuente y el filtro:

Fuente	Filtro	Clase de sonidos
Periódica	Fijo, oral	vocales orales
Aperiódica continua	Fijo, oral	fricativas sordas
Aperiódica impulsional	Variable, oral	occlusivas sordas
Aperiódica continua + periódica	Fijo, oral	fricativas sonoras
Aperiódica impulsional + periódica	Variable, oral	occlusivas sonoras
Aperiódica impulsional + periódica	Variable, nasal	nasales
Aperiódica continua + periódica	Variable, oral	laterales y vibrantes

Tabla 7: Clasificación acústica de los sonidos del habla en función de la fuente y el filtro (de Landercy y Renard, 1977).

LA REPRESENTACIÓN ACÚSTICA DE LOS SONIDOS DEL HABLA

Las características de la fuente y el filtro se reflejan en las representaciones acústicas de los sonidos del habla que se utilizan habitualmente en fonética experimental. Hasta ahora se han presentado espectros de tonos complejos en los que se pone de manifiesto la amplitud y la frecuencia de cada uno de los armónicos. En la figura 20 puede observarse el espectro de las vocales [i] (20a) y [u] (20b), y el de la fricativa sorda [s] (20c).

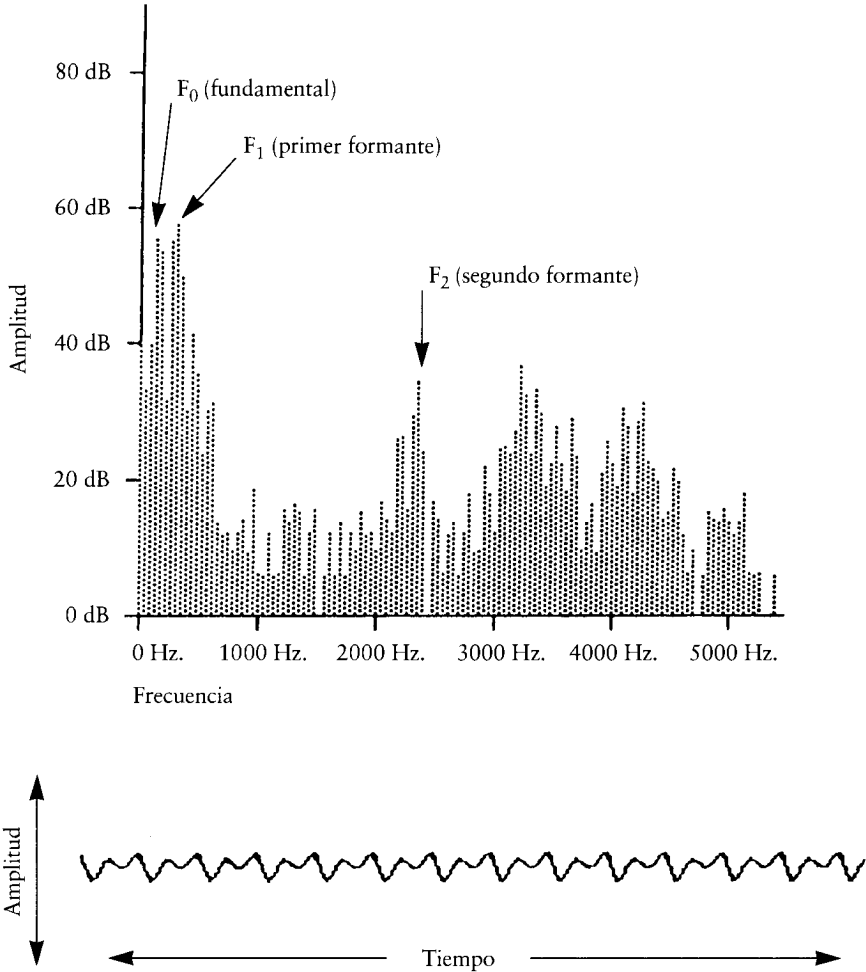


Figura 20a: Espectro y oscilograma de la vocal [i].

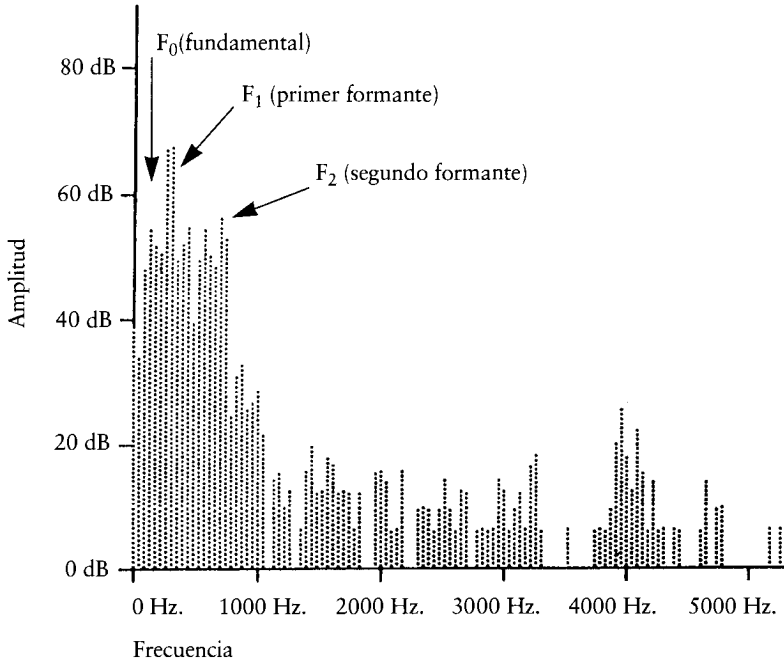


Figura 20b: Espectro de la vocal [u].

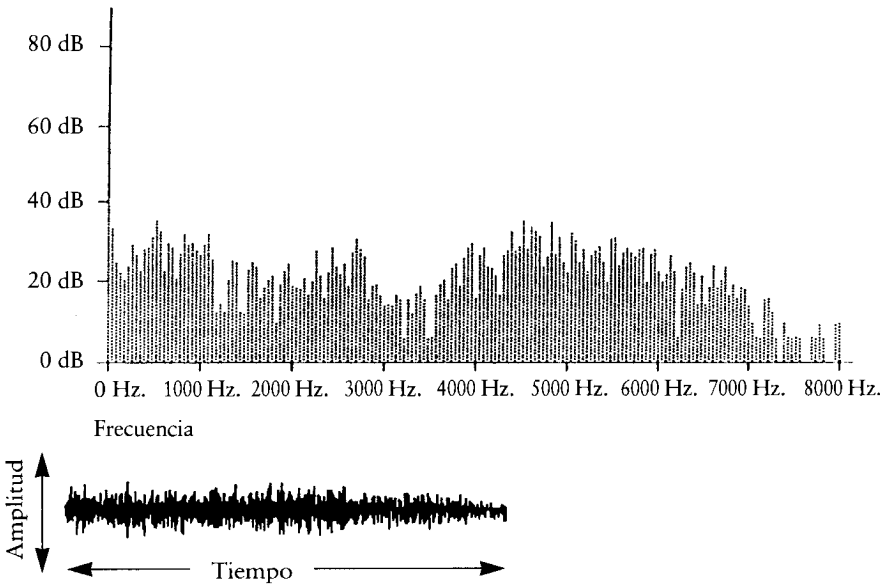


Figura 20c: Espectro y oscilograma de la fricativa sorda [s].

Desde el punto de vista de las características de la fuente, se observa que las vocales presentan una estructura espectral con armónicos y fundamental, ausente de la configuración de la fricativa sorda. También en la representación oscilográfica es patente la periodicidad de la onda sonora en la vocal (figuras 20a y 20b) y la ausencia de un ciclo que se repita en la fricativa (figura 20c). El efecto del filtro se manifiesta en el perfil superior del espectro: en el caso de las vocales presenta una estructura formántica, que no es visible de manera tan clara en la fricativa. En cambio, se observan también en la fricativa ciertas zonas de mayor intensidad de la energía sonora, producidas por el efecto de la forma y volumen de la cavidad en la que tiene lugar la turbulencia.

Una representación habitual en fonética es el *espectrograma*. En él se visualiza la evolución de la frecuencia y la amplitud a lo largo del tiempo. En el eje horizontal se representa el tiempo, y en el vertical, la frecuencia; la intensidad viene señalada por la gama de grises que aparece en el documento, de forma que las zonas más intensas se acercan más al negro. La figura 21 muestra un espectrograma de banda ancha de la realización [sa], junto con el oscilograma correspondiente:

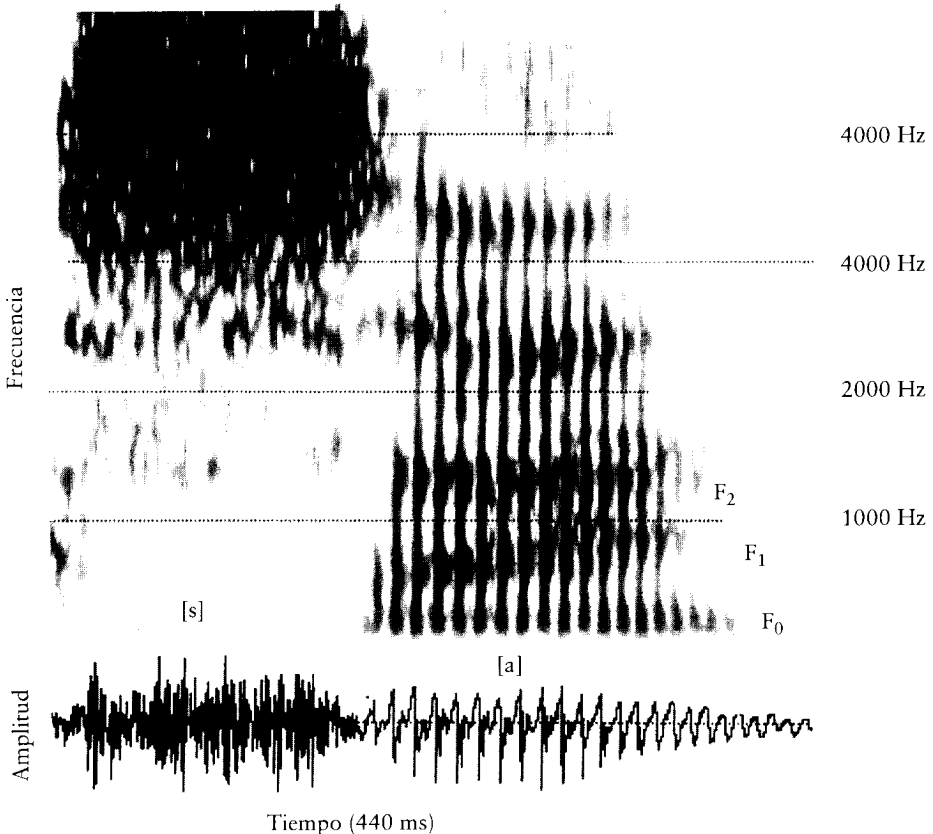


Figura 21: Espectrograma y oscilograma de la realización [sa].

En la figura 21 puede observarse, en primer lugar, una división clara entre los sonidos producidos con una fuente periódica —[a]— y los producidos con una fuente aperiódica —[s]—; en los primeros se aprecian estrías verticales que corresponden a la vibración de las cuerdas vocales, de modo que cada estría representa un ciclo; en los segundos no aparece esta estructura periódica, observándose una dispersión de la energía sonora en una determinada banda de frecuencia que es específica para cada consonante. El efecto del filtro puede observarse en las barras horizontales que constituyen bandas frecuenciales en las que el sonido presenta una intensidad mayor, correspondientes a los formantes (F_1 y F_2); la frecuencia a la que se sitúan corresponde a la de los picos identificados como formantes en los espectros de las figuras 20a y 20b, y el nivel de gris que indica su intensidad corresponde a la altura del pico en el espectro. En lo que respecta a la consonante fricativa, puede apreciarse que la energía sonora se distribuye en una banda de frecuencia determinada por su punto de articulación, banda que coincide con las zonas del espectro de la figura 20c en las que se aprecia una mayor intensidad sonora, correspondiente a la altura de los diferentes picos.

En el análisis acústico del habla se utiliza la información procedente de los tres tipos de documentos presentados: el oscilograma proporciona especialmente datos sobre aspectos temporales; en el espectro puede observarse la intensidad de la energía sonora en las diferentes bandas de frecuencias, permitiendo la determinación de la estructura formántica; en el espectrograma se encuentra la combinación de información temporal, frecuencial y de intensidad en un mismo documento. El modelo de la fuente y el filtro es indispensable para interpretar esos documentos y relacionar las características acústicas con las articulatorias.

CARACTERÍSTICAS ACÚSTICAS Y CARACTERÍSTICAS ARTICULATORIAS EN LA DESCRIPCIÓN DE LOS SONIDOS DEL HABLA

De todo lo visto hasta ahora, se desprende que los sonidos del habla pueden ser descritos en función de cuatro parámetros acústicos: *frecuencia del fundamental*, *configuración espectral* —o estructura formántica—, *amplitud* y *tiempo*. Estas características están, a su vez, relacionadas con propiedades articulatorias ya que, como se ha señalado anteriormente, el modo de producción de los sonidos es responsable de sus propiedades acústicas. Las relaciones entre ambas características se resumen en la tabla 8:

Características articulatorias	Características acústicas
Vibración de las cuerdas vocales	Frecuencia del fundamental (F_0)
Configuración de las cavidades supraglóticas	Configuración espectral y estructura formántica
Fuerza espiratoria	Amplitud
Duración de la espiración	Tiempo

Tabla 8: Relación entre las características articulatorias y las características acústicas de los sonidos del habla.

En resumen, el proceso de la producción del habla puede entenderse como un mecanismo mediante el cual una representación de las unidades lingüísticas que componen el mensaje se transforma, por la acción del aparato fonador, en una onda sonora transmitida al receptor. En el próximo apartado se presenta sucintamente el proceso que lleva a la interpretación de la onda sonora como un mensaje dotado de estructura lingüística.

4. LA PERCEPCIÓN DE LOS SONIDOS DEL HABLA

Este apartado se dedica a la etapa del proceso de la comunicación que corresponde al receptor, es decir, la audición y la percepción de los sonidos del habla. Se empieza describiendo el funcionamiento del sistema auditivo, para abordar a continuación una clasificación de los sonidos del habla basada en la manera como los procesa dicho sistema. Finalmente, se relacionan las características acústicas con las características perceptivas a fin de explicar el proceso de decodificación de las unidades lingüísticas presentes en la onda sonora.

4.1. EL PROCESO DE LA AUDICIÓN

EL OÍDO EXTERNO

La audición puede concebirse como un mecanismo que tiene como finalidad la transformación de los cambios de presión del aire en impulsos nerviosos que, al llegar al cerebro, se interpretan en términos lingüísticos como unidades integrantes de un mensaje. El órgano que lleva a cabo esta transformación es el oído, cuya estructura interna puede observarse de forma esquematizada en la figura 22. El proceso empieza en el llamado oído externo, formado por el pabellón auditivo y el canal auditivo externo. Una de las funciones del *pabellón auditivo* —o aurícula— es ayudar a la localización del sonido, mientras que el *canal auditivo* externo protege, con sus folículos pilosos y las glándulas ceruminosas, el resto del sistema auditivo.

EL OÍDO MEDIO

El proceso de transformación al que se aludía empieza propiamente en el oído medio, a cuya entrada se encuentra el *tímpano*. Éste es una membrana que vibra en función de la presión de las moléculas del aire y que comunica su movimiento a una cadena de tres huesecillos —*martillo*, *yunque* y *estribo*— conectada a su vez a otra membrana conocida como ventana oval. Tal cadena actúa como un sistema de palancas destinado, por una parte, a amplificar las vibraciones de la membrana timpánica; por otra, protege también las delicadas estructuras del oído interno, amortiguando las vibraciones demasiado intensas. La cadena de huesecillos se halla en una cavidad llena de aire —el oído medio— comunicada con la faringe mediante un canal denominado *trompa de Eustaquio*. Esto permite igualar la presión en el interior del oído con la presión exterior, impidiendo, por ejemplo, la sensación de taponamiento del oído que se produce en ciertas ocasiones.

EL OÍDO INTERNO

El oído medio acaba en una pared con otras dos membranas: la ventana oval y la ventana redonda. Mientras que la *ventana oval* recoge las vibraciones del último hueso de la cadena —el estribo—, la *ventana redonda* actúa compensando la presión en el interior del oído interno. En el oído interno se sitúan tanto las estructuras relacionadas con el equilibrio como las propiamente dedicadas a la audición. Su característica principal reside en ser un medio líquido —a diferencia del oído externo y del medio, en el que la transmisión del sonido se realiza por vía aérea— formado por el laberinto óseo y el laberinto membranoso, separados entre sí por el líquido perilinfático.

Tal vez la estructura anatómica más importante en la audición sea la cóclea, denominada así por tratarse de un canal óseo en forma de espiral de unos 35 mm de largo y con un área de 4 mm² en la base y 1 mm² en el extremo. La cóclea es también un medio líquido, en el que se recogen las vibraciones que llegan a la ventana oval a través del tímpano y de la cadena de huesecillos.

La cóclea está dividida en canales separados por membranas, en una de las cuales —la membrana basilar— se sitúa el *órgano de Corti*. Éste es un conjunto de células ciliares que se mueven siguiendo los desplazamientos de la membrana basilar producidos por el líquido contenido en la cóclea. En la base de tales células se encuentran las terminaciones nerviosas que forman las fibras del nervio auditivo, responsables de transmitir la información hasta las zonas del cerebro especializadas en el procesamiento de estímulos auditivos.

El mecanismo de esta transmisión es complejo, y se basa en el hecho de que distintas zonas de la cóclea responden a diferentes frecuencias de vibración del líquido contenido en su interior. La cóclea puede describirse como un analizador frecuencial, que, partiendo de las vibraciones de las moléculas de aire que llegan al tímpano y se transmiten a través del oído medio, las transforma en estímulos nerviosos —es decir, potenciales eléctricos— codificados en función de la frecuencia y la intensidad gracias a la acción de las células ciliares del órgano de Corti.

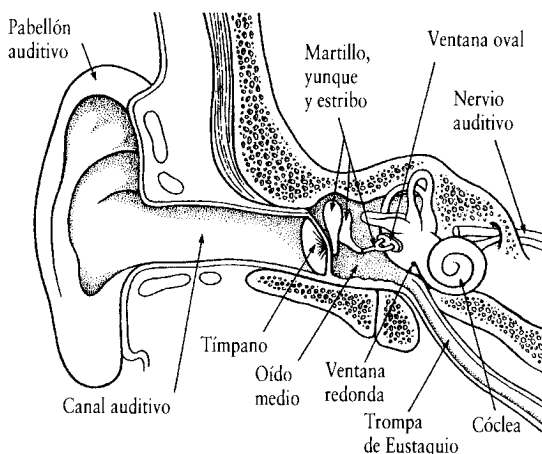


Figura 22: Representación esquemática del oído (de P. B. Denes & E. N. Pinson [1963]), *The speech chain; the physics and biology of spoken language*. Anchor Press/Doubleday, Garden City, N.Y. W. H. Freeman, Nueva York, 1993, 2.^a ed.

AUDICIÓN Y PERCEPCIÓN

El *nervio auditivo* recoge las terminaciones nerviosas del órgano de Corti y se encarga de hacer llegar a los dos hemisferios cerebrales los estímulos procedentes de la cóclea para que sean tratados en el *córtex auditivo* del cerebro. A partir de aquí empieza el proceso de la percepción, por el cual se interpretan lingüísticamente los estímulos recibidos. Hay que destacar que, si bien la audición puede estudiarse de forma directa, no es factible acceder de este modo a la percepción; su estudio experimental debe basarse, en cambio, en las respuestas subjetivas de oyentes que, en diversos tipos de pruebas de percepción, identifican y categorizan como sonidos del habla las señales acústicas que les llegan a través del oído.

4.2. LAS CARACTERÍSTICAS PERCEPTIVAS DE LOS SONIDOS DEL HABLA

LA CLASIFICACIÓN AUDITIVA DE LOS SONIDOS DEL HABLA

Así como las características acústicas de los sonidos del habla dependen del modo en que son producidos por el aparato fonador, también las propiedades auditivas del sonido dependen de su estructura acústica. La tabla 9 resume la relación entre características acústicas y auditivas:

Características acústicas	Características auditivas
Frecuencia del fundamental (F_0)	Altura (<i>pitch</i>) agudo - grave
Configuración espectral y estructural formántica	Timbre (<i>quality</i>): claro - oscuro
Amplitud	Intensidad (<i>loudness</i>): fuerte - flojo
Tiempo	Duración: largo-breve

Tabla 9: Relación entre las características acústicas y las características auditivas de los sonidos del habla.

La percepción de la *altura (pitch)* —agudo o grave— de un sonido depende esencialmente de la frecuencia de su fundamental (F_0), mientras que el *timbre (quality)* viene determinado por la distribución en el espectro de las bandas de frecuencias que presentan mayor intensidad. El hecho de percibir un sonido como fuerte o flojo depende de su *intensidad (loudness)*, mientras que la percepción de la *duración* viene dada por el tiempo. Así, desde el punto de vista auditivo, un sonido del habla puede ser definido en función de su altura, su timbre, su intensidad y su duración.

INDICIOS ACÚSTICOS E INDICIOS PERCEPTIVOS DE LOS SONIDOS DEL HABLA

En el apartado 3.2. se ha mostrado que los sonidos del habla presentan una realización acústica condicionada por su configuración articulatoria. En la percepción, la actividad del receptor consiste esencialmente en extraer de la onda sonora aquella informa-

ción que indica a qué clase de sonido del habla pertenece cada uno de los segmentos de la onda. Por tanto, puede imaginarse la producción como un proceso mediante el cual se codifican en forma de onda sonora las propiedades de cada una de las clases de los sonidos del habla, y la percepción como un mecanismo de decodificación de estas propiedades a fin de relacionar la onda sonora con una representación lingüística.

Es posible asociar propiedades acústicas de la señal sonora a características de ciertas clases de sonidos, como por ejemplo la explosión en el caso de las consonantes oclusivas, la sonoridad, o la estructura formántica de las vocales; estas propiedades se conocen como *indicios acústicos* (*acoustic cues*) y se utilizan en la descripción acústica de los sonidos del habla. En el momento de la percepción, el sistema de procesamiento del habla es capaz de detectar determinadas propiedades de la señal, que constituyen *indicios perceptivos* (*perceptual cues*) a partir de los cuales se lleva a cabo la asignación de un segmento de la onda sonora a una clase de sonidos. Estos indicios han sido estudiados de modo experimental desde finales de la década de los 50 y constituyen la base de los mecanismos perceptivos que nos permiten reconstruir la representación lingüística del mensaje. En el espectrograma de la figura 23 se muestran algunos de los indicios utilizados en la percepción de las consonantes oclusivas sordas.

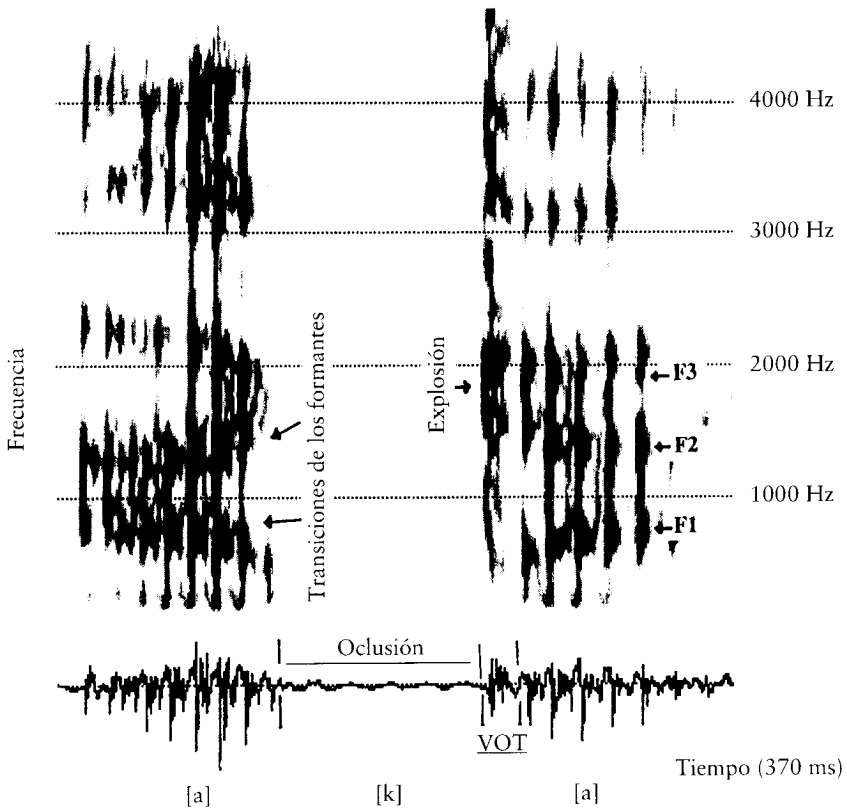


Figura 23: Indicios acústicos utilizados en la percepción de las consonantes oclusivas sordas.

Dos de los principales indicios que permiten identificar una oclusiva sorda son la oclusión y la explosión; mientras que la primera se manifiesta acústicamente en forma de ausencia de energía sonora, la segunda se presenta como una impulsión, un sonido aperiódico cuya distribución espectral depende del lugar de articulación de la consonante; también se considera un indicio perceptivo importante para identificar el lugar de articulación el período comprendido entre la explosión y el inicio de la sonoridad, conocido con las siglas de *VOT* (*Voice Onset Time* o tiempo de inicio de la sonoridad); éste aumenta progresivamente en duración de las labiales a las velares. Finalmente, los movimientos del aparato fonador en el paso gradual de la configuración propia de la vocal a la que es característica de la consonante se reflejan en variaciones frecuenciales de los formantes en la frontera entre la vocal y la consonante y la vocal; tales cambios se conocen con el nombre de *transición*, y constituyen también un indicio importante para la identificación del lugar de articulación de las oclusivas.

5. LOS ELEMENTOS SUPRASEGMENTALES

Hasta ahora se han considerado únicamente los elementos denominados segmentales —vocales y consonantes—, pero no debe olvidarse que en la producción del habla existen fenómenos fonéticos que afectan a más de un segmento, trátese de la sílaba, la palabra, el enunciado o el discurso. Por ello tales fenómenos se conocen como elementos *suprasegmentales* o como rasgos o elementos *prosódicos*.

En este apartado se definen en primer lugar los principales elementos suprasegmentales, examinando brevemente sus correlatos articulatorios, acústicos y perceptivos; para cada uno de ellos se presentan algunas indicaciones sobre su papel en la comunicación.

5.1. ACENTO

El *acento* puede definirse como la prominencia de una sílaba en contraste con las que la rodean. Diversos estudios experimentales han mostrado que el acento no presenta un único correlato articulatorio, ya que refleja la actividad de los músculos implicados en la respiración, la fuerza de salida del aire de los pulmones —es decir, el nivel de la presión subglótica— y, para ciertos autores, la fuerza articulatoria.

La situación es similar desde el punto de vista acústico, ya que las manifestaciones fonéticas del acento pueden traducirse en un aumento de la frecuencia del fundamental, de la amplitud o del tiempo en una sílaba; también en determinadas lenguas —por ejemplo, en las que existe reducción vocálica— el acento está relacionado con la estructura formántica de las vocales. Lo mismo sucede en el nivel perceptivo, ya que tanto la altura, como la intensidad, la duración o el timbre pueden contribuir a la percepción de una sílaba como acentuada.

En la descripción prosódica de determinadas lenguas suele hacerse referencia a la noción de *grupo acentual* —equiparado, por algunos autores, al *grupo de intensidad*—, formado por una sílaba acentuada y las sílabas no acentuadas que la siguen o la preceden; por ejemplo, el artículo y el nombre forman un grupo acentual en castellano. Para referirse al acento propio de las palabras suele usarse el término *acento léxico*, mientras que se habla de *acento de frase* en relación a una prominencia acentual que pone en relieve una determinada parte en el marco de todo el enunciado.

Las lenguas suelen dividirse entre las que presentan el acento léxico siempre en la misma posición en la palabra —denominadas *lenguas de acento fijo*—, entre las cuales se cuenta el francés, el checo o el polaco, y aquellas en las que el acento puede aparecer en cualquier posición dentro de la palabra —o *lenguas de acento libre*— como el castellano, catalán, italiano o inglés.

5.2. TONO

En las lenguas denominadas *tonales* existen variaciones de la frecuencia fundamental que producen cambios en el significado de las palabras, es decir, son fonológicamente pertinentes. Estas variaciones —o *tonos*— están ligadas a cambios en la frecuencia de vibración de las cuerdas vocales, y se perciben como variaciones en la altura. Los tonos se clasifican en función de la dirección del movimiento —ascendente, descendente o plano— y de la altura percibida —alto o bajo—. Por ejemplo, en una lengua tonal como el chino mandarín, la palabra [ma] adquiere un significado diferente según el tono con que sea pronunciada: con un tono alto plano significa «madre», con un tono alto ascendente «cáñamo», con tono alto descendente «regañar» y con tono bajo descendente-ascendente «caballo»; estas diferencias se reflejan en diferentes grafías —ideogramas en este caso— para cada una de las palabras.

5.3. MELODÍA Y ENTONACIÓN

La *melodía* es un elemento suprasegmental que se manifiesta en el nivel del enunciado. Desde el punto de vista articulatorio, la melodía se debe a las variaciones en la frecuencia de abertura y cierre de las cuerdas vocales que se producen en el habla, y que tienen como resultado acústico la variación de la frecuencia del fundamental (F_0) en el tiempo; estas variaciones melódicas se perciben como cambios en la altura del enunciado.

La representación acústica de la melodía viene dada por la evolución temporal de la frecuencia del fundamental, tal como se refleja en la figura 24.

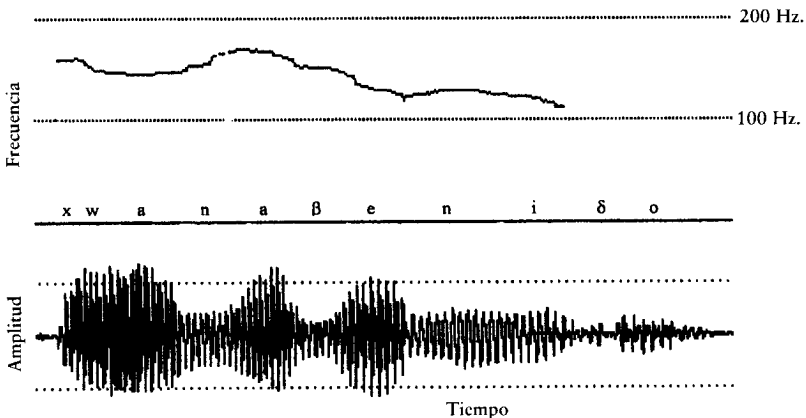


Figura 24: Curva melódica y oscilograma del enunciado «Juan ha venido» realizado con modalidad enunciativa.

La representación de la melodía que se muestra en la figura 24 se denomina *curva melódica* (*pitch contour*) y puede obtenerse a partir de sistemas de análisis acústico que detectan, mediante diversas técnicas, las variaciones de frecuencia fundamental a lo largo de la señal. También es posible obtener una representación de la curva melódica recogiendo directamente las vibraciones de las cuerdas vocales mediante electrodos de contacto aplicados externamente a la laringe; ésta es la base de sistemas como el laringógrafo o el electroglotógrafo.

Desde una perspectiva puramente lingüística, diversos autores han propuesto métodos para representar de forma más abstracta las variaciones de fundamental en un enunciado. Algunos sistemas clásicos emplean una serie de *niveles*, de forma que a cada sílaba se le asigna una altura tonal en relación a un conjunto de cuatro o cinco niveles. Es también clásica la representación de las variaciones tonales mediante *cifras* que representan el nivel de cada sílaba y *flechas* ascendentes o descendentes indicando la dirección del tono al final del enunciado. La representación de la melodía empleando la notación musical es un procedimiento que ha quedado ya desfasado.

La melodía tiene una *función lingüística* estrechamente relacionada con la sintaxis, y en lenguas como el español puede ser un procedimiento para la expresión de la modalidad oracional. En las figuras 25a y 25b se presentan las curvas melódicas del mismo enunciado de la figura 24 realizado con modalidad interrogativa (a) y con modalidad exclamativa (b).

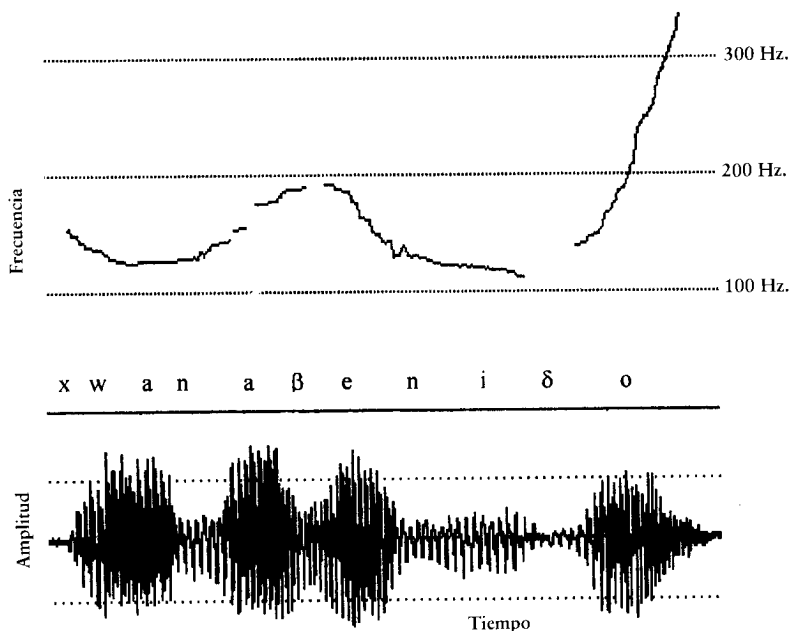


Figura 25a: Curva melódica y oscilograma del enunciado «Juan ha venido» realizado con modalidad interrogativa.

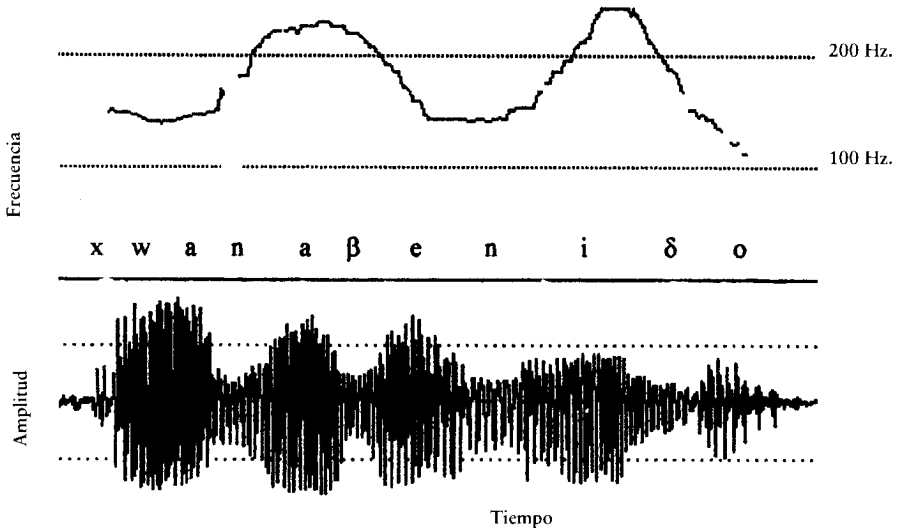


Figura 25b: Curva melódica y oscilograma del enunciado «Juan ha venido» realizado con modalidad exclamativa.

La melodía se emplea también para poner en relieve determinados constituyentes de la frase, por ejemplo en procesos como la dislocación relacionados con la distribución de la información en el enunciado. Igualmente, puede usarse como una marca para señalar la información nueva —foco— en un enunciado.

La melodía posee igualmente una *función extralingüística*; es conocida la anécdota relativa a las cuarenta variaciones melódicas que pedía Stanislavsky a sus actores sobre el enunciado «esta tarde», de manera que cada una de ellas transmitiera un contenido diferente.⁵ El estudio de las funciones de la melodía y de otros elementos fonéticos como portadores de información de tipo emocional relacionada con el estado de ánimo o la intención comunicativa del hablante corresponde a la *fonoestilística*.

La *entonación* es, para ciertos autores, la integración a nivel perceptivo del conjunto de los rasgos suprasegmentales. Esto puede justificarse fácilmente al observar que, en una curva melódica, las variaciones de frecuencia del fundamental reflejan simultáneamente el acento y la melodía. En cambio, para otros especialistas, la entonación equivale al fenómeno que en este apartado se ha denominado melodía.

5. R. JAKOBSON, (1960) «Linguistics and Poetics», en T. A. Sebeok, ed., *Style in Language*. Nueva York. (Trad. cast. de A. M. Gutiérrez y estudio preliminar de F. Abad: *Lingüística y poética*. pp. 34-35. Cátedra, Madrid, 1981. 4.ª ed.: 1988.)

5.4. PAUSA

Desde un punto de vista fonético, las *pausas* constituyen interrupciones en la fonación. Delimitan los denominados *grupos fónicos*, o unidades entre dos pausas. Una de las causas de la aparición de pausas es de tipo fisiológico, ya que éstas se relacionan con la necesidad de respirar; pero, en la mayoría de las ocasiones, la pausa actúa como elemento que delimita la estructura sintáctica o la estructura informativa de los enunciados y del discurso, por lo que puede asignársele una función lingüística. La pausa igualmente aparece en casos en que el hablante necesita planificar su discurso o cuando se producen interrupciones motivadas por el proceso de elaboración del mensaje, denominándose en este caso *pausas de duda* (*hesitation pause*).

En el análisis del discurso y de la conversación es habitual distinguir entre las pausas sin otra manifestación fonética que el silencio —*pausas vacías* (*empty pauses*)— y aquellas en las que se introducen elementos como «ehm», «eh», llamadas *pausas llenas* (*filled pauses*). La pausa cumple también una función estilística y retórica en el discurso oral, tal como puede observarse, por ejemplo, en la oratoria política.

5.5. VELOCIDAD DE ELOCUCIÓN

La *velocidad de elocución* o *tempo* (*speech rate*) consiste en el número de segmentos —en general sílabas— emitido por unidad de tiempo. Se habla de *velocidad de articulación* cuando en el cómputo se descuenta el tiempo dedicado a las pausas. La velocidad de elocución incide en las características articulatorias y acústicas de los segmentos, ya que está relacionada con fenómenos como la reducción vocálica o el grado de coarticulación.

5.6. RITMO

Finalmente, cabe considerar que el habla se produce a lo largo del decurso temporal, y que se percibe como una sucesión de prominencias —acentos—, interrupciones —pausas— y cambios tonales —melodía—. Cuando esta sucesión se repite a intervalos determinados aparece una estructura rítmica, producto de la alternancia entre partes del discurso realzadas y partes no marcadas. Aunque el ritmo está presente en toda manifestación hablada, se utiliza de forma especial como recurso expresivo en la lengua literaria.

5.7. LA TRANSCRIPCIÓN DE LOS ELEMENTOS SUPRASEGMENTALES

El Alfabeto Fonético Internacional ofrece también un sistema simbólico para la transcripción de los elementos suprasegmentales. Como puede verse en la figura 26, por una parte existe un conjunto de símbolos para la transcripción del acento, de la duración y de fenómenos relacionados con fronteras prosódicas; por otra, es posible transcribir tanto tonos como acentos de palabra atendiendo a la altura (nivel) o al contorno tonal. Es necesario destacar que, en el AFI, el acento debe aparecer antes de la sílaba tónica.

SUPRASEGMENTALES		TONOS Y ACENTOS DE PALABRA			
		NIVEL		CONTORNO	
ˈ	Acento principal	é	↗ Extraalto	ě	↗ Ascendente
ˌ	Acento secundario	é	↖ Alto	ê	↘ Descendente
ː	Larga	ē	↔ Medio	ě	↗ Alto ascendente
ˑ	Mediolarga	è	↘ Bajo	ě	↘ Bajo descendente
ˑ	Extrabreve	ẽ	↘ Extrabajo	ẽ	↗ Ascendente-descendente, etc.
•	Frontera silábica	↓	Descenso	↗	Ascenso global
	Grupo menor (pie)	↑	Ascenso	↘	Descenso global
	Grupo (entonativo) mayor				
˘	Enlace (sin separación)				

Figura 26: Símbolos para la transcripción de los elementos suprasegmentales utilizados en el Alfabeto Fonético Internacional (de IPA [1993], «IPA chart, revised to 1993», *Journal of the International Phonetic Association*, 23[1]).

6. LAS TECNOLOGÍAS DEL HABLA

Como se ha dicho al empezar el capítulo, uno de los ámbitos de aplicación de la fonética es la tecnología del habla. El objetivo de tal disciplina es lograr que la interacción con los múltiples sistemas informáticos utilizados habitualmente se lleve a cabo mediante el habla, prescindiendo del teclado y la pantalla y permitiendo, por tanto, el acceso por vía telefónica a servicios que pueden ofrecerse gracias al surgimiento de estas nuevas posibilidades.

Las *tecnologías del habla* constituyen un ámbito muy amplio que cubre esencialmente la producción de habla por parte de un sistema informático —síntesis— y el reconocimiento automático de un enunciado realizado por un hablante humano. Aunque aquí se hará referencia únicamente a estos dos aspectos, el campo cubre otras aplicaciones como la identificación o la verificación de la identidad del locutor o los sistemas que permiten establecer un diálogo entre un ordenador y un usuario humano. También la codificación del habla para hacer más eficaz su almacenamiento o su transmisión a distancia es un campo de trabajo integrado en las tecnologías del habla, aunque más cercano al procesamiento de las señales.

6.1. LA SÍNTESIS DEL HABLA

Existen en la actualidad diversos procedimientos para conseguir que un sistema informático proporcione información oral. La *síntesis del habla* puede realizarse por medio de varias estrategias, entre las cuales destacan tres: la codificación, la parametrización de unidades y la conversión de texto a habla.

En cada una de estas técnicas es necesario determinar, en primer lugar, el tipo de unidades que van a utilizarse en la generación de mensajes hablados. Habitualmente se contemplan desde unidades de alto nivel como enunciados enteros o palabras, hasta unidades de nivel más bajo como mitades de sílaba —*semisílabas*—, unidades formadas por varios alófonos —difonemas, trifonemas o polifonemas—, o incluso partes de alófonos. Una vez tomada esta decisión, es preciso constituir un *diccionario de unidades de síntesis* que contenga todas las unidades que sean necesarias de modo que, mediante su combinación, se creen todos los mensajes requeridos por el usuario. Así, los sistemas de síntesis pueden dividirse entre los que permiten generar cualquier mensaje —como en el caso de los conversores de texto a habla— y aquellos con los que sólo se pueden producir los enunciados que han sido previamente decididos —como en los sistemas de codificación.

En un sistema de síntesis basado en la *codificación de la onda sonora*, suelen elegirse palabras o enunciados enteros que son grabados por un hablante humano, almacenados y posteriormente combinados para producir los mensajes deseados. Ésta es la técnica empleada en sistemas como los que en las estaciones anuncian la salida de los trenes. Sin embargo, la onda sonora correspondiente a un enunciado almacenado en formato digital en un ordenador ocupa una gran cantidad de memoria, y es preciso introducir sistemas de codificación que reduzcan el tamaño de los mensajes almacenados. Esto se consigue con técnicas de procesamiento digital de señales con las cuales están familiarizados los ingenieros de telecomunicación. Al codificar se produce una reducción de la información sobre la onda sonora almacenada, información que se recupera en el momento de la decodificación cuando se necesita que el sistema produzca un determinado enunciado. Por ello, los sistemas de síntesis basados en la codificación pueden producir únicamente los mensajes que han sido previamente grabados, codificados y almacenados.

Mucho más interesantes desde el punto de vista de la fonética son aquellos sistemas basados en la concatenación de un conjunto de unidades parametrizadas. Habitualmente se trabaja con *difonemas* —segmentos que abarcan desde la mitad de un alófono a la mitad del siguiente— o con *trifonemas* —en los cuales se inserta un alófono entre los dos que aparecen divididos por la mitad—, aunque en algunos casos se introducen combinaciones más largas de unidades como los *polifonemas*. En el momento de realizar la síntesis de un mensaje, estas unidades se concatenan de manera que una unidad se una con la siguiente por un punto en el que no tengan lugar grandes variaciones acústicas: por ello, un difonema consta de la mitad de la parte estable del primer segmento, la transición hacia el siguiente y la mitad de la parte estable del segundo.

El primer paso para la construcción de un sistema de síntesis mediante unidades parametrizadas consiste, pues, en la realización del inventario de segmentos de la lengua y en el estudio de las posibilidades de combinación entre ellos para formar difonemas u otras unidades mayores; a continuación se realiza una grabación de las unidades por parte de un hablante humano, a la que se aplican técnicas de *parametrización*. El objetivo es almacenar en la memoria del ordenador únicamente la información que permite reconstruir cada unidad. Por ejemplo, en una vocal puede considerarse la duración y la frecuencia de sus formantes; en una oclusiva la duración de la fase de oclusión, las características temporales, frecuenciales y de intensidad de la explosión y la transición de los formantes hacia la vocal. Así, cada unidad se reduce a un conjunto de valores de parámetros que en el momento de la síntesis son interpretados por un programa que

modela la función del tracto vocal, generando la onda sonora correspondiente a cada unidad a partir de los valores numéricos para cada parámetro que se han almacenado previamente.

Esta función puede realizarla un *sintetizador por formantes*, consistente esencialmente en una fuente y un filtro que reproducen en términos acústicos el funcionamiento del tracto vocal. En los parámetros relativos a la fuente se especifica, por ejemplo, el carácter sonoro o sordo de cada unidad, y en los parámetros relativos al filtro la frecuencia de sus formantes. Sin embargo, existen otras posibilidades como la codificación de las unidades mediante la técnica conocida como *predicción lineal* (LPC, *Linear Predictive Coding*), mediante la cual pueden parametrizarse las unidades de síntesis en términos de una serie de coeficientes que resumen sus propiedades acústicas y que sirven para el control de un sintetizador que las interpreta. Con este tipo de sistemas basados en la parametrización y concatenación de unidades es posible generar un número infinito de mensajes.

En muchas de las aplicaciones de la síntesis interesa llegar a la realización oral de un texto almacenado en forma escrita. Para ello se han desarrollado múltiples sistemas de *conversión de texto a habla*, cuya utilidad es la lectura de textos escritos y guardados en soporte informático. Tales sistemas se basan también en un conjunto de unidades previamente almacenadas en forma de valores de parámetros, pero requieren además un módulo que convierta la representación escrita en una representación fonética en la cual el sistema de síntesis reconozca las unidades que tiene almacenadas para así proceder a su concatenación.

El primer paso en la conversión de texto a habla es el *pretratamiento del texto*, convirtiendo en una secuencia de letras las cifras, fechas, horas, abreviaturas y otros símbolos especiales. A continuación se realiza la *transcripción fonética automática* o conversión de grafía a alófonos, para pasar después a buscar en el diccionario las unidades necesarias para constituir un mensaje. Para la buena lectura de un texto es necesario también asignar correctamente pausas, dividiendo el enunciado en grupos fónicos, decidir adecuadamente la duración de cada segmento en función de su contexto y, muy especialmente, producir una curva melódica adecuada. Todo ello se realiza en diversas etapas que suelen formar parte del *módulo prosódico*, complementado en muchos casos por un *módulo lingüístico* que realiza un análisis morfológico o sintáctico especialmente orientado a determinar las características prosódicas de la realización oral del texto.

Las aplicaciones de la conversión de texto a habla son múltiples, desde el acceso telefónico a sistemas de información en los que los datos proporcionados al usuario varían frecuentemente —por ejemplo en el caso de noticias de actualidad, boletines meteorológicos, carteleras de espectáculos— hasta la lectura de pantallas de ordenador para que personas con deficiencias visuales puedan utilizar sin dificultad herramientas informáticas.

6.2. EL RECONOCIMIENTO DEL HABLA

El *reconocimiento del habla* puede considerarse el proceso inverso a la síntesis: si en ésta se genera un mensaje oral, en el reconocimiento se trata de transformar un enunciado hablado a fin de convertirlo en texto escrito de modo que pueda ser utiliza-

do en un entorno informático. El reconocimiento no debe confundirse con la comprensión del habla, ya que éste no implica una interpretación del contenido del mensaje, sino una traducción a una representación escrita o a comandos que permitan el control de un sistema informático.

Habitualmente suelen distinguirse los sistemas que reconocen *palabras aisladas* de los que pueden tratar *habla conectada* o *continua*. En los primeros, se reconocen en general órdenes o comandos que controlan un sistema como si se tratara del menú que se encuentra en muchos programas, mientras que los segundos abordan tareas de mayor complejidad como el reconocimiento de números de teléfono pronunciados sin separación entre los dígitos; por su parte, los sistemas de reconocimiento de habla continua son necesarios en aplicaciones como el dictado automático, que permite convertir en un texto escrito cualquier enunciado oral.

Para llevar a cabo el reconocimiento se utilizan, al igual que en la síntesis, diversas estrategias. El reconocimiento de palabras aisladas suele hacerse por *comparación de patrones*; en este caso, es preciso grabar primero las palabras que deben reconocerse para que después pueda llevarse a cabo la comparación entre estos modelos y las producciones de cada locutor. En cambio, el reconocimiento del habla continua requiere ya un análisis de la señal sonora extrayendo una serie de unidades —similares en algunos casos a las utilizadas en la síntesis— y deduciendo de su concatenación el enunciado originalmente emitido. Para ello es preciso conocer también las reglas de combinación de unidades y las probabilidades de aparición de cada una en combinación con las demás. La primera fase se lleva a cabo mediante un proceso conocido como *decodificación acústico-fonética*, en el que se intenta obtener una representación fonética de la señal sonora; la segunda suele realizarse con la ayuda de un *modelo de lenguaje*, consistente en información de tipo probabilístico o estadístico sobre la combinación de las unidades.

Un segundo parámetro utilizado en la descripción de los sistemas de reconocimiento de habla es el *número de locutores* que es capaz de reconocer. Esto está ligado a la noción de entrenamiento del sistema. Si, por ejemplo, un sistema debe reconocer palabras aisladas pronunciadas siempre por la misma persona, el entrenamiento se lleva a cabo mediante diversas realizaciones de cada una de estas palabras de modo que pueda crearse un modelo o plantilla de cada palabra; tal plantilla se utiliza posteriormente como referencia para comparar con las palabras que el usuario grabado previamente pronuncia al trabajar con el sistema. Si, en cambio, el sistema debe reconocer el habla de cualquier persona, es preciso entonces recoger realizaciones fonéticas de un gran número de hablantes. En el primer caso se trata de *sistemas unilocutor*, mientras que los segundos se conocen como sistemas *multilocutor*.

Cabe considerar, finalmente, el *número de enunciados* que el sistema es capaz de reconocer. Al igual que en la síntesis, existen sistemas que reconocen únicamente aquellas palabras con las cuales han sido entrenados, y sistemas que son capaces de reconocer cualquier enunciado; mientras que los primeros suelen operar comparando modelos de palabras, los segundos requieren procedimientos como la decodificación acústico-fonética acompañada de un modelo de lenguaje tal como se ha descrito más arriba.

Las aplicaciones del reconocimiento son diversas; por una parte, cabe mencionar aquellas que se centran en el control de sistemas como ordenadores o teléfonos mediante comandos vocales; en el otro extremo se sitúan los sistemas de dictado auto-

mático mediante los cuales se consigue obtener por escrito un texto producido oralmente. En algunas aplicaciones, como en los llamados *sistemas de diálogo*, es importante disponer simultáneamente de la síntesis y del reconocimiento, de modo que sea posible llevar a cabo una interacción completa entre un usuario y un sistema informático; por ejemplo, se utilizan sistemas de diálogo para reservas de billetes o información sobre horarios de transporte por vía telefónica sin que sea necesaria la intervención de un operador humano. Los sistemas de diálogo son un componente importante, junto con la síntesis y el reconocimiento, de la *traducción oral automática*, en la cual cada uno de los interlocutores de una conversación se expresa en su propia lengua y escucha la respuesta también en esta lengua; estos sistemas requieren además la existencia de métodos de traducción automática, desarrollados por quienes se dedican al tratamiento de textos escritos en el campo conocido como procesamiento del lenguaje natural.

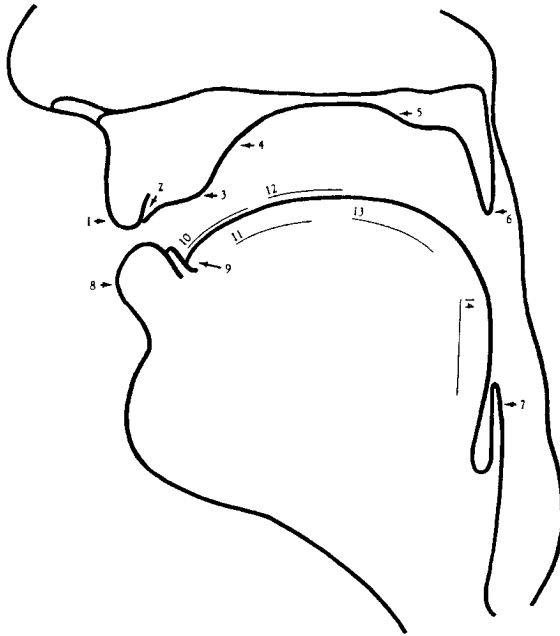
7. CONCLUSIÓN

En este capítulo se ha presentado la fonética como un campo de trabajo interdisciplinar cuyo objetivo es la descripción y explicación de los procesos implicados en la producción y la percepción de los sonidos del habla en el marco de la comunicación humana. Por ello, la fonética se integra en las denominadas ciencias del habla sin dejar de ser una disciplina en estrecha relación con la lingüística. Se han descrito sucintamente tres perspectivas de análisis —la articulatoria, la acústica y la perceptiva—, haciendo hincapié en la necesidad de considerarlas conjuntamente, puesto que mientras que en la producción del habla se codifica un conjunto de unidades lingüísticas portadoras de un mensaje en forma de señal sonora, en la percepción se realiza la operación inversa.

Entre los ámbitos a los que la fonética puede realizar aportaciones se ha abordado especialmente el de las tecnologías del habla, como paradigma de una interacción entre diversos saberes que permite desarrollar aplicaciones de utilidad para mejorar la comunicación entre personas y entre personas y sistemas informáticos. Con ello ha querido mostrarse también que la colaboración de especialistas en fonética en equipos multidisciplinares es esencial para llegar a una mejor comprensión de la comunicación oral en todas sus facetas.

EJERCICIOS 6

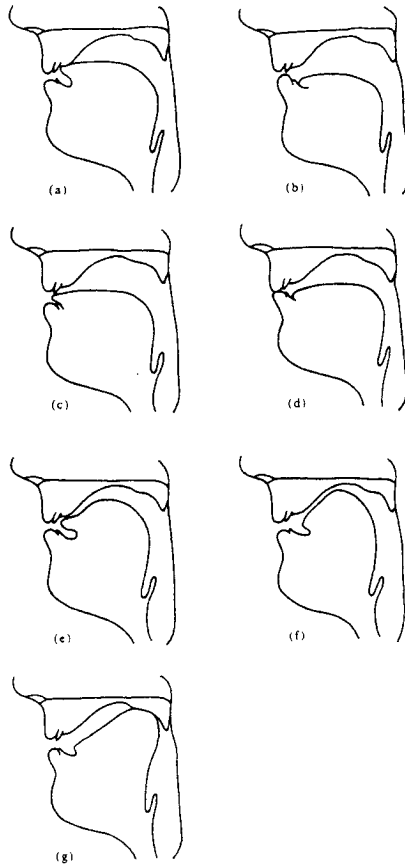
1. ¿Puede afirmarse que la lengua oral tiene prioridad sobre la lengua escrita? Argumente la respuesta, ayudándose del trabajo de Murillo (1979) citado en la nota 2.
2. ¿Es posible transcribir fonéticamente un enunciado completo en una lengua desconocida para la persona que hace la transcripción? ¿Es posible leer con acento perfectamente nativo un fragmento en transcripción fonética de una lengua que no se conoce? Argumente las respuestas.
3. Identifique cada una de las partes del aparato fonador que aparecen numeradas en la siguiente figura:



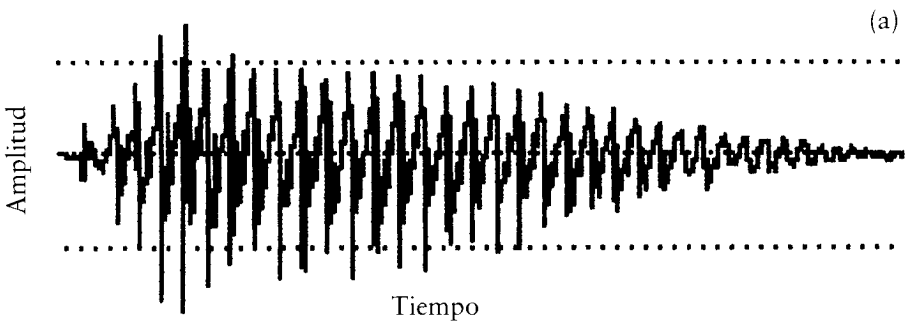
4. Describa articulatoriamente, indicando el modo de articulación, el lugar de articulación y el estado de la glotis, las consonantes que aparecen en el siguiente enunciado: «El chocolate es necesario para fortalecer el organismo».

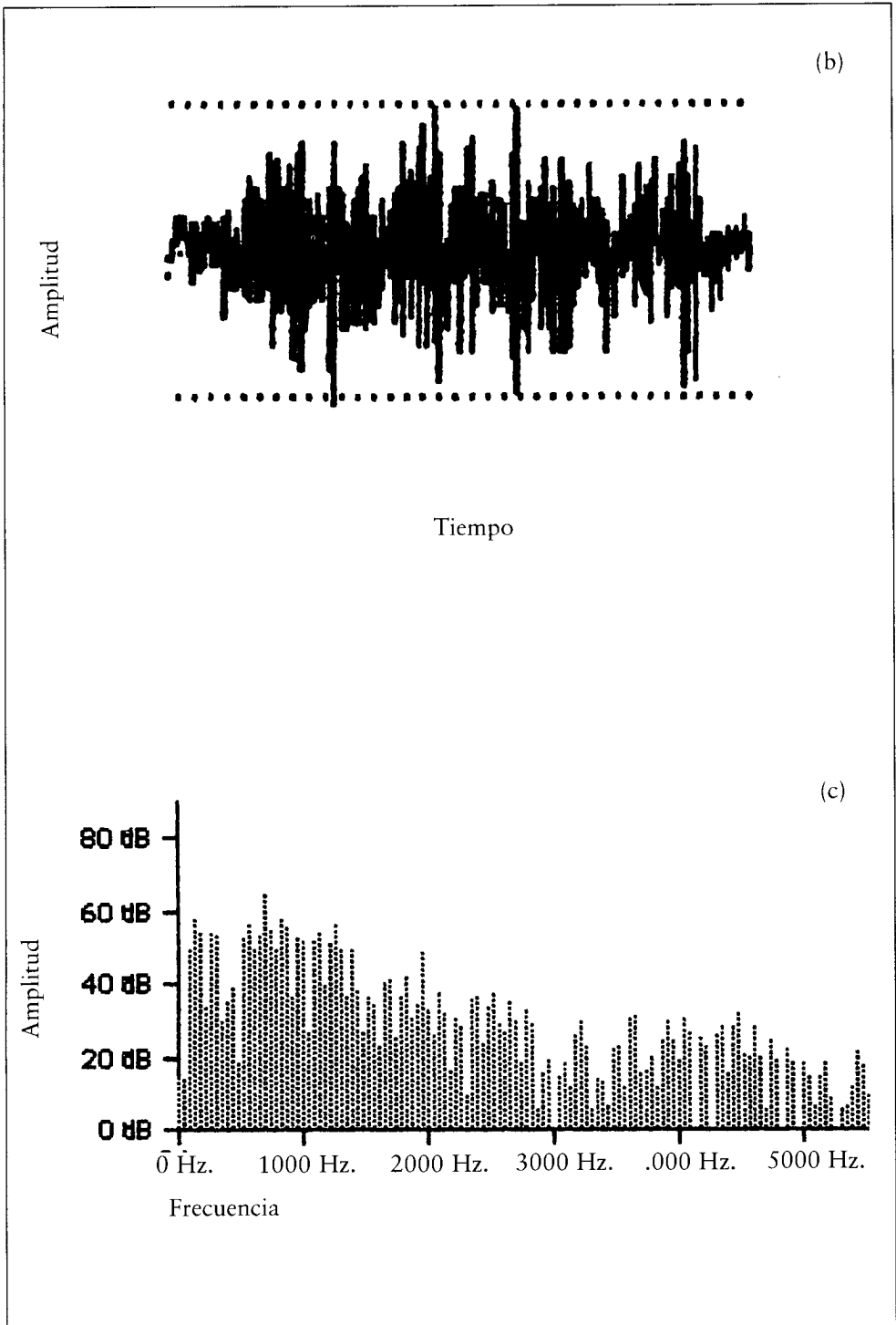
6. Algunos de los ejercicios que se presentan a continuación están adaptados de Ladefoged (1975) y de Landercy-Renard (1977).

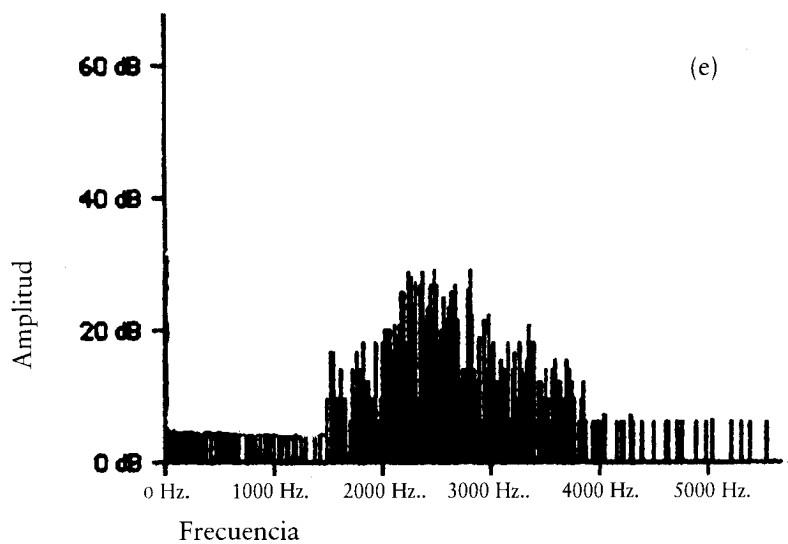
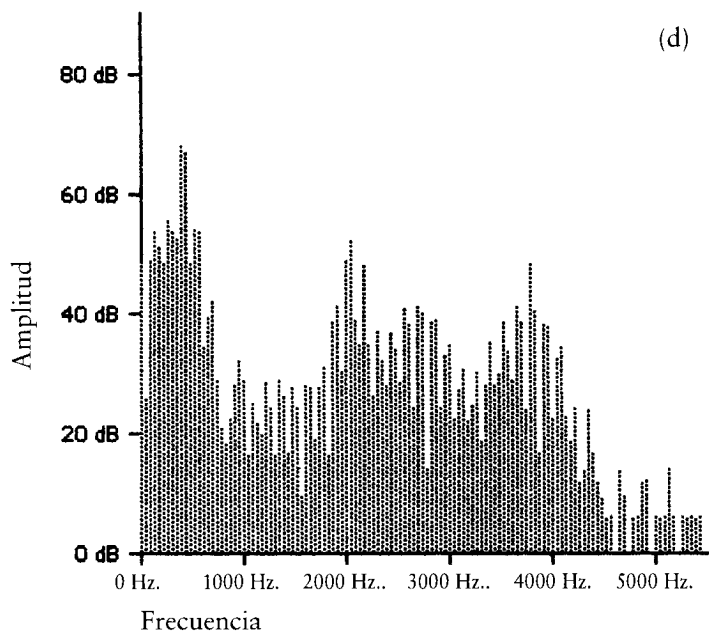
5. Indique a qué clase de sonidos, precisando el punto y el modo de articulación, pueden corresponder las siguientes representaciones:

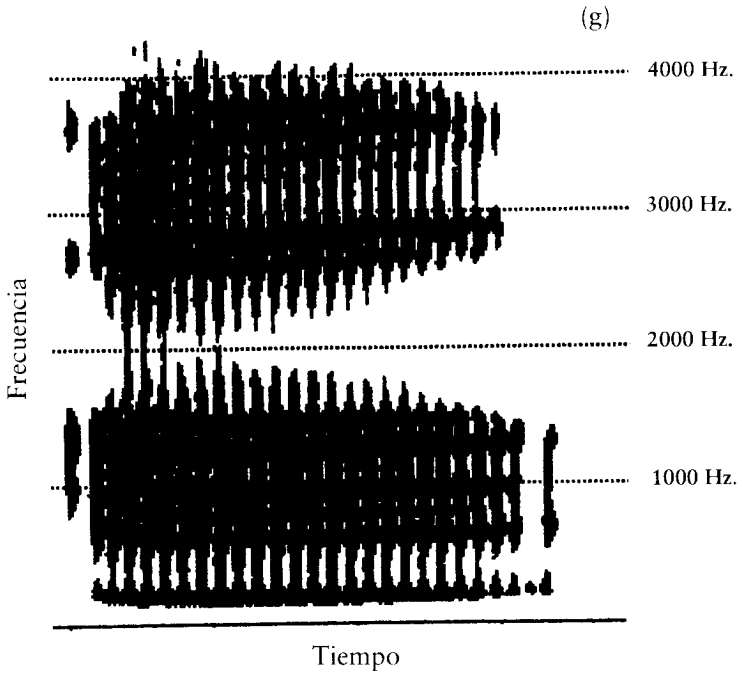
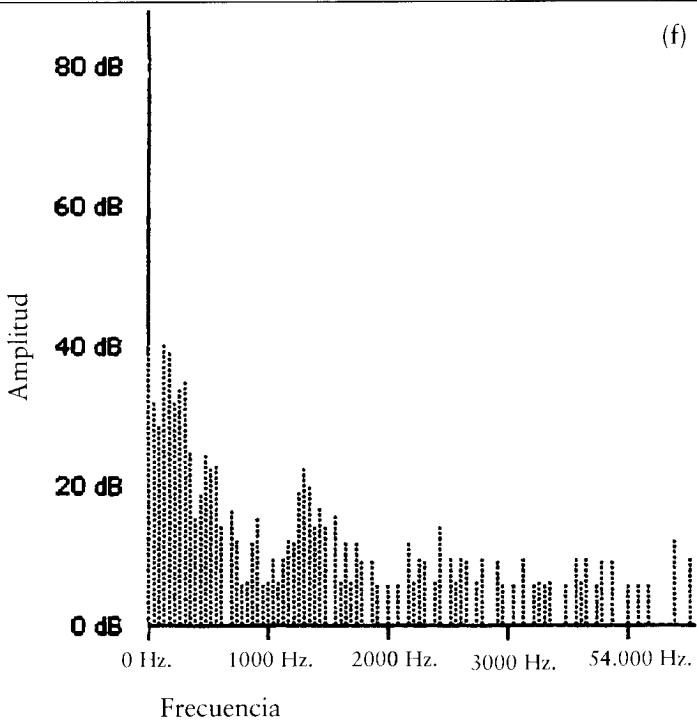


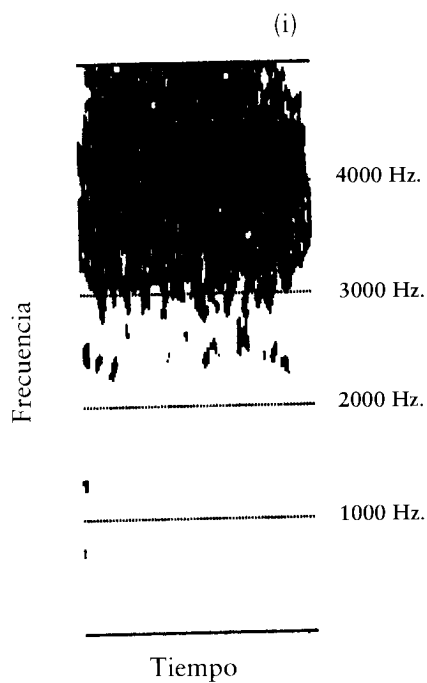
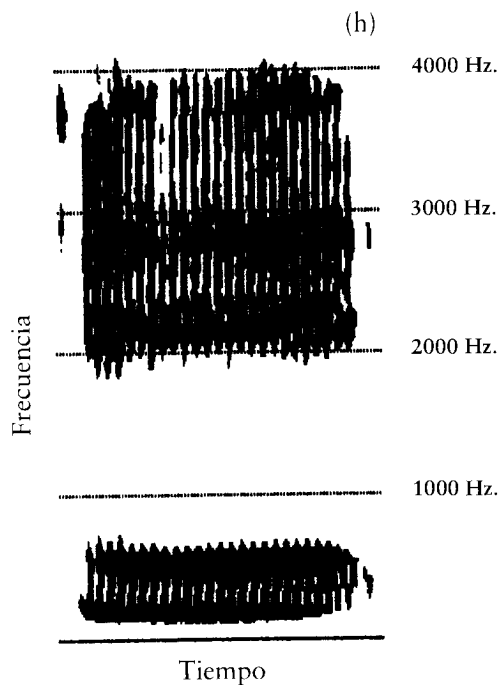
6. Observe las representaciones acústicas correspondientes a las realizaciones fonéticas de [a], [e], [s], [ma] y conteste a las preguntas que vienen a continuación:

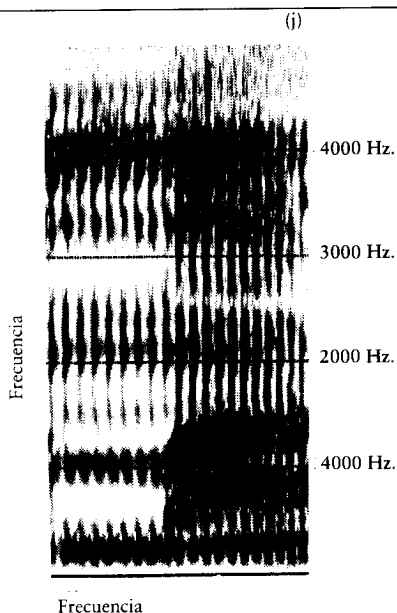












En (a) y (b) puede observarse la periodicidad de los sonidos.

En (a) y (b) se distingue la sonoridad.

En las representaciones como las de (a) y (b) es posible observar diferencias de duración entre los sonidos.

En (a) puede observarse la estructura formántica de la vocal.

La estructura formántica de la vocal [a] puede observarse en (c) pero no en (g).

La duración de cada sonido es observable en (c), (d), (e) y (f).

Tanto (f) como (j) nos permiten determinar el carácter sonoro o sordo de los sonidos representados.

Las transiciones de los formantes pueden observarse en (j) y en (f).

¿A qué sonido y a qué tipo de documento de análisis acústico corresponde cada una de las representaciones de la figura?:

	Documento de análisis	Sonido
(a)		
(b)		
(c)		
(d)		
(e)		
(f)		
(g)		
(h)		
(i)		
(j)		

¿Cómo pueden caracterizarse los sonidos [a], [e], [s] y [m] en términos de la fuente y el filtro? ¿Qué información sobre cada una de estas características ofrecen los distintos documentos presentados en la figura?

7. Complete la tabla siguiente:

Correlato articulatorio	Correlato acústico	Correlato perceptivo
		Altura
Configuración de las cavidades supraglóticas		
	Amplitud	
		Duración

8. Complete la siguiente tabla utilizando los términos mencionados a continuación: membrana basilar, estribo, conversión en impulsos nerviosos, pabellón auditivo, recogida de las vibraciones del aire, órgano de Corti, conducto auditivo, martillo, cóclea, conversión en vibraciones mecánicas, yunque, tímpano:

	Oído externo	Oído medio	Oído interno
Función			
Órganos			

REFERENCIAS BIBLIOGRÁFICAS

BORDEN, G. J. & HARRIS, K. S. (1980), *Speech science primer. Physiology, acoustics and perception of speech*. Williams & Wilkins, Baltimore.

Excelente manual para profundizar en las nociones presentadas en los apartados 2, 3 y 4. Especialmente recomendable para estudiantes de segundo ciclo.

CLARK, J. & YALLOP, C. (1990), *An introduction to phonetics and phonology*. Basil Blackwell, Oxford.

Los capítulos de fonética son muy adecuados al nivel de primer curso, aunque fundamentalmente se basa en la descripción del inglés. Contiene, además, un buen tratamiento de los elementos suprasegmentales.

DANILOFF, R., SCHUCKERS, G. & FETH, L. (1980), *The physiology of speech and hearing*. Prentice-Hall, Englewood Cliffs, N. J.

Permite profundizar en la descripción de los mecanismos articulatorios y auditivos, por lo que es muy adecuado para estudiantes de segundo ciclo.

DENES, P. B. & PINSON, E. N. (1963), *The speech chain; the physics and biology of spoken language*. Anchor Press/Doubleday, Garden City, N.Y. W. H. Freeman, Nueva York, 1993, 2.^a ed.

Constituye una muy buena introducción al estudio de la comunicación oral desde una perspectiva interdisciplinaria, considerando tanto la fonética como las tecnologías del habla.

GIL FERNÁNDEZ, J. (1988), *Los sonidos del lenguaje*. Síntesis, Madrid.

Muy recomendable para primer curso, es un manual que presenta de manera clara las nociones esenciales de fonética articulatoria, acústica y perceptiva, junto con un capítulo dedicado a la transcripción fonética. Contiene un glosario.

HOLMES, J. N. (1988), *Speech synthesis and recognition*. Van Nostrand Reinhold, Wokingham.

Presenta de forma mucho más detallada y técnica los contenidos del apartado 6., y constituye una buena introducción a las tecnologías del habla, aunque de un nivel algo más avanzado que el de un primer curso.

LADEFOGED, P. (1971), *Preliminaries to linguistic phonetics*. The University of Chicago Press, Chicago.

Especialmente útil en lo que se refiere a la clasificación articulatoria de los sonidos del habla presentada en el apartado 2.2.

LADEFOGED, P. (1975), *A course in phonetics*. Harcourt, Brace, Jovanovich, Nueva York. 2.^a ed., 1982, 3.^a ed., 1993.

Aunque basado especialmente en la descripción del inglés, es un manual muy pedagógico, recomendable para primer curso. Contiene, además, ejercicios al final de cada capítulo y un glosario.

P. LADEFOGED (1995), *New elements of acoustic phonetics*. The University of Chicago Press, Chicago.

Ampliación de un trabajo más sucinto de 1962, contiene las nociones básicas de fonética acústica expuestas con gran claridad.

LANDERCY, A. & RENARD, R. (1977), *Éléments de phonétique*. Didier, Bruselas.

Manual muy pedagógico, con una excelente colección de ejercicios al final y un glosario. Es muy recomendable especialmente en lo que se refiere a la fonética acústica.

LASS, N. J., ed. (1996), *Principles of experimental phonetics*. Mosby, St. Louis.

Excelente recopilación de trabajos realizados por prestigiosos especialistas que recogen diversos aspectos de la producción y la percepción del habla y de los métodos experimentales utilizados en fonética. Especialmente adecuado para estudiantes de segundo ciclo.

LAVER, A. J. (1994), *Principles of phonetics*. Cambridge University Press, Cambridge.

Manual exhaustivo y avanzado que recoge la labor del autor en el campo de la descripción fonética. Se presta especial atención a aspectos como el contexto comunicativo, la descripción y clasificación articulatoria, la organización prosódica y la transcripción.

LIEBERMAN, A & BLUMSTEIN, S. (1988), *Speech physiology, speech perception and acoustic phonetics*. Cambridge University Press, Cambridge.

Excelente manual, especialmente útil a estudiantes de segundo ciclo que necesiten una presentación más detallada de los conceptos expuestos en este capítulo.

LLISTERRI, J. (1991), *Introducción a la fonética: el método experimental*. Anthropos, Barcelona.

Dedicado especialmente al método de trabajo en fonética experimental, puede ser útil a quienes deseen iniciarse en la investigación en este campo. El capítulo primero desarrolla más ampliamente los contenidos presentados en el apartado 1.

MILLER, G. A. (1981), *Language and speech*. W. H. Freeman, San Francisco. (Trad. cast. de J. C. Gómez: *Lenguaje y habla*. Alianza, Madrid, 1985.)

Los capítulos 4. (La teoría de la fuente y el filtro), 5. (El habla) y 6. (La percepción del habla) abordan de manera muy asequible algunos de los temas discutidos aquí. Es un libro muy recomendado no sólo por su facilidad de lectura sino también por la perspectiva interdisciplinar del autor.

MINIFIE, F. D., HIXON, T. J. & WILLIAMS, F., eds. (1973), *Normal aspects of speech, hearing and language*. Prentice-Hall, Englewood Cliffs, N.J.

Obra colectiva con capítulos dedicados a la respiración, la fonación, la articulación, la acústica del habla y la audición, muy adecuados para estudiantes de segundo ciclo.

O'SHAUGHNESSY, D. (1987), *Speech communication. Human and machine*. Addison Wesley, Reading, Mass.

Manual de nivel avanzado que combina la presentación de los mecanismos de producción y percepción del habla con las aplicaciones tecnológicas.

PICKETT, J. M. (1980), *The sounds of speech communication. A primer of acoustic phonetics and speech perception*. University Park Press, Baltimore/Pro-Ed, Austin.

Especialmente recomendado para estudiantes de segundo ciclo. Presenta un tratamiento excelente de las características acústicas y perceptivas de los sonidos del habla.

PULLUM, G. K & LADUSAW, W. A. (1986), *Phonetic symbol guide*. The University of Chicago Press, Chicago.

Diccionario de símbolos utilizados en diversos alfabetos fonéticos, muy útil para consultar la utilización de los símbolos en diferentes tradiciones y su descripción fonética.

ROACH, P. (1992), *Introducing phonetics*. Penguin, Londres.

Organizado en forma de diccionario, es muy útil por la presentación concisa y de fácil consulta de los principales conceptos utilizados en fonética.